ISSN(online) 2959-5894
ISSN (print) 2959-5886

Абай атындағы Қазақ ұлттық педагогикалық университеті
Казахский национальный педагогический университет имени Абая
Abai Kazakh National Pedagogical University

ХАБАРШЫ

«Физика-математика ғылымдары» сериясы
Серия «Физико-математические науки»
Series of Physics & Mathematical Sciences
№3(95)

Алматы, 2026

Абай атындағы Қазақ ұлттық педагогикалық университеті

ХАБАРШЫ
«Физика-математика ғылымдары» сериясы №3 (95), 2026 ж.

Бас редактор:
ф.-м.ғ.д., профессор **М.А. Бектемесов**

Редакция алқасы:
Бас ред.орынбасары:
п.ғ.д., профессор **Е.Ы. Бидайбеков,**
ф.-м.ғ.д., профессор **В.Н. Косов**

Жауапты хатшылар:
п.ғ.к., қауым. профессор,
Ш.Т. Шекербекова,
п.ғ.к., қауым. профессор,
Г.А. Абдулкаримова

Ғылыми редактор
тех.ғ.д., профессор, **Н.С. Заурбеков,**
PhD доктор, қауымд. профессор м.а.
А.Р. Рысқан

Редакциялық алқа мүшелері:
PhD, Professor **A. Cabada** *(Spain),*
ф.-м.ғ.д., профессор **Ф.Ф. Комаров** *(Республика Беларусь),*
PhD, Professor **M. Ruzhansky** *(Belgium),*
PhD, **E. Kovatcheva** *(Bulgaria),*
ф.-м.ғ.д., профессор, РҒА корреспондент-мүшесі, **С.И. Кабанихин** *(Ресей),*
п.ғ.д., **профессор В.В. Гриншкун** *(Ресей),*
ф.-м.ғ.д., профессор **В.М. Лисицин** *(Ресей),*
п.ғ.д., профессор **Н.И. Пак** *(Ресей),*
ф.-м.ғ.д., профессор **А.Л. Семенов** *(Ресей),*
п.ғ.д., профессор, РБА академигі
А.Е. Абылқасымова,
PhD, профессор **К. Алимхан**,
т.ғ.д., профессор **Б.С. Ахметов**,
ф.-м.ғ.д., профессор **А.С. Бердышев**,
т.ғ.д., профессор **К. Бисембаев**,
ф.-м.ғ.д., профессор **М.Н. Калимолдаев**,
т.ғ.д., профессор **М.К. Кулбек**,
ф.-м.ғ.д., профессор **С.Т. Мухамбетжанов**,
PhD, қауым., профессор м.а.
Ж.М. Нурмухамедова,
п.ғ.д., профессор **Б.Д. Сыдықов**,
т.ғ.д., профессор **А.К. Тулешов**

© Абай атындағы Қазақ ұлттық педагогикалық университеті, 2026

Қазақстан Республикасының Ақпарат министрлігінде тіркелген
№ 4824 – Ж - 15.03.2004
(Журнал бір жылда 4 рет шығады)
2000 жылдан бастап шығады

Басуға 27.09.2026 қол қойылды
Пішімі 60x84 $^{1}/_{8}$. Көлемі 46 е.б.т.
Таралымы 300 дана. Тапсырыс 621.

050010, Алматы қаласы,
Достық даңғылы, 13
Абай атындағы ҚазҰПУ-ің "Ұлағат" баспасы

М а з м ұ н ы
С о д е р ж а н и е
C o n t e n t

МАТЕМАТИКА ЖӘНЕ ФИЗИКАЛЫҚ ПРОЦЕСТЕР МЕН МЕХАНИКАЛЫҚ ЖҮЙЕЛЕРДІ МАТЕМАТИКАЛЫҚ МОДЕЛЬДЕУ

МАТЕМАТИКА И МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ ФИЗИЧЕСКИХ ПРОЦЕССОВ И МЕХАНИЧЕСКИХ СИСТЕМ

MATHEMATICS AND MATHEMATICAL MODELING OF PHYSICAL PROCESSES AND MECHANICAL SYSTEMS

Abdullayeva B.Zh., Eralieva A.R., Amanzholov T.E., Tungatarova M.S., Aizhulov D.Y.
Sensitivity analysis of design and operating parameters affecting flat-plate solar collector efficiency in BTES……..... 7

Wang Y., Sydyk D., Omirbekov S.
A Legendre-Fenchel transform framework for mathematical modeling of nonlinear capillary flow of generalized Newtonian fluids …… 18

Жамбулатова Ә.Ж., Әлімхан Қ.
Сызықтық емес ауыспалы жүйелерді шекті уақыт ішінде тұрақтандыру әдістері……..……. 40

Жаңабергенова Н.С., Темирханова А.М.
Квазисызықты интегралдық Харди типті теңсіздіктер 49

Жетписбаева А.Е., Джумабаева А.А.
Неравенство типа Ульянова для смешанных модулей гладкости обобщённых производных Лиувилля-Вейля …….. 65

Marlambekov D., Mukhambetzhanov S., Ibraheem S.
PPO-based resource allocation for fifth-generation network slicing 74

Минглибаев М.Дж., Сапарова М.Р.
Уравнения движения в оскулирующих переменных в координатах Якоби…… 83

Рысқан А.Р., Абдрахман Б.Қ., Айнакеева Н.Ж.
Гипергеометриялық функциялары үшін екінші ретті дербес туындылы дифференциалдық теңдеулер жүйесін шешу …….. 94

Shapen K.M., Rakhymova A.T., Akhmedov A., Baishemirov Zh.D.
Bayesian causal demand modeling with latent ranking factors in e-commerce 103

Казахский национальный педагогический университет имени Абая

ВЕСТНИК
Серия «Физико-математические науки»
№3 (95), 2026 г.

Главный редактор:
д.ф.-м.н., профессор **М.А. Бектемесов**

Редакционная коллегия:

Зам.главного редактора:
д.п.н., профессор **Е.Ы. Бидайбеков,**
д.ф.-м.н., профессор **В.Н. Косов**

Ответ. секретари:
к.п.н., асс. профессор **Ш.Т. Шекербекова,**
к.п.н., асс. профессор **Г.А. Абдулкаримова**

Научный редактор
д.т.н., профессор, **Н.С. Заурбеков**,
PhD д., и.о. профессора **А.Р. Рысқан**

Члены редколлегии:
PhD, **A.Cabada** *(Spain),*
д.ф.-м.н., профессор **Ф.Ф. Комаров** (Республика Беларусь),
PhD, **M. Ruzhansky** *(Belgium),*
PhD, **E. Kovatcheva** *(Bulgaria),*
д.ф.-м.н., профессор, член-корр. РАН **С.И. Кабанихин** *(Россия),*
д.п.н., профессор **В.В. Гриншкун** *(Россия),*
д.ф.-м.н., профессор **В.М. Лисицин** *(Россия),*
д.п.н., профессор **Н.И. Пак** *(Россия),*
д.ф.-м.н., профессор **А.Л. Семенов** *(Россия),*
д.п.н., профессор, академик РАН **А.Е. Абылкасымова,**
PhD, профессор **К. Алимхан,**
д.т.н., профессор **Б.С. Ахметов**,
д.ф.-м.н., профессор **А.С. Бердышев**,
д.т.н., профессор **К. Бисембаев**,
д.ф.-м.н., профессор **М.Н. Калимолдаев**,
д.т.н., профессор **М.К. Кулбек**,
д.ф.-м.н., профессор **С.Т. Мухамбетжанов**,
PhD, асс., профессор и.о. **Ж.М. Нурмухамедова**,
д.п.н., профессор **Б.Д. Сыдықов**,
д.т.н., профессор **А.К. Тулешов**

© Казахский национальный педагогический университет им. Абая, 2026

Зарегистрирован в Министерстве информации
Республики Казахстан,
№ 4824 - Ж - 15.03.2004
(периодичность – 4 номера в год)
Выходит с 2000 года

Подписано в печать 27.09.2026.
Формат 60x84 1/8. Об. 46 уч.-изд.л.
Тираж 300 экз. Заказ 621.

050010, г. Алматы, пр. Достык, 13,
Издательство «Ұлағат» КазНПУ им. Абая

МАТЕМАТИКАНЫ ОҚЫТУ ӘДІСТЕМЕСІ

МЕТОДИКА ПРЕПОДАВАНИЯ МАТЕМАТИКИ

METHODS OF TEACHING MATHEMATICS

Биргебаев А.Б., Қыдырақын М.Е.
Функционалдық анализден есептер шығарудың ерекшеліктері 117

Mamadaliev T.S.
Digital educational resources in inclusive mathematics teaching … 128

ИНФОРМАТИКА

COMPUTER SCIENCE

Abdildayeva A., Taganova G., Tazhibeyev A., Kural A., Askarov S.
Leakage-controlled residual-calibrated hybrid model for one-hour-ahead photovoltaic generation forecasting 141

Абдрахманов Р.Е., Алимова А.Н., Соколов С.В., Чалова Н.Р., Нурсеитов Д.Б.
Набор данных и бенчмарк для тонкозернистой классификации охотничьих водоплавающих птиц……..... 150

Aibagarov S.Zh., Mukhanbet A.A.
Benchmarking handover control in multi-user 5G networks…… 168

Бегимбетов О.К., Нурсеитова А.Т., Болтайханова Т.Т., Тармышов К.Б., Тургумбаева Г.Т.
Стратегия максимизации пассажирооборота на социально значимых железнодорожных маршрутах в условиях детерминированных тарифов 176

Дүйсенбекқызы Ж., Тұрсынхан А., Газиз Г., Ілесова Б.
Трансферлік оқыту әдісі негізінде қазақ тілді балалардың сөйлеуін автоматты тану жүйесін модификациялау..... 194

Ергеш Н.С., Капалова Н.А.
Модифицированная схема разделения секрета для защиты лицевых биометрических шаблонов..... 201

Junussova M., Rakisheva D.
Machine learning for reservoir analysis and forecasting in Kazakhstan 211

Зиятбекова Г.З., Шекербаева Ш.Н., Акжолова А.И., Сыдыкова М.М., Полатова Ж.Б.
Суицидтік мінез-құлық белгілері бар қазақ тіліндегі мәтіндер моделін және қосымшасын құру……..... 220

Lebedev D., Otarbay Zh., Kyzyrkanov A.
A reproducible 5G SA testbed for O-RAN machine learning 232

Ospan A.G., Sarsembayeva T.S., Sailau A.Zh., Temirkhanova A.B., Felde I.
Optimizing chunking strategies and inference acceleration for Kazakh low-resource RAG systems. 240

Sakypbekova M., Aubakirova R.
Corpus and baseline models for Kazakh toxic comment detection 252

Abai Kazakh National Pedagogical University

BULLETIN
Ser. Physics & Mathematical Sciences

№3(95), 2026.

Editor-in-Chief
Dr. Sci. **M.A. Bektemesov**

Deputy Editor-in-Chief:
Dr. Sci. (Ped.), **Ye.Y. Bidaibekov,**
Dr. Sci. **V.N. Kosov**

Responsible editorial secretary:
Cand. Sci. (Ped.) **Sh.T. Shekerbekova**,
Cand. Sci. (Ped.) **G.A. Abdulkarimova**

Scientific editor
PhD, professor, **N.S. Zaurbekov**,
PhD, associate Professor **A.R. Ryskan**

Editorial board:
PhD, **A. Cabada** *(Spain),*
Dr. Sci. **F.F. Komarov** *(Republic of Belarus),*
PhD, **M. Ruzhansky** (Belgium),
PhD, **E. Kovatcheva** (Bulgaria),
Dr.Sc. **S.I. Kabanikhin** (Russia),
Dr.Sci. **V.V. Grinshkun** (Russia),
Dr. Sci. **V.M. Lisicin** (Russia),
Dr. Sci. (Ped.) **N.I. Pak** (Russia),
Dr. Sci. **A.L. Semenov** (Russia),
Dr. Sci. (Ped.) **A.Ye. Abylkasymova**,
*Dr.Sci***. K. Alimhan**,
Dr.Sci. **B.S. Akhmetov**,
Dr. Sci. **A.S. Berdyshev**,
Dr. Sci. **K. Bisembaev**,
Dr. Sci. **M.N. Kalimoldayev**,
Dr.Sci.(Engineering) **M.K. Kulbek**,
Dr. Sci. **S.T. Mukhambetzhanov**,
PhD, **Zh.M. Nurmukhamedova**,
Dr. Sci. (Ped.) **B.D. Sydykov**,
Dr.Sci.(Engineering) **A.K. Tuleshov**

© Abai Kazakh National Pedagogical University, 2026

Registered in the Ministry of Information of the Republic of Kazakhstan,
№ 4824 - Ж - 15.03.2004
(Periodicity: 4 issues per year)
Published since 2000

Signed to print 27/09/2026
Format 60x84 $^{1}/_{8}$. Vol. 46 p.
Printing 300 copies. Order 621.

Publishing and Editorial:
050010, 13 Dostyk av., Almaty, Kazakhstan
Publisher "Ulagat" Abai KazNPU

ИНФОРМАТИКАНЫ ОҚЫТУ ӘДІСТЕМЕСІ.
БІЛІМ БЕРУДІ АҚПАРАТТАНДЫРУ

МЕТОДИКА ПРЕПОДАВАНИЯ ИНФОРМАТИКИ.
ИНФОРМАТИЗАЦИЯ ОБРАЗОВАНИЯ

METHODS OF TEACHING COMPUTER SCIENCE.
INFORMATIZATION OF EDUCATION

Ажибеков К.Ж., Мадияров Н.К., Утеулиев Н.С., Алтынбеков Ш.Е., Ажибеков М.Ж.
Жасанды интеллектінің өзіндік жұмыстарда қолданылуы: педагогикалық бақылаудың рөлі 265

Камалова Г.Б., Турашова Ш.П., Ерсин М.
Концептуальная модель организации и цифрового сопровождения деятельности smart-аудитории в системе педагогического образования ……. 278

Найманова Д.С., Даутова А.З., Алимова Ж.С., Оспанова Н.Н.
Разработка интеллектуальной аналитической системы для анализа образовательных программ на базе LLM 295

Нурбеков Б.Ж., Байарыстан А.О., Майдисарова Д.С., Сантеева С.Ә., Толегенова Г.Б.
Жасанды интеллект негізінде жоғары математиканы оқытудың әдістемелік моделі 310

Сағымбаева А.Е., Сабраев Қ.Ж.
Жасанды интеллект жағдайында болашақ информатика мұғалімдерінің логикалық ойлау дағдыларын қалыптастырудағы педагогикалық мәселелер 322

Сыдыкова Ж.К., Ерженбек Б., Жанбекова Г.И., Нұрмухамедова Ж.М., Нурбаева Д.М.
Оқушылардың функционалдық сауаттылығын қалыптастыруда жасанды интеллектіні қолдану әдістемесі 333

Шекербекова Ш.Т., Базарбаева Ү.Қ., Қалданов Б.Б.
Болашақ информатика мұғалімдерін ақпараттық үдерістерді оқытуға дайындауда жасанды интеллект негізіндегі инфографиканы қолдану 345

Шетиева К.Ж., Салыкова А.Н., Есейкызы А., Мамбетакунов У.Э.
Системный анализ симуляторов дополненной реальности в обучении физике …… 357

МАТЕМАТИКА ЖӘНЕ ФИЗИКАЛЫҚ ПРОЦЕСТЕР МЕН МЕХАНИКАЛЫҚ ЖҮЙЕЛЕРДІ МАТЕМАТИКАЛЫҚ МОДЕЛЬДЕУ

МАТЕМАТИКА И МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ ФИЗИЧЕСКИХ ПРОЦЕССОВ И МЕХАНИЧЕСКИХ СИСТЕМ

MATHEMATICS AND MATHEMATICAL MODELING OF PHYSICAL PROCESSES AND MECHANICAL SYSTEMS

IRSTI 28.23.29, 28.17.23

https://doi.org/10.51889/2959-5894.2026.95.3.001

B.Zh. Abdullayeva[1], A.R. Eralieva[2], T.E. Amanzholov[2]*, M.S. Tungatarova[1], D.Y. Aizhulov[1]

[1] K.I. Satpayev Kazakh National Research Technical University, Almaty, Kazakhstan[1]
[2]Al-Farabi Kazakh National University, Almaty, Kazakhstan[2]
[*]*e-mail: tannur.amanzholov@gmail.com*

SENSITIVITY ANALYSIS OF DESIGN AND OPERATING PARAMETERS AFFECTING FLAT-PLATE SOLAR COLLECTOR EFFICIENCY IN BTES

Abstract

Flat-plate solar collectors are crucial in Borehole Thermal Energy Storage systems for bridging the gap between renewable energy supply and consumer heat demand. Because most studies rely on idealized models, this work develops a Python-based transient mathematical model of a "solar collector–heat exchanger" system using energy balance equations and real commercial SchücoSol collector specifications. With this verified model, a parametric and sensitivity analysis (±20%) evaluated how daily efficiency is impacted by optical efficiency, heat losses, heat exchanger transfer coefficient, fluid mass flow rate, and ambient temperature. Results demonstrate that ambient temperature, optical efficiency, linear heat losses, and heat exchanger characteristics most significantly influence efficiency. These findings quantitatively assess the impact of design and operating parameters, offering practical recommendations for designing and optimizing real seasonal thermal energy storage systems.

Keywords: solar collector, borehole thermal energy storage (BTES), seasonal thermal energy storage (STES), heat exchanger, modeling, optical efficiency.

Б.Ж. Абдуллаева[1], А.Р. Ералиева[2], Т.Е. Аманжолов[2], М.С. Тунгатарова[1], Д.Е. Айжулов[1]
[1]Казахский национальный исследовательский технический университет имени К. И. Сатпаева, г.Алматы, Казахстан
[2]Казахский национальный университет имени аль-Фараби, г.Алматы, Казахстан

АНАЛИЗ ЧУВСТВИТЕЛЬНОСТИ ЭФФЕКТИВНОСТИ ПЛОСКОГО СОЛНЕЧНОГО КОЛЛЕКТОРА В ГТА К КОНСТРУКТИВНЫМ И ЭКСПЛУАТАЦИОННЫМ ПАРАМЕТРАМ

Аннотация

Плоский солнечный коллектор является ключевым элементом систем грунтового теплового аккумулирования, необходимых для компенсации дисбаланса между потребительским спросом на тепловую энергию для нужд отопления и ГВС и предложением энергии из возобновляемых источников. Выполняя сбор солнечной энергии, коллектор значительно влияет на общую эффективность всей системы. Поскольку большинство исследований опирается на идеализированные модели, в данной работе разработана нестационарная математическая модель системы «солнечный коллектор – теплообменник» на базе уравнений баланса энергии. Модель реализована в Python с

использованием характеристик реального промышленного коллектора SchücoSol. На основе верифицированной модели выполнены параметрическое исследование и анализ чувствительности (±20 %) влияния оптического КПД, тепловых потерь, коэффициента теплопередачи теплообменника, расхода теплоносителя и температуры среды на суточную эффективность. Установлено, что наибольшее влияние на КПД оказывают температура среды, оптический КПД, линейные теплопотери и характеристики теплообменника. Результаты позволяют количественно оценить влияние конструктивных и режимных параметров и сформулировать рекомендации по проектированию и оптимизации реальных систем сезонного аккумулирования энергии.

Ключевые слова: солнечный коллектор, грунтовый тепловой аккумулятор (ГТА), сезонное хранение тепловой энергии, теплообменник, моделирование, оптический КПД

Б.Ж. Абдуллаева[1], А.Р. Ералиева[2], Т.Е. Аманжолов[2], М.С. Тунгатарова[1], Д.Е. Айжулов[1]
[1]Қ.И. Сәтбаев атындағы Қазақ ұлттық зерттеу техникалық университеті, Алматы қ., Қазақстан
[2]Әл-Фараби атындағы Қазақ ұлттық университеті, Алматы қ., Қазақстан

ЖЕР АСТЫ ЖЫЛУ САҚТАУ ЖҮЙЕСІНДЕГІ ЖАЗЫҚ КҮН КОЛЛЕКТОРЫНЫҢ ТИІМДІЛІГІНІҢ КОНСТРУКЦИЯЛЫҚ, ПАЙДАЛАНУ ПАРАМЕТРЛЕРІНЕ СЕЗІМТАЛДЫҒЫН ТАЛДАУ

Аңдатпа

Жазық күн коллекторы – жаңартылатын энергия мен тұтынушылық сұраныс арасындағы теңгерімсіздікті жоюға арналған жерасты жылу сақтау жүйелерінің негізгі элементі. Күн энергиясын жинақтай отырып, ол бүкіл жүйенің тиімділігіне айтарлықтай әсер етеді. Көптеген зерттеулер идеалдандырылған модельдерге сүйенетіндіктен, бұл жұмыста «күн коллекторы – жылу алмастырғыш» жүйесінің стационарлық емес математикалық моделі әзірленді. Модель нақты өнеркәсіптік SchücoSol коллекторының сипаттамалары қолданылып, Python тілінде жүзеге асырылды. Верификацияланған модель негізінде оптикалық ПӘК, жылу шығындары, жылу алмастырғыштың жылу беру коэффициенті, жылу тасымалдағыш шығыны және қоршаған орта температурасының тәуліктік тиімділікке әсерін бағалау үшін параметрлік және сезімталдық (±20%) талдауы жүргізілді. Нәтижесінде ПӘК-ке қоршаған орта температурасы, оптикалық ПӘК, сызықтық жылу шығындары және жылу алмастырғыш сипаттамалары басты әсер ететіні анықталды. Алынған мәліметтер құрылымдық және режимдік параметрлердің әсерін сандық бағалауға, сондай-ақ маусымдық энергия жинақтаудың нақты жүйелерін жобалау мен оңтайландыру бойынша практикалық ұсыныстар әзірлеуге мүмкіндік береді.

Түйін сөздер: күн коллекторы, жер асты жылу аккумуляторы, жылу энергиясын маусымдық жинақтау, жылу алмастырғыш, модельдеу, оптикалық ПӘК.

Introduction

Heating and cooling account for approximately half of Europe’s energy consumption, with 85% supplied by fossil fuels [1], while buildings consume up to 40% of the EU’s total energy [2]. In Canada, space and water heating generate over 80% of residential greenhouse gas emissions [3]. These figures highlight the need for low-carbon heating systems. Solar energy offers a promising alternative, but its availability is seasonally mismatched with heating demand: solar radiation peaks in summer, whereas heat demand peaks in winter [4]. Seasonal Thermal Energy Storage (STES) addresses this imbalance by storing surplus renewable heat collected in summer for winter use [5].

Among underground sensible heat storage technologies, Borehole Thermal Energy Storage (BTES) offers flexibility, scalability, and cost-effectiveness [6]. It uses soil or rock as the storage medium (Figure 1), exchanging heat through arrays of vertical boreholes, typically 30–200 m deep, equipped with U-shaped tubes carrying a heat-transfer fluid [6, 7]. Unlike tank storage or Aquifer Thermal Energy Storage (ATES), BTES requires no artificial reservoirs and is independent of aquifer availability. Its modular design allows expansion by drilling additional boreholes [7].

Integrating solar collectors with seasonal storage forms the basis of Central Solar Heating Plants with Seasonal Storage [2]. Collectors supply surplus summer heat, which a circulating fluid transfers through heat exchangers to BTES for storage until winter. A notable example is Canada’s Drake Landing Solar Community (DLSC), completed in 2006. It combines 800 flat-plate collectors installed

on garage roofs with 144 boreholes, each 35 m deep [8]. After five years of operation, DLSC achieved a solar fraction of 97%, meeting nearly all the heating needs of 52 homes with stored summer solar energy [6]. Collector selection strongly influences overall STES performance. Flat-plate solar collectors (FPCs) are widely used because of their reliability and suitability for low-grade heat [9]. However, their efficiency decreases substantially at fluid temperatures of 70–95°C, often required for effective seasonal storage charging [10].

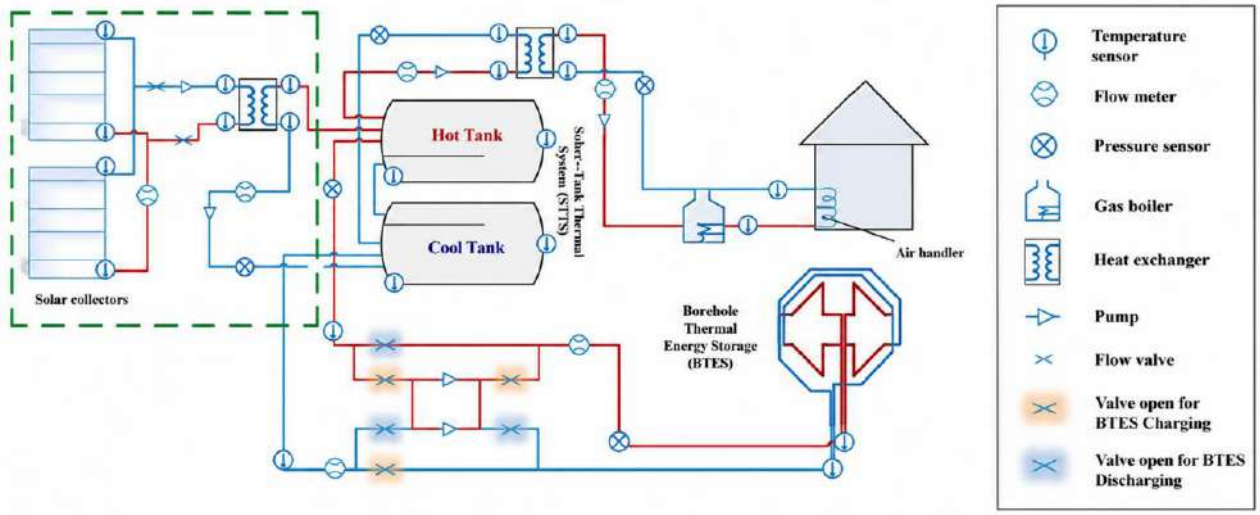


Figure 1. Schematic diagram of a borehole thermal energy storage system

Hybrid configurations improve solar heat collection at high temperatures. At the Taars plant in Denmark, flat-plate collectors preheat the fluid, while parabolic trough collectors (PTCs) raise its temperature to approximately 95°C for storage or district heating [10].

Solar–STES integration nevertheless faces several challenges. As the borehole field warms, higher collector inlet temperatures reduce collection efficiency, while high storage temperatures increase ground heat losses. At DLSC, approximately 60% of the injected heat was lost [2]. Insufficient summer replenishment also causes progressive ground cooling, reducing heat pump performance and increasing electricity consumption [11]. Furthermore, drilling and collector installation costs hinder commercial deployment without substantial subsidies [6]. Low-temperature BTES coupled with ground-source heat pumps offers a potential solution. Storage temperatures below approximately 35°C reduce heat losses and improve collector efficiency; heat pumps then raise the temperature for building heating [6]. Phase-change materials can increase storage density and moderate temperature fluctuations [12], while hybrid FPC–PTC systems improve heat generation across different temperature ranges [10]. However, many simulations inadequately represent practical operation. Electricity-price-based control of heat pumps and storage charging remains insufficiently explored [11]. Balancing collector array size, borehole storage costs, and thermal performance to minimize the levelized cost of heat is another unresolved challenge [10, 13]. Detailed analysis using commercial collector specifications is therefore needed to assess how actual design characteristics affect seasonal efficiency. This study investigates the sensitivity of collector thermal efficiency to design and operating parameters. The objectives are to compare model results with published calculations [14], use commercial collector specifications, and assess subsystem performance relevant to STES integration. Ground storage dynamics and seasonal BTES efficiency are outside the present model.

Research methodology

Thermal state model of a solar collector. The modeled subsystem comprises a flat-plate collector, a primary circulation loop, and a heat exchanger within the solar circuit shown in Figure 1. It supplies heat to the downstream storage system; the borehole field itself is not modeled. The collector comprises an insulated housing, a transparent cover, an absorber, and copper tubes (Figure 2a). Solar

radiation absorbed by the plate is transferred through the tubes to the circulating heat-transfer fluid, while conduction, convection, and radiation cause heat losses (Figure 2b).

The heat exchanger contains primary and secondary sections separated by a heat-transfer partition (Figure 2c). The primary fluid releases heat received from the collector, while the secondary fluid transports it to the storage tank.

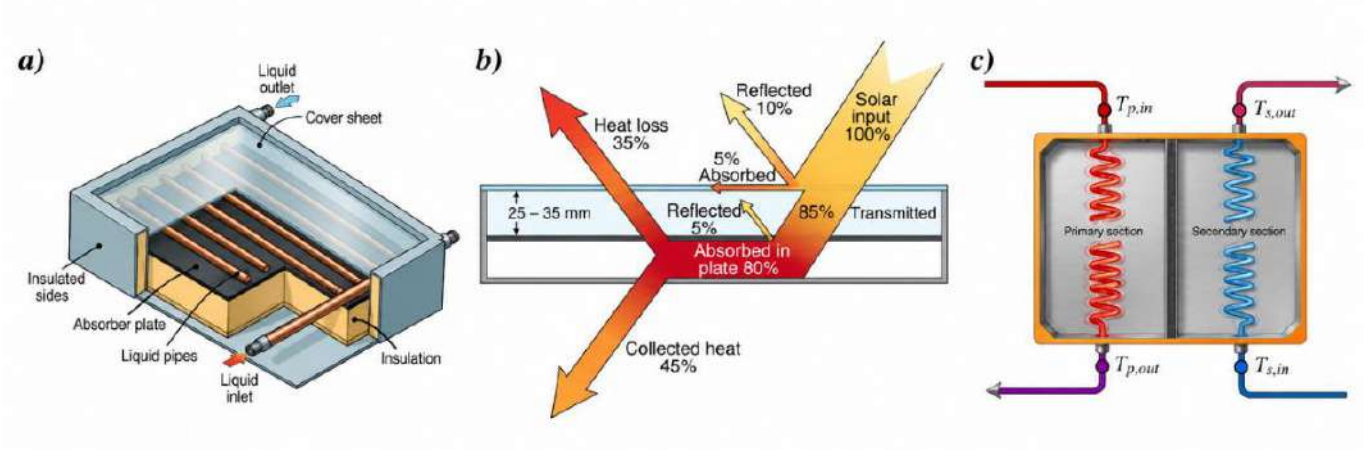


Figure 2. Solar thermal system components: (a) flat-plate solar collector construction; (b) solar energy distribution and heat losses in the collector; (c) heat exchanger schematic showing the primary and secondary sections and fluid inlet and outlet temperatures.

The model predicts the time-dependent collector outlet temperature and the heat exchanger outlet temperatures in both circuits, T_1 and T_2, for a prescribed daily solar radiation profile. The following assumptions apply:

1) Each modeled fluid volume is perfectly mixed, with uniform temperature and no stratification.
2) Fluid thermal properties and flow rates in both circuits remain constant.
3) The absorber temperature equals approximately the mean collector inlet and outlet fluid temperature.
4) Connecting pipelines are perfectly insulated; their heat losses are neglected.

The collector energy balance includes absorbed solar power, thermal accumulation, environmental losses, and heat removal by the circulating fluid:

$$\dot{Q}_{\text{st}} = \dot{Q}_{\text{in}} - \dot{Q}_{\text{out}} + \dot{Q}_{\text{g}} \tag{1}$$

where Qst is the thermal accumulation rate, Qin is absorbed solar power, Qout comprises environmental losses and fluid heat removal, and Qg is auxiliary heating power (all in W). Here, Qg = 0.

Thermal accumulation in the collector fluid is expressed as:

$$\dot{Q}_{\text{st}} = \rho c \frac{dT_{\text{c,out}}}{dt} V \tag{2}$$

where ρ [kg/m³] is fluid density, cp [J/(kg·K)] is specific heat capacity, V [m³] is fluid volume, and Tc,out [°C] is collector outlet temperature.

The solar energy arriving at the collector surface is determined by expression (3).

$$\dot{Q}_{\text{sol}} = IA \tag{3}$$

The absorbed solar power accounts for cover transmittance and absorber absorptance:

$$\dot{Q}_{\text{in}} = \dot{Q}_{\text{abs}} = I\eta_0 A \tag{4}$$

where I $[W \cdot m^{-2}]$ – solar radiation falling on the collector surface, A $[m^2]$ – collector area, $\eta_0 = \tau\alpha$ – optical efficiency of the collector, τ – transmittance of a transparent coating, α – absorption coefficient of the absorber.

The total heat leaving the collector control volume comprises environmental losses and useful heat transferred to the fluid:

$$\dot{Q}_{\text{out}} = \dot{Q}_{\text{loss}} + \dot{Q}_{\text{F}} \tag{5}$$

where $\dot{Q}_{\text{loss}}$ [W], the heat losses of the collector to the environment are described as:

$$\dot{Q}_{\text{loss}} = U_{\text{L}} A (T_{\text{abs}} - T_{\text{amb}}) \tag{6}$$

where U_{L} $[W \cdot m^{-2} \cdot °C^{-1}]$ – heat loss coefficient, T_{abs} [°C] – absorber temperature, T_{amb} [°C] – ambient temperature.

The thermal power $\dot{Q}_{\text{F}}$ [W] transferred to the coolant is determined by the expression:

$$Q_{\text{F}} = \dot{m} c (T_{\text{c,out}} - T_{\text{c,in}}) \tag{7}$$

where ṁ [kg/s] is mass flow rate, c [J/(kg·K)] is fluid specific heat capacity, and $T_{\text{c,in}}$ [°C] is collector inlet temperature.

Based on the systems of equations (1)–(7), a differential equation for the heat balance of the collector is compiled:

$$C \frac{dT_{\text{c,out}}}{dt} = IA\eta_0 - U_{\text{L}} A (T_{\text{abs}} - T_{\text{amb}}) - \dot{m} c (T_{\text{c,out}} - T_{\text{c,in}}) \tag{8}$$

where C = ρcpV [J/K] is the fluid heat capacity and t [s] is time.

Equation (8) describes coolant temperature evolution under solar input, environmental losses, and fluid flow. Only dominant upper-surface heat losses are considered. Coolant density, specific heat capacity, and volume are assumed constant [6], while the absorber temperature equals the mean coolant temperature:

$$T_{\text{abs}} = T_{\text{avg}} = \frac{T_{\text{c,in}} + T_{\text{c,out}}}{2} \tag{9}$$

where T_{abs} [°C] – average temperature of the coolant in the collector. Substituting (9) into (8) yields the governing equation for the flat-plate solar collector:

$$\frac{dT_{\text{c,out}}}{dt} = \frac{IA\eta_0}{C} - \frac{U_{\text{L}} A}{C} (T_{\text{avg}} - T_{\text{amb}}) - \frac{\rho F_{\text{c}} c}{\text{C}} (T_{\text{c,out}} - T_{\text{c,in}}) \tag{10}$$

where Fc [m³/s] is the primary volumetric flow rate.

The heat exchanger is represented by two interacting, perfectly mixed fluid volumes, coupled through thermal conductance UA. The primary outlet temperature satisfies:

$$\frac{dT_{\text{p,out}}}{dt} = \frac{F_{\text{c}}}{V_1} (T_{\text{p,in}} - T_{\text{p,out}}) - \frac{UA}{\rho c V_1} (T_{\text{p,out}} - T_{\text{s,out}}) \tag{11}$$

where $T_{\text{p,in}}$ and $T_{\text{p,out}}$ are primary inlet and outlet temperatures, $T_{\text{s,out}}$ is secondary outlet temperature, V_1 is primary fluid volume, and UA [W/K] is heat exchanger conductance. With insulated connections, $T_{\text{p,in}} = T_{\text{c,out}}$ and $T_{\text{p,out}} = T_{\text{c,in}}$.

The change in the coolant temperature at the outlet of the secondary circuit is determined by the expression:

$$\frac{dT_{s,out}}{dt} = \frac{F_1}{V_2}(T_{s,in} - T_{s,out}) + \frac{UA}{\rho c V_2}(T_{p,out} - T_{s,out}) \quad (12)$$

where V_2 is secondary fluid volume, Ts,in is secondary inlet temperature, and F_1 is secondary volumetric flow rate [m³/s]. A variable solar insolation profile was used to simulate the collector's daily operating mode. The daily variation in solar radiation was approximated by a sinusoidal function between sunrise and sunset:

$$I(t) = I_{max} \sin\left[\frac{\pi(t-t_{rise})}{t_{set}-t_{rise}}\right] \quad (13)$$

where I $[W \cdot m^{-2}]$ – intensity of solar radiation at a given moment in time t, I_{max} $[W \cdot m^{-2}]$ – the maximum intensity of solar radiation corresponding to a clear summer day ($I_{max} = 900\ W \cdot m^{-2}$), t_{rise} – sunrise time, t_{set} – sunset time. The instantaneous efficiency of the collector was determined by the formula:

$$\eta = \eta_0 - k_1 \frac{T_{avg}-T_{amb}}{I} - k_2 \frac{(T_{avg}-T_{amb})|T_{avg}-T_{amb}|}{I} \quad (14)$$

where η – instantaneous collector efficiency, k_1 $[W \cdot m^{-2} \cdot °C^{-1}]$ – linear heat loss coefficient, k_2 $[W \cdot m^{-2} \cdot °C^{-2}]$ – coefficient of square heat loss. The daily efficiency of the collector was estimated by the integral efficiency:

$$\eta_{day} = \frac{\int (Q_{sol}-Q_{loss})dt}{A \int I dt} \quad (15)$$

where η_{day} – daily efficiency of the system. The initial conditions for the numerical simulation were set based on the assumption that at the initial moment of time the entire system is in thermal equilibrium with the environment:

$$T_{c,out}(0) = T_{c,in}(0) = T_{p,in}(0) = T_{p,out}(0) = T_{s,in}(0) = T_{s,out}(0) = T_a \quad (16)$$

The model was implemented in Python using NumPy. Figure 3 compares collector and heat exchanger temperatures with published calculations [14]. In this verification case, circulation is initially absent ($F_c = 0$), producing a collector temperature peak; pump activation then transfers heat to the exchanger. This comparison checks implementation consistency rather than providing experimental validation. The subsequent daily simulations use SchücoSol K specifications and a separate solar input profile.

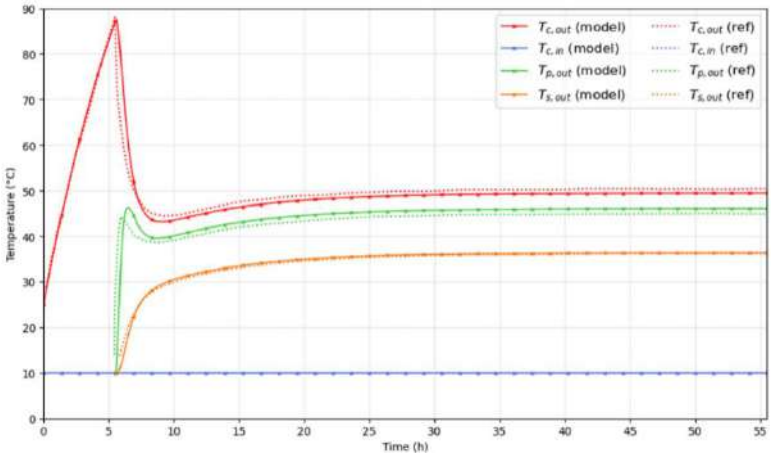


Figure 3. Comparison of the calculated temperature data of the model with the data of work [14]

Discussion of Results

Collector specifications were taken from the SchücoSol K product sheet [15] after comparison with published calculations [14]. Each of six parameters was varied independently while the others remained at their baseline values (Table 1). Daily collector efficiency was the evaluation criterion. The full-range efficiency spread is defined as:

$$\Delta\eta = \eta_{\max} - \eta_{\min} \tag{17}$$

Table 1. Research parameters and variation ranges

Parameter	*Base value*	*Variation range*	*Unit*
Optical efficiency η_0	*0,799*	*0,60 – 0,92*	–
Linear losses k_1	*3,97*	*1,5 – 8,0*	$W/(m^2 \cdot K)$
Square-law losses k_2	*0,016*	*0,004 – 0,040*	$W/(m^2 \cdot K^2)$
Pump flow rate F_c	*2,5*	*0,5 – 8,0*	l/min
UA of the heat exchanger	*45*	*10 – 150*	W/K
Air temperature T_a	*10*	*−10 – +40*	°C

Figure 4 presents the baseline daily simulation using η_0 = 0.799, k_1 = 3.97 W/(m²·K), k_2 = 0.016 W/(m²·K²), and A = 2.32 m². Solar collection occurs between 06:00 and 18:00, with the collector outlet reaching approximately 66°C (Figure 4a). The heat exchanger temperatures remain nearly constant after circulation stops, while the collector cools. Figure 4b shows absorbed solar power, environmental heat losses, and useful heat collection, expressed in W.

Figure 5 shows the influence of six parameters on daily collector efficiency, with a common baseline of 48.5%. Increasing η0 from 0.60 to 0.92 raises efficiency from 36.9% to 55.4%; increasing Ta from −10 to 40°C raises it from 41.3% to 58.9%.

UA and F_c show diminishing returns. Efficiency rises from 30.4% at UA = 10 W/K to 54.3% at 150 W/K, while increasing flow from 0.5 to 8 L/min raises it from 44.7% to 49.2%. Higher flow therefore offers limited additional thermal benefit.

Increasing k_1 from 1.5 to 8.0 W/(m²·K) reduces efficiency from 58.9% to 37.2%; increasing k_2 from 0.005 to 0.040 W/(m²·K²) reduces it from 50.5% to 45.3%. The linear loss coefficient has the larger influence over these ranges.

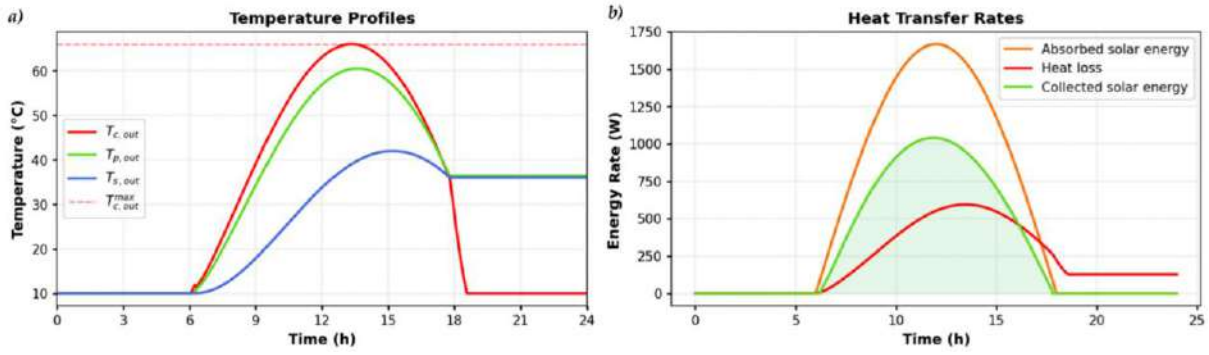


Figure 4. Baseline daily operation: (a) fluid temperatures; (b) absorbed solar power, heat losses, and useful heat collection.

Figure 5 summarizes the influence of the full investigated parameter ranges on daily collector efficiency. The largest efficiency spread is associated with the heat exchanger conductance UA (23.9 percentage points), followed by the linear heat loss coefficient k_1 (21.7), optical efficiency η_0 (18.5), and ambient temperature T_a (17.6). The corresponding spreads for the quadratic heat loss coefficient k_2 and volumetric flow rate F_c are 5.2 and 4.5 percentage points, respectively.

These results characterize the overall influence of the selected parameter ranges and should not be interpreted as normalized local sensitivities.

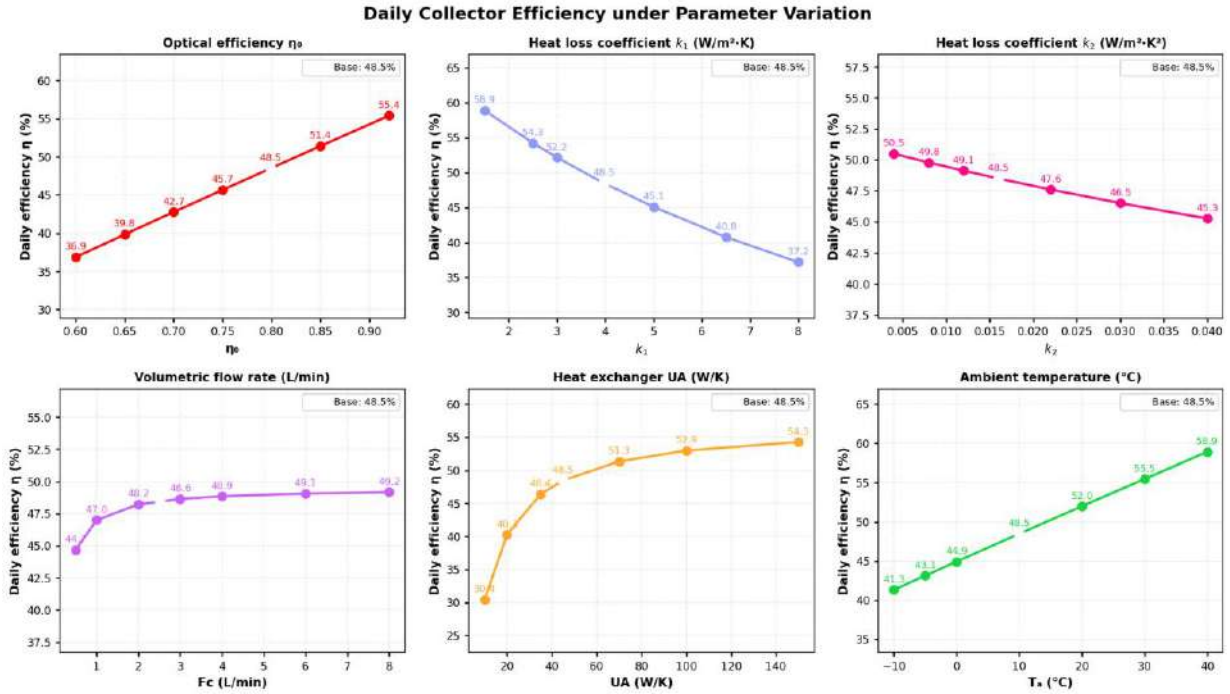


Figure 5. Dependence of the daily efficiency of a solar collector on changes in key operational, design parameters and external factors

Figure 6a presents a local sensitivity analysis of daily collector efficiency, in which each parameter is varied by ±20% around its baseline value. This analysis evaluates the response of collector efficiency to moderate parameter variations under baseline operating conditions and complements the broader parametric analysis presented in Figure 5.

Figure 6b compares the operating trajectory with the instantaneous efficiency characteristic. During high insolation, operating points occupy the low reduced-temperature region, with efficiency approximately 0.55–0.73.

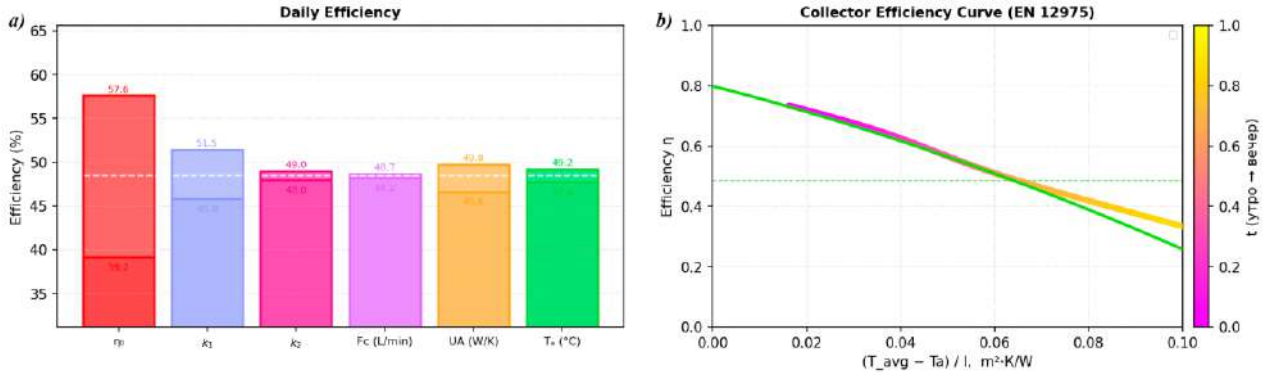


Figure 6. a) Daily efficiency ranges under ±20% variations in the model parameters; (b) instantaneous efficiency characteristic and operating trajectory.

Figure 7 compares useful collected heat and maximum collector outlet temperature. Increasing optical efficiency or ambient temperature increases both quantities, whereas increasing either heat

loss coefficient reduces them. The useful heat gain with higher optical efficiency is accompanied by a higher peak temperature.

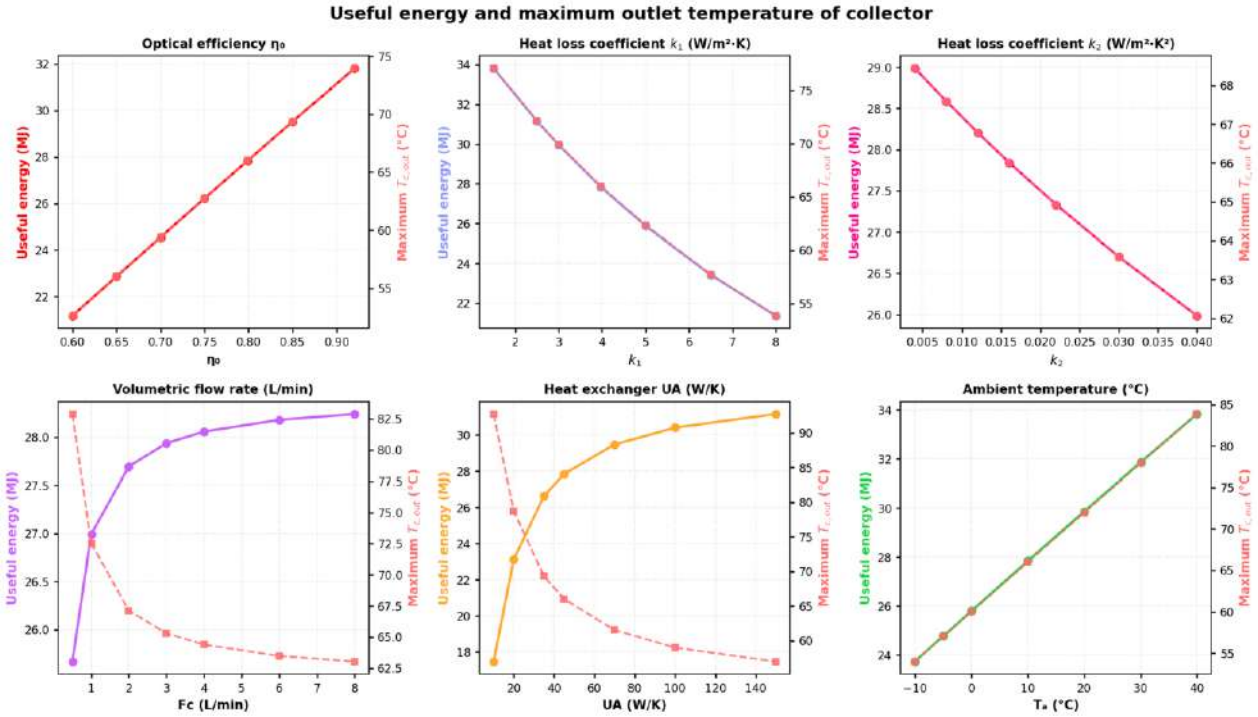


Figure 7. Results of the coupled analysis of useful thermal energy and peak temperature conditions of a solar collector

Fc and UA instead produce opposing energy and temperature trends. Increasing F_c from 0.5 to 8 L/min raises useful heat from approximately 25.7 to 28.2 MJ while reducing the peak temperature from 83 to 63°C. Increasing UA from 10 to 150 W/K raises useful heat from approximately 17.5 to 31 MJ and lowers the peak temperature from 93 to 58°C. Improved heat removal therefore limits overheating, although gains diminish at higher flow and conductance.

Table 2 summarizes the daily efficiency endpoints displayed in Figure 5. Its spreads describe the selected full variation ranges, which differ from the local perturbations in Figure 8. They do not establish a universal sensitivity ranking or an economic optimum.

Table 2. Daily collector efficiency ranges read from Figure 5

Parameter	*ηmin (%)*	*ηmax (%)*	*Δη (percentage points)*
Air temperature T_a	*41,3*	*58,9*	*17,6*
Optical efficiency η_0	*36,9*	*55,4*	*18,5*
Linear heat loss coefficient k_1	*37,2*	*58,9*	*21,7*
Heat exchanger conductance UA	*30,4*	*54,3*	*23,9*
Pump flow rate F_c	*44,7*	*49,2*	*4,5*
Quadratic heat loss coefficient k_2	*45,3*	*50,5*	*5,2*

Figure 8 presents local sensitivity around the 48.5% baseline. Optical efficiency has the largest reported effect, followed by U_{eff} and UA; T_a and F_c have smaller effects. This five-parameter analysis differs from the six-parameter sweeps in Figures 5–6 because it uses an effective loss coefficient U_{eff}. Local sensitivities describe behavior near the baseline and cannot be ranked directly against the full-range efficiency spreads.

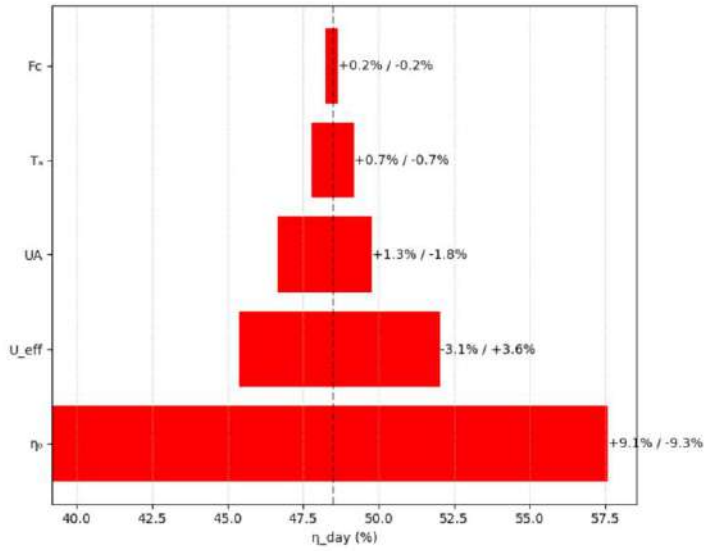


Figure 8. Reported local sensitivity of daily efficiency to ±20% parameter variations

Conclusions

In this paper, the heat transfer processes of a flat-plate solar collector in a water heating system were studied using dynamic mathematical modeling. The following results were obtained and the proposed problems were solved.

A non-stationary mathematical model of the solar collector was developed based on the equations of EN 12975 and implemented numerically using the Euler method with a 10-s step size. The model takes into account the thermal inertia of the system (C = 19,502 J/K), heat transfer between the elements, and the dynamics of the storage tank. Verification showed good agreement between the instantaneous efficiency and the EN 12975 specification (deviation ≤ 8 10^{-4}). A parametric analysis based on the Hottel–Whillier–Bliss equation revealed the following dynamics of parameter influences on daily system efficiency:

- Across full physical parameter ranges, daily efficiency varied from 30.4% to 58.9%. Under these conditions, the largest efficiency spread ($\Delta\eta$) was observed when varying heat exchanger conductance (UA, up to 23.9 percentage points), linear heat loss coefficient (k_1, 21.7 percentage points), optical efficiency (η_0, 18.5 percentage points), and ambient temperature (T_a, 17.6 percentage points). The influence of the quadratic heat loss coefficient (k_2) and coolant flow rate (F_c) remained minor ($\Delta\eta \leq 5.2$ percentage points).
- Local sensitivity analysis ($\pm 20\%$ variation): Optical efficiency (η_0) emerged as the dominant factor, causing an efficiency spread of 18.4 percentage points (ranging between 39.2% and 57.6%). The impact of other parameters under local variations was less pronounced: k_1 yielded a spread of 5.7 percentage points, UA 3.1 percentage points, T_a 1.4 percentage points, k_2 1.0 percentage point, and coolant flow rate F_c only 0.4 percentage points.

Based on these findings, practical design recommendations for cold-climate regions (specifically Kazakhstan) include prioritizing collectors with low linear heat losses ($k_1 \leq 2.5$ W/(m² K)), high optical efficiency ($\eta_0 \geq 0.85$), and sufficient heat exchanger conductance ($UA \geq 50\ W/K$). Furthermore, operating at moderate coolant flow rates ($F_c \approx 2.5 - 3\ L/min$) strikes an optimal balance between thermal performance and pumping power requirements. Future research will focus on extending the dynamic model by integrating hybrid energy sources, phase-change materials, and optimizing the thermal capacity of storage systems.

Acknowledgment

This research has been funded by the Committee of Science of the Ministry of Science and Higher Education of the Republic of Kazakhstan (Grant No. AP26103573).

References

[1] Borko, K. Insights into Aquifer and Borehole Thermal Energy Storage Systems for Slovenia's Energy Transition [Text] / K. Borko [et al.] // Energies. – 2025. – Vol. 18, no. 5. – P. 1019. – DOI: 10.3390/en18051019.

[2] Pavlov, G. K. Seasonal Ground Solar Thermal Energy Storage - Review of Systems and Applications [Text] / G. K. Pavlov, B. W. Olesen // Proceedings. – 2011.

[3] Mesquita, L. Drake Landing Solar Community: 10 Years of Operation [Text] / L. Mesquita [et al.] // Proc. ISES Solar World Congress 2017 / IEA SHC International Conference on Solar Heating and Cooling for Buildings and Industry. – Abu Dhabi, UAE, 2017. – DOI: 10.18086/swc.2017.06.09.

[4] Kamali, M. Solar-Assisted Seasonal Aquifer Thermal Energy Storage in a Relatively Deep Geothermal Aquifer for Urban Heating: A Canadian Case Study [Text] / M. Kamali [et al.] // Processes. – 2026. – Vol. 14, no. 10. – P. 1636. – DOI: 10.3390/pr14101636.

[5] Ghoreishi-Madiseh, S. A. Application of Borehole Thermal Energy Storage in Waste Heat Recovery from Diesel Generators in Remote Cold Climate Locations [Text] / S. A. Ghoreishi-Madiseh [et al.] // Energies. – 2019. – Vol. 12, no. 4. – P. 656. – DOI: 10.3390/en12040656.

[6] Rad, F. M. An integrated model for designing a solar community heating system with borehole thermal storage [Text] / F. M. Rad, A. S. Fung, M. A. Rosen // Energy for Sustainable Development. – 2017. – Vol. 36. – P. 6–15. – DOI: 10.1016/j.esd.2016.10.003.

[7] Nilsson, E. Borehole Thermal Energy Storage Systems for Storage of Industrial Excess Heat: Performance Evaluation and Modelling [Text] : Licentiate Thesis / E. Nilsson, P. Rohdin. – Linköping University, Linköping, Sweden, 2020. – DOI: 10.3384/lic.diva-163789.

[8] Wong, B. Drake Landing Solar Community: Financial Summary and Lessons Learned [Text] / B. Wong, L. Mesquita // Proceedings of the ISES Solar World Congress 2019. – 2019. – DOI: 10.18086/swc.2019.11.03.

[9] Ayompe, L. M. Analysis of the thermal performance of a solar water heating system with flat plate collectors in a temperate climate [Text] / L. M. Ayompe, A. Duffy // Applied Thermal Engineering. – 2013. – Vol. 58, no. 1–2. – P. 447–454. – DOI: 10.1016/j.applthermaleng.2013.04.062.

[10] Tian, Z. Solar heating plants with flat plate collectors and parabolic trough collectors [Text] / Z. Tian. – Technical University of Denmark, Department of Civil Engineering, 2018. – (B Y G D T U. Rapport ; No. R-394).

[11] Behzadi, A. Dynamic seasonal energy management of borehole thermal energy storage and smart heat pump synergies in fossil-free, ultra-efficient buildings [Text] / A. Behzadi [et al.] // Applied Energy. – 2026. – Vol. 406. – P. 127261. – DOI: 10.1016/j.apenergy.2025.127261.

[12] Xiao, H. Application of Photovoltaic and Solar Thermal Technologies in Buildings: A Mini-Review [Text] / H. Xiao [et al.] // Coatings. – 2024. – Vol. 14, no. 3. – P. 257. – DOI: 10.3390/coatings14030257.

[13] Beaufait, R. Techno-Economic Analysis of a Seasonal Thermal Energy Storage System with 3-Dimensional Horizontally Directed Boreholes [Text] / R. Beaufait [et al.] // Thermo. – 2022. – Vol. 2, no. 4. – P. 453–481. – DOI: 10.3390/thermo2040030.

[14] Mastný, P. Mathematical Modeling of Basic Parts of Heating Systems with Alternative Power Sources [Text] / P. Mastný, J. Morávek, J. Pitron // Recent Advances in Fluid Mechanics and Thermal Engineering. – Salerno, Italy, 2015. – P. 126–131.

[15] Flat−plate solar collectors, SchücoSol K [Electronic resource] : product information sheet. – URL: https://www.construction21.org/espana/data/sources/users/1651/schucosol-solar-thermal-compact-line-product-information-sheet.pdf

IRSTI 27.35.21

https://doi.org/10.51889/2959-5894.2026.95.3.002

Y. Wang[1]*, D. Sydyk[1], S. Omirbekov[1]

[1]Nazarbayev University, Astana, Kazakhstan
e-mail: yanwei.wang@nu.edu.kz

A LEGENDRE-FENCHEL TRANSFORM FRAMEWORK FOR MATHEMATICAL MODELING OF NONLINEAR CAPILLARY FLOW OF GENERALIZED NEWTONIAN FLUIDS

Abstract

A Legendre-Fenchel transform framework is developed for pressure-driven capillary flow of monotone generalized Newtonian fluids. A monotone nonlinear flow curve defines a convex dissipation potential, and its Legendre-Fenchel transform gives a complementary potential in the stress variable. Since the stress distribution is known explicitly in fully developed capillary flow, the velocity profile can be obtained from this complementary potential without requiring the inverse constitutive relation to be written explicitly. This reformulates nonlinear capillary-flow modeling from a model-by-model constitutive inversion problem into a convex-duality construction. The method recovers the shear-rate-parametrized Carreau-Yasuda solution, gives compact average-velocity relations for circular and planar geometries, and extends to several generalized Newtonian models, including Papanastasiou-regularized yield-stress fluids. In the ideal yield-stress limit, plug formation follows naturally from the subdifferential structure of the dissipation potential and the corresponding flat interval of the complementary potential. The framework provides a unified mathematical structure for analytical and semi-analytical pressure-drop-flow-rate relations in nonlinear capillary flow.

Keywords: Legendre-Fenchel transform; capillary flow; generalized Newtonian fluids; convex duality; non-Newtonian rheology; pressure-driven flow; yield-stress fluids.

Я. Ванг[1], Д. Сыдық[1], С. Омирбеков[1]
[1]Назарбаев Университет, г.Астана, Казахстан

ПРЕОБРАЗОВАНИЕ ЛЕЖАНДРА-ФЕНХЕЛЯ ДЛЯ МАТЕМАТИЧЕСКОГО МОДЕЛИРОВАНИЯ НЕЛИНЕЙНОГО КАПИЛЛЯРНОГО ТЕЧЕНИЯ ОБОБЩЁННЫХ НЬЮТОНОВСКИХ ЖИДКОСТЕЙ

Аннотация

В работе разработана схема преобразования Лежандра-Фенхеля для напорного капиллярного течения монотонных обобщённых ньютоновских жидкостей. Монотонная нелинейная кривая течения задаёт выпуклый потенциал диссипации, а её преобразование Лежандра-Фенхеля формирует дополнительный потенциал в переменной напряжения. Поскольку распределение напряжения в полностью развитом капиллярном течении известно из уравнения импульса, профиль скорости может быть получен непосредственно из дополнительного потенциала без явного обращения определяющего соотношения. Такой подход переводит моделирование нелинейного капиллярного течения из задачи обращения отдельных реологических моделей в единую конструкцию выпуклой двойственности. Метод воспроизводит параметризованное по скорости сдвига решение Карро-Ясуды, даёт компактные соотношения для средней скорости в круглой и плоской геометриях и распространяется на ряд обобщённых ньютоновских моделей, включая регуляризованные модели сред с пределом текучести. В пределе идеальной среды с пределом текучести образование пробковой зоны естественно следует из субдифференциальной структуры потенциала диссипации. Предложенная схема обеспечивает единое математическое описание аналитических и полуаналитических зависимостей между перепадом давления и расходом в нелинейном капиллярном течении.

Ключевые слова: преобразование Лежандра-Фенхеля; капиллярное течение; обобщённые ньютоновские жидкости; выпуклая двойственность; неньютоновская реология; напорное течение; жидкости с пределом текучести.

Я. Ванг[1], Д. Сыдық[1], С. Өмірбеков[1]
[1]Назарбаев университеті, Астана қ., Қазақстан

ЖАЛПЫЛАНҒАН НЬЮТОНДЫҚ СҰЙЫҚТАРДЫҢ СЫЗЫҚТЫҚ ЕМЕС КАПИЛЛЯРЛЫҚ АҒЫНЫН МАТЕМАТИКАЛЫҚ МОДЕЛЬДЕУГЕ АРНАЛҒАН ЛЕЖАНДР-ФЕНХЕЛЬ ТҮРЛЕНДІРУІ

Аңдатпа

Бұл жұмыста монотонды жалпыланған ньютондық сұйықтардың қысым әсерінен болатын капиллярлық ағынын сипаттау үшін Лежандр–Фенхель түрлендіруіне негізделген тәсіл ұсынылады. Монотонды сызықтық емес ағын қисығы дөңес диссипация потенциалын анықтайды, ал оның Лежандр-Фенхель түрлендіруі кернеу айнымалысындағы қосымша потенциалды береді. Толық дамыған капиллярлық ағында кернеудің таралуы импульс теңдеуінен белгілі болғандықтан, жылдамдық профилін анықтаушы теңдеуді айқын түрде кері шығармай-ақ осы қосымша потенциал арқылы алуға болады. Осылайша, сызықтық емес капиллярлық ағынды модельдеу жеке реологиялық модельдерді кері айналдыру мәселесінен дөңес дуалдылыққа негізделген бірыңғай математикалық құрылымға көшіріледі. Әдіс Карро-Ясуда моделінің ығысу жылдамдығымен параметрленген шешімін қалпына келтіреді, дөңгелек және жазық геометриялар үшін орташа жылдамдықтың ықшам қатынастарын береді және Папанастасиу бойынша регуляризацияланған аққыштық шегі бар сұйықтарды қоса алғанда бірнеше жалпыланған ньютондық модельдерге қолданылады. Идеал аққыштық шегі бар сұйық шегінде тығын тәрізді аймақтың пайда болуы диссипация потенциалының субдифференциалдық құрылымынан табиғи түрде шығады. Ұсынылған тәсіл сызықтық емес капиллярлық ағындағы қысым түсуі мен шығын арасындағы аналитикалық және жартылай аналитикалық тәуелділіктерді сипаттайтын біртұтас математикалық негіз береді.

Түйін сөздер: Лежандр-Фенхель түрлендіруі; капиллярлық ағын; жалпыланған ньютондық сұйықтар; дөңес дуалдылық; ньютондық емес реология; қысыммен қозғалатын ағын; аққыштық шегі бар сұйықтар.

1 Introduction

Main provisions

A monotone flow curve of a generalized Newtonian fluid can be represented by a convex dissipation potential, while its Legendre-Fenchel conjugate provides a complementary potential in the stress variable. For fully developed capillary flow, the stress distribution is known from momentum balance; therefore, the velocity profile follows directly from the complementary potential without explicit inversion of the constitutive law. The framework recovers the Carreau-Yasuda solution and gives compact average-velocity relations for circular and planar geometries. The formulation also applies to Cross, Sisko, Ellis, Bingham–Papanastasiou and Herschel–Bulkley–Papanastasiou fluids, including ideal yield-stress limits through subdifferentials.

Pressure-driven flow in capillary geometries is a classical problem in fluid mechanics, transport phenomena, and rheology [1,2]. In its simplest form, the problem asks how a prescribed pressure gradient determines the velocity profile and the corresponding relation between flow rate and pressure drop. For a Newtonian fluid, the linear relation between shear stress and shear rate leads directly to the Hagen–Poiseuille solution. For a generalized Newtonian fluid, the same momentum balance remains simple, but the constitutive relation between shear stress and shear rate becomes nonlinear.

Capillary flow of non-Newtonian fluids has a long history in rheometry and polymer processing. Classical studies of plastic and non-Newtonian tube flow established the importance of pressure-drop/flow-rate relations, wall effects, and apparent viscosity corrections [3–6]. Modern treatments have further developed analytical, numerical, and semi-analytical descriptions of generalized Newtonian flow in pipes, slits, and narrow channels [7–11]. The present work is connected to this literature, but its purpose is more specific: it identifies a convex-duality structure underlying shear-rate-parametrized capillary-flow solutions.

The essential mathematical difficulty is the nonlinear constitutive relation. In fully developed capillary flow, the shear-stress distribution is determined explicitly by the momentum equation. In a

circular capillary of radius R, for example, the shear stress is a linear function of the radial coordinate. The shear rate is then obtained from the nonlinear flow curve

$$\tau = T(\dot{\gamma}), \tag{1}$$

where T maps the shear-rate magnitude to the shear-stress magnitude. Once $\dot{\gamma}$ is known as a function of position, the velocity follows by integration. Thus, the force balance gives the stress field directly; the remaining task is to invert, or otherwise represent, the nonlinear mapping from shear rate to stress.

In Ref. [12], this difficulty was avoided for the Carreau-Yasuda model by using the shear rate itself as the independent variable. In normalized variables, the stress balance gives the radial coordinate directly as a function of the normalized shear rate, and the velocity profile is then obtained in parametric form. This procedure led to closed expressions involving the Gauss hypergeometric function and provided benchmark solutions for numerical calculations. Related approaches have also used stress, shear rate, velocity, or flow rate as organizing variables for laminar tube-flow problems [9,10].

The present paper shows that this shear-rate parametrization can be placed in a broader Legendre-Fenchel transform framework. The Legendre-Fenchel transform is a standard construction in convex analysis [13,14]. For a monotone flow curve T, we define the dissipation potential

$$\mathcal{D}(\dot{\gamma}) = \int_0^{\dot{\gamma}} T\,(q)\ \mathrm{d}q. \tag{2}$$

The constitutive relation between shear stress and shear rate is then the derivative, or more generally the subgradient, of a convex function. The Legendre-Fenchel transform of $\mathcal{D}$, denoted by $\mathcal{D}^*$, gives a complementary dissipation potential in the stress variable. In this dual representation, the inverse constitutive relation is encoded by the derivative or subgradient of $\mathcal{D}^*$.

This observation leads to a compact capillary-flow solution. Since the stress distribution is known explicitly, the velocity profile can be obtained by evaluating the complementary dissipation potential along that stress distribution. For a circular capillary with no slip at the wall, the result is

$$u(r) = \frac{R}{\tau_w}\left[\mathcal{D}^*(\tau_w) - \mathcal{D}^*\left(\tau_w \frac{r}{R}\right)\right], \tag{3}$$

where τ_w is the wall shear stress. Equation (3) is the central result of this work. It reformulates the capillary-flow problem as a convex-duality problem: the geometry determines the stress distribution, while the rheology enters through the complementary dissipation potential.

Unlike traditional capillary-flow analyses, which typically require a model-specific inversion of the constitutive relation, the present formulation reduces the problem to the construction of a complementary dissipation potential. This provides a unified treatment of smooth shear-thinning models, rheological laws naturally specified in inverse form, and nonsmooth ideal yield-stress limits within the same framework.

This formulation has several advantages. First, it separates the geometric part of the problem from the constitutive part. Second, it avoids the need to write the inverse constitutive relation explicitly when a parametric representation is sufficient. Third, it applies not only to smooth shear-thinning models, such as Carreau-Yasuda and Cross fluids [15–17], but also to regularized and ideal yield-stress models [3,18–20]. In the ideal yield-stress limit, the use of Legendre-Fenchel conjugacy is especially natural because nonsmooth constitutive behavior can be represented through subdifferentials rather than ordinary derivatives.

The analysis in this paper is restricted to steady, fully developed, isothermal, pressure-driven flow of monotone generalized Newtonian fluids in capillary geometries. Within this setting, the aim is to clarify the mathematical structure behind shear-rate-parametrized solutions and to show how several

common rheological models fit into the same framework. The paper is organized as follows. Section 2 recalls the Legendre-Fenchel transform ideas needed in this work and applies them to generalized Newtonian flow curves. Section 3 derives the general capillary-flow solution in terms of the complementary dissipation potential. Section 4 recovers the Carreau-Yasuda solution of Ref. [12]. Section 5 applies the same framework to other generalized Newtonian models, including Cross, Sisko, Ellis, Bingham-Papanastasiou, Herschel-Bulkley-Papanastasiou, and ideal yield-stress fluids. Section 6 summarizes the main findings and discusses the scope and possible extensions of the formulation.

2 Research methodology

Legendre-Fenchel transform and generalized Newtonian fluids

2.1 Definition and basic properties

The Legendre-Fenchel transform is a standard construction in convex analysis [13,14]. It is useful here because it provides a natural way to pass between a constitutive relation written in terms of shear rate and a complementary relation written in terms of stress.

To define the transform, it is convenient to allow functions that may take the value $+\infty$. Thus, instead of considering only real-valued functions, we consider functions of the form

$$f: X \to \mathbb{R} \cup \{+\infty\}. \tag{4}$$

Here X is the space of the variable x, while $f(x)$ is either a finite real number or $+\infty$. The value $+\infty$ is useful because it allows constraints to be included directly in the function. For example, if a variable is only allowed to lie in a set C, one can assign a finite value to $f(x)$ for $x \in C$ and assign $f(x) = +\infty$ for $x \notin C$. Points outside the allowed set are then automatically excluded from minimization problems.

In the present work, most calculations involve scalar variables such as shear-rate magnitude and shear-stress magnitude. Nevertheless, the extended-real notation is useful because it also covers constrained and nonsmooth cases, including ideal yield-stress fluids.

The Legendre-Fenchel transform of f, also called the convex conjugate of f, is defined by

$$f^*(p) = \sup_{x \in X} \{\langle p, x\rangle - f(x)\}. \tag{5}$$

Here x is the original variable, p is the dual variable, and $\langle p, x\rangle$ denotes the duality pairing. In finite-dimensional Euclidean space, this pairing is simply the dot product.

Equation (5) should be compared with the classical Legendre transform. If f is differentiable and strictly convex, one introduces the dual variable by

$$p = f'(x), \tag{6}$$

solves this relation for $x = x(p)$, and writes

$$f^*(p) = p\,x(p) - f(x(p)). \tag{7}$$

Thus, in the differentiable and strictly convex case, the Legendre-Fenchel transform reduces to the usual Legendre transform. The advantage of Eq. (5) is that it does not require differentiability or strict convexity. It is therefore the appropriate transform for nonsmooth functions, constrained functions, and yield-stress limits.

A useful way to understand the transform is through supporting lines. For a convex function, the affine function

$$x \mapsto \langle p, x\rangle - f^*(p) \tag{8}$$

lies below $f(x)$ for all x. The number $f^*(p)$ therefore determines the intercept of the best affine lower support with slope p. This is the geometric reason why the Legendre-Fenchel transform is naturally connected with convexity.

The fundamental inequality associated with Eq. (5) is the Fenchel-Young inequality,

$$f(x) + f^*(p) \geq \langle p, x \rangle. \tag{9}$$

Equality holds precisely when p is a subgradient of f at x:

$$p \in \partial f(x). \tag{10}$$

Equivalently,

$$x \in \partial f^*(p). \tag{11}$$

If both functions are differentiable, these conditions reduce to

$$p = f'(x), \qquad x = (f^*)'(p). \tag{12}$$

Thus, the derivative of the transformed function gives the inverse of the derivative of the original function. This inverse relation is the key mathematical mechanism used in the capillary-flow problem below. Two comments are useful for the present application. First, the Legendre-Fenchel transform is always convex as a function of the dual variable. Second, the transform is not automatically self-inverse. The double transform recovers the original function when the original function is closed and convex; otherwise, it gives the closed convex envelope. In the present work, the relevant functions are dissipation potentials generated by monotone flow curves, so convexity is built into the formulation.

2.2 Dissipation and complementary dissipation potentials

We now apply the Legendre-Fenchel transform to generalized Newtonian fluids. Let $\dot{\gamma} \geq 0$ denote the shear-rate magnitude. A generalized Newtonian fluid is described by a scalar flow curve

$$\tau = T(\dot{\gamma}), \tag{13}$$

where $\tau \geq 0$ is the shear-stress magnitude. For a viscosity model $\eta(\dot{\gamma})$, the flow curve is

$$T(\dot{\gamma}) = \eta(\dot{\gamma})\dot{\gamma}. \tag{14}$$

We assume throughout that the flow curve is monotone nondecreasing:

$$T'(\dot{\gamma}) \geq 0. \tag{15}$$

For differentiable viscosity functions, this condition is equivalent to

$$\eta(\dot{\gamma}) + \dot{\gamma}\eta'(\dot{\gamma}) \geq 0. \tag{16}$$

Physically, monotonicity means that increasing the shear stress should not reduce the shear rate. Mathematically, it ensures that the flow curve can be generated by a convex potential. All model-specific formulas below are understood to hold for parameter ranges where this monotonicity condition is satisfied. Define the dissipation potential by

$$\mathcal{D}(\dot{\gamma}) = \int_0^{\dot{\gamma}} T\,(q)\ \mathrm{d}q. \tag{17}$$

If T is differentiable, then

$$\mathcal{D}''(\dot{\gamma}) = T'(\dot{\gamma}) \geq 0, \tag{18}$$

so $\mathcal{D}$ is convex. The constitutive relation can therefore be written in the subgradient form

$$\tau \in \partial\mathcal{D}(\dot{\gamma}). \tag{19}$$

For smooth flow curves, this reduces to

$$\tau = \mathcal{D}'(\dot{\gamma}) = T(\dot{\gamma}). \tag{20}$$

The Legendre-Fenchel transform of $\mathcal{D}$ is

$$\mathcal{D}^*(\tau) = \sup_{\dot{\gamma}\geq 0}[\tau\dot{\gamma} - \mathcal{D}(\dot{\gamma})]. \tag{21}$$

We call $\mathcal{D}^*$ the complementary dissipation potential. The original potential $\mathcal{D}$ is written in terms of the shear-rate variable, while the complementary potential $\mathcal{D}^*$ is written in terms of the stress variable.

The inverse constitutive relation is obtained from the subgradient of $\mathcal{D}^*$:

$$\dot{\gamma} \in \partial\mathcal{D}^*(\tau). \tag{22}$$

For smooth strictly monotone flow curves, this becomes

$$\dot{\gamma} = (\mathcal{D}^*)'(\tau). \tag{23}$$

Thus, $\mathcal{D}$ gives stress as a function of shear rate, while $\mathcal{D}^*$ gives shear rate as a function of stress. In nonsmooth cases, such as ideal yield-stress limits, the derivative relation is replaced by the corresponding subgradient inclusion.

This is the point at which the Legendre-Fenchel transform becomes useful for capillary flow. The momentum equation gives the stress distribution directly, whereas the velocity gradient is determined by the shear rate. A formulation in which the shear rate is obtained from the stress variable is therefore well suited to the capillary-flow problem. The complementary dissipation potential provides this inverse constitutive information without requiring the inverse flow curve to be written explicitly in elementary form.

In the following section, this observation is used to obtain the velocity profile directly from $\mathcal{D}^*$.

3 Results of the study

Legendre-Fenchel solution of capillary flow

3.1 Circular capillary

Consider steady, fully developed, pressure-driven flow in a straight circular capillary of radius R. Let

$$G = \frac{P_0 - P_L}{L} > 0 \tag{24}$$

be the pressure drop per unit length. The axial velocity is

$$u = u(r), \tag{25}$$

and the shear-rate magnitude is

$$\dot{\gamma}(r) = -\frac{\mathrm{d}u}{\mathrm{d}r} \geq 0. \tag{26}$$

The minus sign appears because the axial velocity decreases from the centerline to the wall.

The wall shear stress is

$$\tau_w = \frac{GR}{2}. \tag{27}$$

Momentum balance gives the linear stress distribution

$$\tau(r) = \tau_w \frac{r}{R}. \tag{28}$$

This relation is independent of the constitutive model. The rheological model enters only through the relation between τ and $\dot{\gamma}$.

The wall shear rate $\dot{\gamma}_w$ is defined by

$$T(\dot{\gamma}_w) = \tau_w. \tag{29}$$

For many models, this scalar equation must be solved numerically. For some models, such as power-law fluids and Bingham-Papanastasiou fluids, it can be solved explicitly. Throughout this section, we assume a monotone generalized Newtonian flow curve so that the stress–shear-rate relation can be represented by a convex dissipation potential.

3.2 Velocity profile from the complementary potential

Let $\mathcal{D}(\dot{\gamma})$ be the local dissipation potential and let $\mathcal{D}^*(\tau)$ be its Legendre-Fenchel transform,

$$\mathcal{D}^*(\tau) = \sup_{\dot{\gamma}\geq 0}[\tau\dot{\gamma} - \mathcal{D}(\dot{\gamma})]. \tag{30}$$

For a smooth strictly monotone generalized Newtonian fluid, the complementary constitutive relation is

$$\dot{\gamma} = \frac{\mathrm{d}\mathcal{D}^*}{\mathrm{d}\tau}. \tag{31}$$

In nonsmooth cases, this derivative relation should be replaced by the corresponding subgradient inclusion $\dot{\gamma} \in \partial\mathcal{D}^*(\tau)$.

Using Eq. (28), the velocity gradient can be written as

$$-\frac{\mathrm{d}u}{\mathrm{d}r} = \frac{\mathrm{d}\mathcal{D}^*}{\mathrm{d}\tau}\left(\tau_w \frac{r}{R}\right) \tag{32}$$

in the smooth case. With the no-slip condition $u(R) = 0$, the velocity at radius r is obtained by integrating from the wall to the interior:

$$u(r) = \int_r^R \dot{\gamma}\,(s)\ \mathrm{d}s. \tag{33}$$

Since

$$\tau = \tau_w \frac{s}{R}, \qquad \mathrm{d}s = \frac{R}{\tau_w}\ \mathrm{d}\tau, \tag{34}$$

Eq. (33) becomes

$$u(r) = \frac{R}{\tau_w}\int_{\tau_w r/R}^{\tau_w} \frac{\mathrm{d}\mathcal{D}^*}{\mathrm{d}\tau}(\tau)\ \mathrm{d}\tau. \tag{35}$$

Therefore,

$$u(r) = \frac{R}{\tau_w}\left[\mathcal{D}^*(\tau_w) - \mathcal{D}^*\left(\tau_w \frac{r}{R}\right)\right]. \tag{36}$$

Equation (36) is the central formula of this work. It states that the velocity profile is generated by the difference between the complementary dissipation potential evaluated at the wall stress and at the local stress. The geometry enters through the known stress distribution in Eq. (28), while the rheology

enters through the complementary potential $\mathcal{D}^*(\tau)$. The formula is derived here for no-slip circular capillary flow, but the same duality structure will also appear in the dimensionless and planar formulations below.

3.3 Dimensionless formulation

It is useful to rewrite the solution in dimensionless form. Let

$$s = \frac{\dot{\gamma}}{\dot{\gamma}_w}, \qquad \rho = \frac{r}{R} = \frac{\tau}{\tau_w}, \tag{37}$$

where $\dot{\gamma}_w$ is the wall shear rate and τ_w is the wall shear stress. Define the normalized flow curve

$$\Psi(s) = \frac{T(\dot{\gamma}_w s)}{\tau_w}. \tag{38}$$

Because $T(\dot{\gamma}_w) = \tau_w$, one has

$$\Psi(1) = 1. \tag{39}$$

The stress distribution gives

$$\rho = \Psi(s). \tag{40}$$

Thus, the dimensionless radial coordinate ρ is the normalized stress corresponding to the dimensionless shear rate s. Under the monotonicity assumption on T, the function $\Psi(s)$ is monotone nondecreasing and can be used as a parametric representation of the radial coordinate.

The normalized dissipation potential is

$$\Phi(s) = \frac{\mathcal{D}(\dot{\gamma}_w s)}{\tau_w \dot{\gamma}_w}. \tag{41}$$

Using Eq. (17), this can be written as

$$\Phi(s) = \int_0^s \Psi(\xi)\, \mathrm{d}\xi. \tag{42}$$

Therefore,

$$\Phi'(s) = \Psi(s) = \rho. \tag{43}$$

This shows that ρ is the conjugate variable to the normalized shear rate s.

The corresponding normalized complementary potential is

$$\Phi^*(\rho) = \frac{\mathcal{D}^*(\tau_w \rho)}{\tau_w \dot{\gamma}_w}. \tag{44}$$

Equivalently,

$$\Phi^*(\rho) = \sup_{s \geq 0} [\rho s - \Phi(s)]. \tag{45}$$

The dimensionless velocity is defined as

$$\tilde{u}(\rho) = \frac{u(r)}{R\dot{\gamma}_w}. \tag{46}$$

Using Eq. (36), one obtains

$$\tilde{u}(\rho) = \Phi^*(1) - \Phi^*(\rho). \tag{47}$$

If $\rho = \Psi(s)$, then the Fenchel equality gives

$$\Phi^*(\rho) = s\rho - \Phi(s) = s\Psi(s) - \Phi(s). \tag{48}$$

At the wall, $s = 1$ and $\rho = 1$. Since $\Psi(1) = 1$,

$$\Phi^*(1) = 1 - \Phi(1). \tag{49}$$

Therefore, the velocity profile can also be written in shear-rate-parametrized form as

$$\tilde{u}(s) = 1 - \Phi(1) - s\Psi(s) + \Phi(s). \tag{50}$$

Using Eq. (42), this becomes

$$\tilde{u}(s) = 1 - s\Psi(s) - \int_s^1 \Psi(\xi)\ \mathrm{d}\xi. \tag{51}$$

Equation (51) is the general shear-rate-parametrized form of the velocity profile. It is equivalent to the construction used in Ref. [12], where $\Psi(s)$ was written as the dimensionless radial coordinate. The auxiliary function introduced there is the tail integral

$$w(s) = \int_s^1 \Psi(\xi)\ \mathrm{d}\xi. \tag{52}$$

Thus, the earlier shear-rate-parametrized solution can be interpreted as a Legendre-Fenchel transform calculation in normalized variables.

3.4 Average velocity

The dimensionless average velocity for circular capillary flow is obtained by area averaging the dimensionless velocity profile:

$$\bar{\tilde{u}} = 2\int_0^1 \tilde{u}(\rho)\rho\ \mathrm{d}\rho = \int_0^1 \tilde{u}(\rho)\ \mathrm{d}(\rho^2). \tag{53}$$

Using integration by parts, together with the no-slip condition $\tilde{u}(1) = 0$, gives

$$\bar{\tilde{u}} = -\int_0^1 \rho^2\ \mathrm{d}\tilde{u}. \tag{54}$$

From Eq. (51),

$$\frac{\mathrm{d}\tilde{u}}{\mathrm{d}\rho} = -s. \tag{55}$$

Therefore,

$$\bar{\tilde{u}} = \int_0^1 \rho^2 s\ \mathrm{d}\rho. \tag{56}$$

Since $\rho = \Psi(s)$, a second integration by parts gives

$$\bar{\tilde{u}} = \frac{1}{3}\left[1 - \int_0^1 \rho(s)^3\ \mathrm{d}s\right]. \tag{57}$$

Thus, for circular capillary flow,

$$\bar{\tilde{u}} = \frac{1}{3}\left[1 - \int_0^1 \Psi(s)^3\ \mathrm{d}s\right]. \tag{58}$$

The same calculation also applies to planar Poiseuille flow, with the area weighting replaced by the one-dimensional average over half of the channel. Using symmetry about the center plane, one has

$$\bar{\tilde{u}} = \int_0^1 \tilde{u}(\rho)\ \mathrm{d}\rho. \tag{59}$$

Integration by parts gives

$$\bar{\tilde{u}} = \int_0^1 \rho s\ \mathrm{d}\rho = \frac{1}{2}\left[1 - \int_0^1 \rho(s)^2\ \mathrm{d}s\right]. \tag{60}$$

Hence, for planar Poiseuille flow,

$$\bar{\tilde{u}} = \frac{1}{2}\left[1 - \int_0^1 \Psi(s)^2\ \mathrm{d}s\right]. \tag{61}$$

Both results can be written in a unified form by introducing the geometric index

$$d = \begin{cases} 1, & \text{planar Poiseuille flow,} \\ 2, & \text{circular capillary flow.} \end{cases} \tag{62}$$

The dimensionless average velocity is then

$$\bar{\tilde{u}} = \frac{1}{d+1}\left[1 - \int_0^1 \Psi\,(s)^{d+1}\ \mathrm{d}s\right]. \tag{63}$$

For circular capillary flow, the dimensional volumetric flow rate is

$$Q = \pi R^2 \bar{u} = \pi R^3 \dot{\gamma}_w \bar{\tilde{u}}. \tag{64}$$

Therefore,

$$Q = \frac{\pi R^3 \dot{\gamma}_w}{3}\left[1 - \int_0^1 \Psi\,(s)^3\ \mathrm{d}s\right]. \tag{65}$$

Together with

$$\tau_w = \frac{GR}{2}, \qquad \Psi(s) = \frac{T(\dot{\gamma}_w s)}{\tau_w}, \qquad \Psi(1) = 1, \tag{66}$$

Eq. (65) gives the flow-rate–pressure-drop relation once the rheological model $T(\dot{\gamma})$ is specified.

4 Recovery of the Carreau-Yasuda result

We now show how the general Legendre-Fenchel formulation recovers the Carreau-Yasuda capillary-flow result. The Carreau-Yasuda viscosity model is

$$\eta(\dot{\gamma}) = \eta_\infty + (\eta_0 - \eta_\infty)[1 + (\lambda\dot{\gamma})^a]^{(n-1)/a}, \tag{67}$$

where η_0 is the zero-shear viscosity, η_∞ is the infinite-shear viscosity, λ is a characteristic time scale, a is the Yasuda transition parameter, and n is the power-law index [15,16]. The corresponding flow curve is

$$T(\dot{\gamma}) = \eta(\dot{\gamma})\dot{\gamma}. \tag{68}$$

The following derivation assumes parameter ranges and shear-rate intervals for which this flow curve is monotone nondecreasing, so that the associated dissipation potential is convex.

Let $\dot{\gamma}_w$ be the wall shear rate. We define

$$\tilde{\lambda} = \lambda\dot{\gamma}_w, \qquad \eta_w = \frac{\tau_w}{\dot{\gamma}_w} = \eta(\dot{\gamma}_w). \tag{69}$$

The normalized viscosities are

$$\tilde{\eta}_0 = \frac{\eta_0}{\eta_w}, \qquad \tilde{\eta}_\infty = \frac{\eta_\infty}{\eta_w}, \qquad \Delta = \tilde{\eta}_0 - \tilde{\eta}_\infty. \tag{70}$$

Therefore,

$$\tilde{\eta}(s) = \frac{\eta(\dot{\gamma}_w s)}{\eta_w} = \tilde{\eta}_\infty + \Delta\left[1 + (\tilde{\lambda}s)^a\right]^{(n-1)/a}. \tag{71}$$

The wall normalization gives

$$\tilde{\eta}(1) = 1, \qquad \tilde{\eta}_\infty + \Delta\left[1 + \tilde{\lambda}^a\right]^{(n-1)/a} = 1. \tag{72}$$

The normalized flow curve is

$$\Psi(s) = \frac{T(\dot{\gamma}_w s)}{\tau_w} = s\tilde{\eta}(s). \tag{73}$$

Thus,

$$\Psi(s) = s\left\{\tilde{\eta}_\infty + \Delta\left[1 + (\tilde{\lambda}s)^a\right]^{(n-1)/a}\right\}. \tag{74}$$

At the wall, $s = 1$ and $\rho = 1$, so that

$$\Psi(1) = 1. \tag{75}$$

The normalized dissipation potential is

$$\Phi(s) = \int_0^s \Psi(\xi)\ \mathrm{d}\xi. \tag{76}$$

For the Carreau-Yasuda model, this integral contains terms of the form

$$\int_0^s \xi\left[1 + (\tilde{\lambda}\xi)^a\right]^{(n-1)/a}\mathrm{d}\xi. \tag{77}$$

This integral is generally not elementary. It can be expressed in terms of the Gauss hypergeometric function, denoted by ${}_2F_1$ [21]. We use the standard integral identity

$$\int_0^s x^{p-1}(1+\beta x^a)^q\,\mathrm{d}x = \frac{s^p}{p}\,{}_2F_1\left(\frac{p}{a}, -q; 1+\frac{p}{a}; -\beta s^a\right), \tag{78}$$

provided the integral is finite, with analytic continuation understood where applicable [21].
Applying Eq. (78) with

$$p = 2, \qquad \beta = \tilde{\lambda}^a, \qquad q = \frac{n-1}{a}, \tag{79}$$

gives

$$\Phi(s) = \frac{1}{2}\tilde{\eta}_\infty s^2 + \frac{1}{2}\Delta s^2\,{}_2F_1\left(\frac{2}{a}, \frac{1-n}{a}; \frac{a+2}{a}; -(\tilde{\lambda}s)^a\right). \tag{80}$$

The velocity profile follows from the general dimensionless formula. Since $\rho = \Psi(s)$, one has

$$\rho(s) = \Psi(s). \tag{81}$$

The shear-rate-parametrized dimensionless velocity is

$$\tilde{u}(s) = 1 - s\Psi(s) - \int_s^1 \Psi(\xi)\ \mathrm{d}\xi. \tag{82}$$

Using $\Phi(s) = \int_0^s \Psi(\xi)\ \mathrm{d}\xi$, this becomes

$$\tilde{u}(s) = 1 - s\Psi(s) - \Phi(1) + \Phi(s). \tag{83}$$

Therefore, the Carreau-Yasuda velocity profile can be written in the closed parametric form

$$\rho(s) = \Psi(s), \qquad \tilde{u}(s) = 1 - s\Psi(s) - \Phi(1) + \Phi(s), \tag{84}$$

where $\Psi(s)$ is given by Eq. (74) and $\Phi(s)$ is given by Eq. (80). This is the Legendre-Fenchel reformulation of the shear-rate-parametrized velocity solution of Ref. [12].

The average velocity follows from the unified formula derived above. Define, for integer $k \geq 1$, the moment

$$M_k = \int_0^1 \Psi\,(s)^k \,\mathrm{d}s. \tag{85}$$

For the Carreau-Yasuda model,

$$\Psi(s)^k = s^k \left\{\tilde{\eta}_\infty + \Delta\left[1 + (\tilde{\lambda}s)^a\right]^{(n-1)/a}\right\}^k. \tag{86}$$

Expanding the power gives

$$M_k = \sum_{j=0}^{k} \binom{k}{j} \tilde{\eta}_\infty^{k-j} \Delta^j \int_0^1 s^k \left[1 + (\tilde{\lambda}s)^a\right]^{j(n-1)/a} \mathrm{d}s. \tag{87}$$

Using Eq. (78) with

$$p = k + 1, \qquad \beta = \tilde{\lambda}^a, \qquad q = \frac{j(n-1)}{a}, \tag{88}$$

one obtains

$$\begin{aligned} M_k = {} & \frac{1}{k+1} \sum_{j=0}^{k} \binom{k}{j} \tilde{\eta}_\infty^{k-j} \Delta^j \\ & \times {}_2F_1\left(\frac{k+1}{a}, \frac{j(1-n)}{a}; \frac{k+1+a}{a}; -\tilde{\lambda}^a\right). \end{aligned} \tag{89}$$

The dimensionless average velocity is therefore

$$\bar{\tilde{u}} = \frac{1}{d+1}\left[1 - M_{d+1}\right]. \tag{90}$$

For circular capillary flow, $d = 2$ and $k = 3$; for planar Poiseuille flow, $d = 1$ and $k = 2$. Thus,

$$\bar{\tilde{u}}_{\mathrm{pipe}} = \frac{1}{3}\left[1 - M_3\right], \qquad \bar{\tilde{u}}_{\mathrm{plane}} = \frac{1}{2}\left[1 - M_2\right]. \tag{91}$$

For circular capillary flow, the dimensional volumetric flow rate is

$$Q = \pi R^2 \bar{u} = \pi R^3 \dot{\gamma}_w \bar{\tilde{u}}_{\mathrm{pipe}}. \tag{92}$$

Thus,

$$Q = \frac{\pi R^3 \dot{\gamma}_w}{3}\left[1 - M_3\right]. \tag{93}$$

The pressure-gradient dependence enters through

$$\tau_w = \frac{GR}{2} = \eta(\dot{\gamma}_w)\dot{\gamma}_w. \tag{94}$$

For a prescribed pressure gradient G, Eq. (94) determines $\dot{\gamma}_w$, and Eq. (93) gives the corresponding volumetric flow rate.

Equations (84) and (90) recover the pipe-flow and plane-flow Carreau-Yasuda results of Ref. [12]. The present derivation also clarifies why hypergeometric functions appear: they are consequences of analytically evaluating the normalized dissipation potential and the moments of the normalized flow curve, rather than essential ingredients of the Legendre-Fenchel framework. The underlying conjugacy structure,

$$s \leftrightarrow \rho = \Psi(s), \qquad \tilde{u}(\rho) = \Phi^*(1) - \Phi^*(\rho), \tag{95}$$

is independent of the particular viscosity model. The Carreau-Yasuda model is therefore one explicit example within a broader Legendre-Fenchel framework for pressure-driven capillary flow.

5 Other generalized Newtonian models

The preceding sections show that, once the normalized flow curve $\Psi(s) = T(\dot{\gamma}_w s)/\tau_w$ is known, the velocity profile and the average velocity follow from Eqs. (51) and (63). This section records several representative examples. The purpose is not to rederive each tube-flow solution in detail, but to show how different rheological models enter the same Legendre-Fenchel structure through their normalized flow curves, dissipation potentials, and moment integrals.

For elementary models, the framework recovers the classical results directly. For empirical shear-thinning models, such as Sisko, Cross, and Carreau-Yasuda fluids, the same construction gives parametric velocity profiles and average-velocity moments. For models naturally written in inverse form, such as the Ellis model, the complementary potential can be obtained directly. For regularized and ideal yield-stress fluids, the formulation is especially useful because it also handles nonsmooth limits through subgradients.

5.1 Newtonian and power-law limits

For a Newtonian fluid, $T(\dot{\gamma}) = \eta\dot{\gamma}$, and the complementary potential is $\mathcal{D}^*(\tau) = \tau^2/(2\eta)$. Substitution into Eq. (36) gives the classical Hagen–Poiseuille profile,

$$u(r) = \frac{G}{4\eta}(R^2 - r^2). \tag{96}$$

In dimensionless form, $\Psi(s) = s$, and $\bar{\tilde{u}} = 1/4$ for circular capillary flow.

For a power-law fluid,

$$T(\dot{\gamma}) = K\dot{\gamma}^{\,n}, \qquad n > 0, \tag{97}$$

the normalized flow curve is $\Psi(s) = s^n$. Therefore, $\rho = s^n$, and the dimensionless velocity profile is

$$\tilde{u}(\rho) = \frac{n}{n+1}\left[1 - \rho^{(n+1)/n}\right]. \tag{98}$$

The corresponding average velocities are

$$\bar{\tilde{u}} = \frac{n}{3n+1}, \text{circular capillary flow,} \tag{99}$$

$$\bar{\tilde{u}} = \frac{n}{2n+1}, \text{planar Poiseuille flow.} \tag{100}$$

These elementary limits serve as consistency checks on the general formula.

5.2 Sisko fluid

The Sisko model combines a Newtonian high-shear contribution with a power-law contribution:

$$T(\dot{\gamma}) = \eta_\infty\dot{\gamma} + K\dot{\gamma}^{\,n}. \tag{101}$$

Define

$$\mu = \frac{\eta_\infty\dot{\gamma}_w}{\tau_w}, \qquad \kappa = \frac{K\dot{\gamma}_w^n}{\tau_w}. \tag{102}$$

The wall condition gives $\mu + \kappa = 1$. The normalized flow curve and normalized dissipation potential are

$$\Psi(s) = \mu s + \kappa s^n, \qquad \Phi(s) = \frac{\mu}{2}s^2 + \frac{\kappa}{n+1}s^{n+1}. \tag{103}$$

Therefore,

$$\tilde{u}(s) = \frac{\mu}{2}(1 - s^2) + \frac{n\kappa}{n+1}(1 - s^{n+1}). \tag{104}$$

For integer k, the moment integral is elementary:

$$M_k = \sum_{j=0}^{k} \binom{k}{j} \mu^{k-j} \kappa^j \frac{1}{k-j+nj+1}. \tag{105}$$

The average velocity follows from Eq. (63) with M_{d+1}.

5.3 Cross fluid

A common Cross-type viscosity model is

$$\eta(\dot{\gamma}) = \eta_\infty + \frac{\eta_0 - \eta_\infty}{1 + (\lambda\dot{\gamma})^m}. \tag{106}$$

This model is widely used for pseudoplastic and polymeric fluids [17]. As pointed out in Ref. [12], it can also be viewed as a special case of the Carreau-Yasuda form. Comparing Eq. (106) with Eq. (67), the Cross model is recovered by the formal parameter choice

$$a = m, \qquad n = 1 - m. \tag{107}$$

This mapping gives $(n-1)/a = -1$, and therefore reduces the Carreau-Yasuda factor to $[1 + (\lambda\dot{\gamma})^m]^{-1}$. It is a formal mapping at the level of the viscosity function and should not be confused with independent empirical parameter conventions used in some Cross-model literature.

Let

$$\tilde{\lambda} = \lambda\dot{\gamma}_w, \qquad \eta_w = \frac{\tau_w}{\dot{\gamma}_w}, \tag{108}$$

and define

$$\tilde{\eta}_\infty = \frac{\eta_\infty}{\eta_w}, \qquad \Delta = \frac{\eta_0 - \eta_\infty}{\eta_w}. \tag{109}$$

The normalized flow curve is

$$\Psi(s) = s\left[\tilde{\eta}_\infty + \frac{\Delta}{1 + (\tilde{\lambda}s)^m}\right], \tag{110}$$

with wall normalization

$$\tilde{\eta}_\infty + \frac{\Delta}{1 + \tilde{\lambda}^m} = 1. \tag{111}$$

Using the hypergeometric integral identity introduced in Sec. 4, the normalized dissipation potential is

$$\begin{aligned} \Phi(s) = {} & \frac{1}{2}\tilde{\eta}_\infty s^2 \\ & + \frac{1}{2}\Delta s^2 \, {}_2F_1\left(\frac{2}{m}, 1; \frac{m+2}{m}; -(\tilde{\lambda}s)^m\right). \end{aligned} \tag{112}$$

The velocity profile is then given by Eq. (50). For integer k, define $M_k = \int_0^1 \Psi(s)^k \, \mathrm{d}s$. Expanding the integrand and using the same hypergeometric identity gives

$$\begin{aligned} M_k = {} & \frac{1}{k+1}\sum_{j=0}^{k} \binom{k}{j} \tilde{\eta}_\infty^{k-j} \Delta^j \\ & \times {}_2F_1\left(\frac{k+1}{m}, j; \frac{k+1+m}{m}; -\tilde{\lambda}^m\right). \end{aligned} \tag{113}$$

The average velocity follows from Eq. (63). Thus, the Cross model gives the same type of hypergeometric structure as the Carreau-Yasuda model, but with the exponent fixed at -1.

5.4 Ellis fluid

The Ellis model is particularly illustrative from the Legendre-Fenchel viewpoint because it is naturally specified in the dual stress variable. As a result, the complementary potential can be obtained directly without first inverting the constitutive relation.

The Ellis model is often written in inverse-flow-curve form:

$$\dot{\gamma}(\tau) = \frac{\tau}{\eta_0}\left[1 + \left(\frac{\tau}{\tau_{1/2}}\right)^{\alpha-1}\right]. \tag{114}$$

This form is especially natural in the present framework because the complementary potential is obtained directly by integration in the stress variable:

$$\mathcal{D}^*(\tau) = \int_0^\tau \dot{\gamma}\,(q)\ \mathrm{d}q = \frac{\tau^2}{2\eta_0} + \frac{\tau^{\alpha+1}}{(\alpha+1)\eta_0\tau_{1/2}^{\alpha-1}}. \tag{115}$$

Substitution into Eq. (36) gives

$$u(r) = \frac{R\tau_w}{2\eta_0}(1-\rho^2) + \frac{R\tau_w^{\alpha}}{(\alpha+1)\eta_0\tau_{1/2}^{\alpha-1}}(1-\rho^{\alpha+1}), \qquad \rho = \frac{r}{R}. \tag{116}$$

This example highlights an advantage of the conjugacy viewpoint: if a model is naturally specified as $\dot{\gamma}(\tau)$, the complementary potential can be obtained without inverting the constitutive law.

5.5 Bingham–Papanastasiou fluid

Yield-stress fluids provide a useful test case because ideal yield-stress laws are nonsmooth, while Papanastasiou regularization gives a smooth approximation that approaches the ideal law as the regularization parameter increases [18]. For a Bingham–Papanastasiou fluid,

$$T(\dot{\gamma}) = \eta_p\dot{\gamma} + \tau_y\left(1 - e^{-m\dot{\gamma}}\right), \tag{117}$$

where η_p is the plastic viscosity, τ_y is the yield stress, and m is the regularization parameter. The corresponding dissipation potential is

$$\mathcal{D}(\dot{\gamma}) = \frac{1}{2}\eta_p\dot{\gamma}^2 + \tau_y\left[\dot{\gamma} - \frac{1-e^{-m\dot{\gamma}}}{m}\right], \tag{118}$$

which is smooth and strictly convex for positive material parameters.

The wall shear rate satisfies

$$\tau_w = \eta_p\dot{\gamma}_w + \tau_y\left(1 - e^{-m\dot{\gamma}_w}\right). \tag{119}$$

This scalar equation can be solved explicitly using the principal branch W_0 of the Lambert W function, defined by $W(z)e^{W(z)} = z$ [22]. The use of the Lambert W function in Papanastasiou and related non-Newtonian flow problems has been discussed in Refs. [23–25]. For the present wall-shear-rate problem, the result is

$$\dot{\gamma}_w = \frac{\tau_w-\tau_y}{\eta_p} + \frac{1}{m}W_0\left[\frac{m\tau_y}{\eta_p}\exp\left(-\frac{m(\tau_w-\tau_y)}{\eta_p}\right)\right]. \tag{120}$$

For positive material parameters, the argument of W_0 is positive, and the principal real branch gives the physical solution.

Define

$$\beta = \frac{\tau_y}{\tau_w}, \qquad \mu = \frac{\eta_p \dot{\gamma}_w}{\tau_w}, \qquad M = m\dot{\gamma}_w. \tag{121}$$

The wall condition is

$$\mu + \beta(1 - e^{-M}) = 1. \tag{122}$$

The normalized flow curve and normalized dissipation potential are

$$\begin{aligned} \Psi(s) &= \mu s + \beta(1 - e^{-Ms}), \\ \Phi(s) &= \frac{\mu}{2}s^2 + \beta\left[s - \frac{1-e^{-Ms}}{M}\right]. \end{aligned} \tag{123}$$

The parametric velocity profile is

$$\begin{aligned} \tilde{u}(s) = \frac{\mu}{2}(1 - s^2) + \beta(se^{-Ms} - e^{-M}) \\ + \frac{\beta}{M}(e^{-Ms} - e^{-M}). \end{aligned} \tag{124}$$

The map $\rho = \Psi(s)$ can also be inverted explicitly:

$$\begin{aligned} s(\rho) = \frac{\rho-\beta}{\mu} \\ + \frac{1}{M}W_0\left[\frac{M\beta}{\mu}\exp\left(-\frac{M(\rho-\beta)}{\mu}\right)\right]. \end{aligned} \tag{125}$$

For circular capillary flow, the average velocity is $\bar{\tilde{u}} = (1 - M_3)/3$, where

$$M_3 = \int_0^1 [\mu s + \beta(1 - e^{-Ms})]^3 \, \mathrm{d}s. \tag{126}$$

This integral is elementary. A compact closed form is

$$\begin{aligned} M_3 = \beta^3 + \frac{3}{2}\beta^2\mu + \beta\mu^2 + \frac{\mu^3}{4} \\ -3\beta[\beta^2 A_0(M) + 2\beta\mu A_1(M) + \mu^2 A_2(M)] \\ +3\beta^2[\beta A_0(2M) + \mu A_1(2M)] \\ -\beta^3 A_0(3M), \end{aligned} \tag{127}$$

Where

$$\begin{aligned} A_0(x) &= \frac{1-e^{-x}}{x}, \\ A_1(x) &= \frac{1-e^{-x}(1+x)}{x^2}, \\ A_2(x) &= \frac{2-e^{-x}(x^2+2x+2)}{x^3}. \end{aligned} \tag{128}$$

The dimensional flow rate is then

$$Q = \pi R^3 \dot{\gamma}_w \frac{1-M_3}{3}. \tag{129}$$

5.6 Herschel–Bulkley–Papanastasiou fluid

The Herschel–Bulkley–Papanastasiou model generalizes the Bingham–Papanastasiou model by replacing the linear viscous contribution with a power-law contribution:

$$T(\dot{\gamma}) = K\dot{\gamma}^{\,n} + \tau_y\left(1 - e^{-m\dot{\gamma}}\right). \tag{130}$$

For $n = 1$ and $K = \eta_p$, Eq. (130) reduces to Eq. (117). The corresponding dissipation potential is

$$\mathcal{D}(\dot{\gamma}) = \frac{K}{n+1}\dot{\gamma}^{\,n+1} + \tau_y\left[\dot{\gamma} - \frac{1-e^{-m\dot{\gamma}}}{m}\right]. \tag{131}$$

For $K > 0$, $n > 0$, $\tau_y \geq 0$, and $m > 0$, the flow curve is monotone nondecreasing and the dissipation potential is convex on $\dot{\gamma} \geq 0$. When $0 < n < 1$, the derivative of the flow curve is singular as $\dot{\gamma} \to 0^+$, but this singularity is compatible with convexity.

The wall shear rate is determined by

$$\tau_w = K\dot{\gamma}_w^{\,n} + \tau_y\left(1 - e^{-m\dot{\gamma}_w}\right). \tag{132}$$

Unlike the Bingham–Papanastasiou case, this equation generally cannot be reduced to the standard Lambert W form when $n \neq 1$, because it contains both a power-law term and an exponential term. It is nevertheless a scalar monotone equation.

Define

$$\alpha = \frac{K\dot{\gamma}_w^{\,n}}{\tau_w}, \qquad \beta = \frac{\tau_y}{\tau_w}, \qquad M = m\dot{\gamma}_w. \tag{133}$$

The wall condition is

$$\alpha + \beta(1 - e^{-M}) = 1. \tag{134}$$

The normalized flow curve and normalized dissipation potential are

$$\begin{aligned} \Psi(s) &= \alpha s^n + \beta(1 - e^{-Ms}), \\ \Phi(s) &= \frac{\alpha}{n+1}s^{n+1} + \beta\left[s - \frac{1-e^{-Ms}}{M}\right]. \end{aligned} \tag{135}$$

The parametric radial coordinate is $\rho(s) = \Psi(s)$. The corresponding velocity profile is

$$\begin{aligned} \tilde{u}(s) &= \frac{\alpha n}{n+1}(1 - s^{n+1}) \\ &+\beta(se^{-Ms} - e^{-M}) + \frac{\beta}{M}(e^{-Ms} - e^{-M}). \end{aligned} \tag{136}$$

The average-velocity moment is

$$M_k = \int_0^1 [\alpha s^n + \beta(1 - e^{-Ms})]^k \, \mathrm{d}s. \tag{137}$$

For integer k, expansion gives

$$\begin{aligned} M_k &= \sum_{j=0}^{k} \binom{k}{j} \alpha^{\,k-j}\beta^{\,j} \\ &\times \sum_{\ell=0}^{j} (-1)^{\ell}\binom{j}{\ell} E_{n(k-j)}(\ell M), \end{aligned} \tag{138}$$

where

$$E_p(x) = \int_0^1 s^p \, e^{-xs} \, \mathrm{d}s. \tag{139}$$

For $x > 0$,

$$E_p(x) = \frac{\gamma(p+1,x)}{x^{p+1}}, \tag{140}$$

where $\gamma(\cdot,\cdot)$ is the lower incomplete gamma function [21]. For $x = 0$, $E_p(0) = 1/(p+1)$. The dimensionless average velocity follows from Eq. (63). For circular capillary flow,

$$Q = \pi R^3 \dot{\gamma}_w \frac{1-M_3}{3}. \tag{141}$$

The Bingham–Papanastasiou result is recovered by setting $n = 1$ and $K = \eta_p$. Then $\alpha = \mu$, and Eqs. (135) and (136) reduce to Eqs. (123) and (124). In the same limit, Eq. (132) reduces to Eq. (119), and the explicit W_0 solution in Eq. (120) is recovered.

5.7 Ideal yield-stress limit

The Legendre-Fenchel formulation is especially useful in the ideal yield-stress limit. Consider $m \to \infty$ in the Herschel–Bulkley–Papanastasiou model. The regularized flow curve tends to the ideal Herschel–Bulkley law

$$T(\dot{\gamma}) = \tau_y + K\dot{\gamma}^{\,n}, \qquad \dot{\gamma} > 0, \tag{142}$$

a classical yield-stress model [19,20]. At $\dot{\gamma} = 0$, the ideal law is not single-valued: stresses below the yield stress correspond to zero shear rate.

Let $\beta = \tau_y/\tau_w$ and assume $0 < \beta < 1$, so that yielding occurs near the wall. The wall shear rate satisfies

$$K\dot{\gamma}_w^n = \tau_w(1-\beta). \tag{143}$$

In the sheared region,

$$\rho = \Psi(s) = \beta + (1-\beta)s^n, \qquad 0 < s \le 1. \tag{144}$$

Thus, $s \to 0^+$ corresponds to $\rho \to \beta$, and the interval $0 \le \rho \le \beta$ is an unsheared plug.

The normalized ideal dissipation potential is

$$\Phi_{\mathrm{HB}}(s) = \beta s + \frac{1-\beta}{n+1}s^{n+1}, \qquad s \ge 0. \tag{145}$$

For the nonnegative stress interval relevant here, the physically relevant subgradient at $s = 0$ is $0 \le \rho \le \beta$. To verify the corresponding conjugate, we compute

$$\Phi_{HB}{}^*(\rho) = sup_{s \ge 0}\left[\rho s - \beta s - \frac{1-\beta}{n+1}s^{n+1}\right]$$

For $0 \le \rho \le \beta$, the supremum is attained at $s = 0$, giving $\Phi_{HB}{}^*(\rho) = 0$. For $\rho > \beta$, the maximizer satisfies

$$\rho - \beta - (1-\beta)s^n = 0,$$

so that

$$s^* = \left(\frac{\rho-\beta}{1-\beta}\right)^{\frac{1}{n}}.$$

Substituting s^* into the supremum yields

$$\Phi^*_{\mathrm{HB}}(\rho) = \begin{cases} 0, & 0 \le \rho \le \beta, \\ \frac{n}{n+1}\frac{(\rho-\beta)^{(n+1)/n}}{(1-\beta)^{1/n}}, & \beta < \rho \le 1. \end{cases} \tag{146}$$

The dimensionless velocity is therefore

$$\tilde{u}(\rho) = \Phi^*_{\mathrm{HB}}(1) - \Phi^*_{\mathrm{HB}}(\rho). \tag{147}$$

Since $\Phi^*_{\mathrm{HB}}(1) = n(1-\beta)/(n+1)$, the plug velocity is

$$\tilde{u}(\rho) = \frac{n}{n+1}(1-\beta), \qquad 0 \le \rho \le \beta. \tag{148}$$

In the sheared region,

$$\tilde{u}(\rho) = \frac{n}{n+1}(1-\beta) - \frac{n}{n+1}\frac{(\rho-\beta)^{(n+1)/n}}{(1-\beta)^{1/n}}, \qquad \beta < \rho \le 1. \tag{149}$$

The average velocity follows from the same moment formula. For integer k,

$$M_k = \int_0^1 [\beta + (1-\beta)s^n]^k \, \mathrm{d}s = \sum_{j=0}^{k} \binom{k}{j} \beta^{k-j}(1-\beta)^j \frac{1}{nj+1}. \tag{150}$$

Although the integral over s runs only over the sheared branch, the plug contribution is included through the boundary term in the integration-by-parts derivation of the average velocity. Hence,

$$\bar{\tilde{u}} = \frac{1-M_{d+1}}{d+1}. \tag{151}$$

For the ideal Bingham case, $n = 1$. The plug velocity becomes

$$\tilde{u}(\rho) = \frac{1-\beta}{2}, \qquad 0 \le \rho \le \beta, \tag{152}$$

and the sheared-region velocity becomes

$$\tilde{u}(\rho) = \frac{1-\beta}{2} - \frac{(\rho-\beta)^2}{2(1-\beta)}, \qquad \beta < \rho \le 1. \tag{153}$$

Thus, the plug region is not imposed separately. It follows directly from the subgradient of the nonsmooth dissipation potential, or equivalently from the flat region of the complementary potential. Plug formation is therefore the spatial manifestation of the dual variational structure of an ideal yield-stress fluid.

6 Discussion

The present work reformulates pressure-driven capillary flow of generalized Newtonian fluids using Legendre-Fenchel conjugacy. The usual difficulty in such problems is the implicit nonlinear relation between shear stress and shear rate. Although the stress distribution is known from momentum balance, the shear rate must still be determined from the nonlinear flow curve $\tau = T(\dot{\gamma})$. For many rheological models, this inversion cannot be performed in elementary functions. The shear-rate parametrization used in Ref. [12] avoided this difficulty by using the shear rate itself as the independent variable. The present manuscript shows that this parametrization is not merely a useful change of variables. It is a manifestation of the conjugate-variable structure generated by the dissipation potential.

In this interpretation, the capillary-flow problem has three distinct parts. The first is geometric: the stress distribution is fixed by momentum balance, as in Eq. (28). The second is constitutive: the flow curve defines the dissipation potential, as in Eq. (17). The third is dual: the velocity profile is generated by the complementary dissipation potential, as shown in Eq. (36). This decomposition separates the known stress field from the rheological response and explains why the same mathematical structure applies to apparently different viscosity models.

The representative models examined in this work are summarized in Table 1. The table is not intended to be an exhaustive catalog of generalized Newtonian fluids. Rather, it illustrates how the same Legendre-Fenchel construction applies to several common models, including Newtonian, power-law, Sisko, Cross, Carreau–Yasuda, Ellis, Bingham-Papanastasiou, Herschel-Bulkley-Papanastasiou, and ideal Herschel–Bulkley fluids. It also emphasizes that the central object in the

dimensionless formulation is the normalized flow curve $\Psi(s)$. Once $\Psi(s)$ is known, the velocity profile and average velocity follow from Eqs. (51) and (63).

Table 1. Representative model catalog in the Legendre-Fenchel capillary-flow framework. The normalized flow curve $\Psi(s) = T(\dot{\gamma}_w s)/\tau_w$ is the central object. The velocity and average velocity follow from Eqs. (51) and (63), respectively.

Model	*Flow curve or inverse law*	*Normalized flow curve*	*Analytical structure*
Newtonian	$T = \eta\dot{\gamma}$	$\Psi = s$	*elementary*
Power-law	$T = K\dot{\gamma}^{\,n}$	$\Psi = s^n$	*elementary*
Sisko	$T = \eta_\infty\dot{\gamma} + K\dot{\gamma}^{\,n}$	$\Psi = \mu s + \kappa s^n$	*elementary*
Cross	$\eta = \eta_\infty + \frac{\eta_0 - \eta_\infty}{1 + (\lambda\dot{\gamma})^m}$	$\Psi = s\left[\tilde{\eta}_\infty + \frac{\Delta}{1 + (\tilde{\lambda}s)^m}\right]$	*hypergeometric*
Carreau-Yasuda	*Eq. (67)*	$\Psi = s\tilde{\eta}(s)$	*hypergeometric*
Ellis	$\dot{\gamma} = \frac{\tau}{\eta_0}\left[1 + \left(\frac{\tau}{\tau_{1/2}}\right)^{\alpha-1}\right]$	*naturally dual*	*elementary*
Bingham-Papanastasiou	*Eq. (117)*	$\Psi = \mu s + \beta(1 - e^{-Ms})$	*Lambert W*
Herschel-Bulkley-Papanastasiou	*Eq. (130)*	$\Psi = \alpha s^n + \beta(1 - e^{-Ms})$	*incomplete gamma*
Ideal Herschel-Bulkley	$T = \tau_y + K\dot{\gamma}^{\,n}$	$\Psi = \beta + (1-\beta)s^n$ *in sheared region*	*nonsmooth conjugate*

As shown in Table 1, the special functions appearing in the final expressions are not fundamental to the method itself. They arise only when the normalized dissipation potential or the average-velocity moments are evaluated analytically for a particular model. The Carreau-Yasuda and Cross models lead to Gauss hypergeometric functions because their normalized potentials involve integrals of rational-power functions. The Cross model is included as a formal special case of the Carreau-Yasuda model, as discussed in Ref. [12], but it is listed separately because its simplified form is widely used. The Bingham–Papanastasiou model leads to exponentials and the Lambert W function because of the exponential regularization term. The Herschel-Bulkley-Papanastasiou model generally does not reduce to the standard Lambert W form for the wall shear rate when $n \neq 1$, but its average-velocity moments can be written using incomplete gamma functions. In contrast, Newtonian, power-law, and Sisko fluids lead to elementary formulas. Ellis fluids are naturally handled in the dual formulation because they are commonly specified by the inverse relation $\dot{\gamma} = \dot{\gamma}(\tau)$.

The ideal yield-stress limit highlights the main conceptual advantage of the Legendre-Fenchel viewpoint. Ideal Bingham and Herschel-Bulkley laws are nonsmooth at zero shear rate. In the present formulation, this nonsmoothness is not an obstacle. It is represented by the subdifferential of the dissipation potential. The corresponding complementary potential contains a flat interval, and this flat interval generates the plug region in capillary flow. Thus, the plug is not introduced as an additional geometric assumption. It follows directly from the dual constitutive structure.

The scope of this study is deliberately limited. The analysis is restricted to steady, fully developed, isothermal, pressure-driven flow of monotone generalized Newtonian fluids in geometries where the stress distribution is known. Within this setting, the Legendre-Fenchel framework gives a compact analytical structure for the velocity profile and the flow-rate relation. The purpose is not to replace numerical simulation or to cover all rheological effects, but to clarify the mathematical structure underlying a class of capillary-flow solutions.

Several extensions follow naturally. Wall slip can be incorporated by modifying the boundary condition, in which case the no-slip solution derived here becomes the bulk contribution and slip adds a stress-dependent velocity offset. Other geometries may also be treated when the stress distribution is available. Planar Poiseuille flow already fits within the same dimensionless formulation. Annular flow is more delicate because the stress field may change sign and the location of zero shear must be determined, but the same principle should apply once the appropriate stress distribution is specified.

The present formulation may also be useful for inverse capillary rheometry. Given pressure-drop/flow-rate data, one can fit the parameters of $T(\dot{\gamma})$ by matching the predicted $Q(G)$ relation. The separation between the wall shear-rate calculation, the normalized flow curve, and the velocity integration may help organize such inverse calculations.

7 Conclusion

In summary, the shear-rate parametrization used previously for the Carreau-Yasuda model is shown here to be a manifestation of Legendre-Fenchel conjugacy. The central result is that the velocity profile of a monotone generalized Newtonian fluid in capillary flow is generated by the complementary dissipation potential evaluated along the known stress distribution. This framework recovers the Carreau-Yasuda hypergeometric solution, explains the structure of several other generalized Newtonian models, and provides a natural treatment of ideal yield-stress limits through subdifferentials and flat regions of the dual potential.

Acknowledgments

This research is funded by the Science Committee of the Ministry of Science and Higher Education of the Republic of Kazakhstan (Grant No. AP26197943).

References

[1] Bird, R.B., Stewart, W.E., Lightfoot, E.N. Transport Phenomena. Revised 2nd ed. New York: John Wiley & Sons, 2006.

[2] Chhabra, R.P., Richardson, J.F. Non-Newtonian Flow and Applied Rheology: Engineering Applications. 2nd ed. Amsterdam: Butterworth-Heinemann, 2008.

[3] Buckingham, E. On plastic flow through capillary tubes // Proceedings of the American Society for Testing Materials. 1921. P. 1154–1156.

[4] Mooney, M. Explicit formulas for slip and fluidity // Journal of Rheology. 1931. Vol. 2, No. 2. P. 210–222. https://doi.org/10.1122/1.2116364.

[5] Meter, D.M., Bird, R.B. Tube flow of non-Newtonian polymer solutions: Part I. Laminar flow and rheological models // AIChE Journal. 1964. Vol. 10, No. 6. P. 878–881. https://doi.org/10.1002/aic.690100619.

[6] Metzner, A.B., Reed, J.C. Flow of non-Newtonian fluids–correlation of the laminar, transition, and turbulent-flow regions // AIChE Journal. 1955. Vol. 1, No. 4. P. 434–440. https://doi.org/10.1002/aic.690010409.

[7] Sochi, T. Using the Euler–Lagrange variational principle to obtain flow relations for generalized Newtonian fluids // Rheologica Acta. 2014. Vol. 53. P. 15–22. https://doi.org/10.1007/s00397-013-0741-3.

[8] Sochi, T. Analytical solutions for the flow of Carreau and Cross fluids in circular pipes and thin slits // Rheologica Acta. 2015. Vol. 54. P. 745–756. https://doi.org/10.1007/s00397-015-0863-x.

[9] Kim, S.K. Flow-rate based method for velocity of fully developed laminar flow in tubes // Journal of Rheology. 2018. Vol. 62, No. 6. P. 1397–1407. https://doi.org/10.1122/1.5041958.

[10] Tozzi, E., Hartt, W. Non-Newtonian laminar flow in pipes using radius, stress, shear rate or velocity as the independent variable // Physics of Fluids. 2021. Vol. 33. Article 103104. https://doi.org/10.1063/5.0067993.

[11] Boyko, E., Stone, H.A. Flow rate–pressure drop relation for shear-thinning fluids in narrow channels: approximate solutions and comparison with experiments // Journal of Fluid Mechanics. 2021. Vol. 923. Article A23. https://doi.org/10.1017/jfm.2021.621.

[12] Wang, Y. Steady isothermal flow of a Carreau-Yasuda model fluid in a straight circular tube // Journal of Non-Newtonian Fluid Mechanics. 2022. Vol. 310. Article 104937. https://doi.org/10.1016/j.jnnfm.2022.104937.

[13] Rockafellar, R.T. Convex Analysis. Princeton: Princeton University Press, 1970.

[14] Boyd, S., Vandenberghe, L. Convex Optimization. Cambridge: Cambridge University Press, 2004.

[15] Carreau, P.J. Rheological equations from molecular network theories // Transactions of the Society of Rheology. 1972. Vol. 16, No. 1. P. 99–127. https://doi.org/10.1122/1.549276.

[16] Yasuda, K., Armstrong, R.C., Cohen, R.E. Shear flow properties of concentrated solutions of linear and star branched polystyrenes // Rheologica Acta. 1981. Vol. 20. P. 163–178. https://doi.org/10.1007/BF01513059.

[17] Cross, M.M. Rheology of non-Newtonian fluids: a new flow equation for pseudoplastic systems // Journal of Colloid Science. 1965. Vol. 20, No. 5. P. 417-437. https://doi.org/10.1016/0095-8522(65)90022-X.

[18] Papanastasiou, T.C. Flows of materials with yield // Journal of Rheology. 1987. Vol. 31, No. 5. P. 385–404. https://doi.org/10.1122/1.549926.

[19] Herschel, W.H., Bulkley, R. Konsistenzmessungen von Gummi-Benzollösungen // Kolloid-Zeitschrift. 1926. Vol. 39. P. 291–300. https://doi.org/10.1007/BF01432034.

[20] Frigaard, I. Simple yield stress fluids // Current Opinion in Colloid & Interface Science. 2019. Vol. 43. P. 80–93. https://doi.org/10.1016/j.cocis.2019.03.002.

[21] Olver, F.W.J., Lozier, D.W., Boisvert, R.F., Clark, C.W. (eds.). NIST Handbook of Mathematical Functions. New York: Cambridge University Press, 2010.

[22] Corless, R.M., Gonnet, G.H., Hare, D.E.G., Jeffrey, D.J., Knuth, D.E. On the Lambert W function // Advances in Computational Mathematics. 1996. Vol. 5. P. 329–359. https://doi.org/10.1007/BF02124750.

[23] You, Z., Huilgol, R.R., Mitsoulis, E. Application of the Lambert W function to steady shearing flows of the Papanastasiou model // International Journal of Engineering Science. 2008. Vol. 46, No. 8. P. 799–808. https://doi.org/10.1016/j.ijengsci.2008.02.002.

[24] Pitsillou, R., Georgiou, G.C., Huilgol, R.R. On the use of the Lambert function in solving non-Newtonian flow problems // Physics of Fluids. 2020. Vol. 32, No. 9. https://doi.org/10.1063/5.0021437.

[25] Housiadas, K.D., Georgiou, G.C. The analytical solution for the flow of a Papanastasiou fluid in ducts with variable geometry // Journal of Non-Newtonian Fluid Mechanics. 2023. Vol. 319. Article 105074. https://doi.org/10.1016/j.jnnfm.2023.105074.

ҒТАХР 28.15.19

https://doi.org/10.51889/2959-5894.2026.95.3.003

Ә.Ж. Жамбулатова[1*], Қ. Әлімхан[1]

[1]Л.Н. Гумилев атындағы Еуразия ұлттық университеті, Астана қ., Қазақстан
**e-mail: aigerimzhambulat@gmail.com*

СЫЗЫҚТЫҚ ЕМЕС АУЫСПАЛЫ ЖҮЙЕЛЕРДІ ШЕКТІ УАҚЫТ ІШІНДЕ ТҰРАҚТАНДЫРУ ӘДІСТЕРІ

Аңдатпа

Бұл мақалада ауысу заңдары еркін болатын сызықтық емес жүйелерді шекті уақыт ішінде тұрақтандыру мәселесі қарастырылады. Мұндай жүйелер басқару теориясында маңызды орын алады және робототехника, химиялық реакторлар, мехатроника және виртуалды өзара әрекеттесу жүйелері сияқты көптеген инженерлік салаларда кездеседі. Классикалық тұрақтылық теориясында жүйе күйінің тепе-теңдік нүктесіне асимптотикалық жақындауы қарастырылса, шекті уақыттағы тұрақтылық жүйе күйінің белгілі бір шектеулі уақыт аралығында нөлге жинақталуын қамтамасыз етеді. Бұл қасиет жүйенің жылдам әрекет етуін және сыртқы әсерлерге төзімділігін арттырады. Мақалада ауысу режимдері бар сызықтық емес жүйелердің математикалық моделін қарастырылып, ортақ Ляпунов функциясына негізделген тұрақтылық шарттары талданады. Сонымен қатар жүйені шекті уақыт ішінде тұрақтандару үшін дәрежелік интеграторды қосу әдісіне негізделген басқару тәсілдері сипатталады. Ұсынылған әдістер ауысу заңдары еркін болатын жағдайларда да жүйе тұрақтылын қамтамасыз етуге мүмкіндік береді. Зерттеудің нәтижелерін тексеру мақсатында бірнеше сандық мысалдар қарастырылып, жүйе MATLAB ортасында моделденді. Алынған графиктер жүйе күй айнымалыларының шекті уақыт ішінде нөлге жинақталатынын көрсетті. Сандық нәтижелер ұсынылған әдістердің тиімділігін және олардың практикалық қолданбаларда қолдануға жарамды екенін дәлелдейді.

Түйін сөздер: коммутацияланатын сызықтық емес жүйелер, шекті уақыттағы тұрақтандыру, Ляпунов функциясы, дәрежелік интегратор әдісі, басқару заңы.

Ә.Ж. Жамбулатова[1*], Қ. Әлімхан[1]
[1]Евразийский национальный университет имени Л.Н.Гумилева, г.Астана, Казахстан

МЕТОДЫ СТАБИЛИЗАЦИИ НЕЛИНЕЙНЫХ ПЕРЕКЛЮЧАЕМЫХ СИСТЕМ ЗА КОНЕЧНОЕ ВРЕМЯ

Аннотация

В статье исследуется задача стабилизации нелинейных переключаемых систем за конечное время при произвольных законах коммутации. Подобные динамические системы широко применяются в различных областях современной техники и управления, включая робототехнические комплексы, мехатронные устройства, химико-технологические процессы и системы виртуального взаимодействия. В отличие от классического подхода, основанного на асимптотической устойчивости, при котором состояние системы лишь постепенно приближается к положению равновесия, концепция устойчивости за конечное время гарантирует достижение равновесного состояния в течение заранее ограниченного промежутка времени. Данное свойство позволяет повысить быстродействие системы и улучшить её устойчивость к внешним возмущениям. В работе рассматривается математическая модель нелинейных систем с переключениями режимов функционирования. Для анализа устойчивости используются условия, основанные на общей функции Ляпунова. Кроме того, представлены методы синтеза управления, построенные на рекурсивной процедуре добавления степенного интегратора, обеспечивающие стабилизацию системы за конечное время. Предлагаемый подход сохраняет работоспособность даже при произвольном характере переключений между подсистемами. Для подтверждения теоретических результатов проведены численные эксперименты с использованием программного пакета MATLAB. Результаты моделирования показывают, что все переменные состояния системы достигают нулевого значения за конечное время. Полученные данные подтверждают эффективность разработанных алгоритмов управления и демонстрируют

перспективность их применения при решении практических задач управления нелинейными переключаемыми системами.

Ключевые слова: переключаемые нелинейные системы, стабилизация за конечное время, функция Ляпунова, метод добавления степенного интегратора, закон управления.

A. Zhambulatova[1*], K. Alimhan[1]
[1] L.N. Gumilyov Eurasian National University, Astana, Kazakhstan

FINITE-TIME STABILIZATION METHODS FOR NONLINEAR SWITCHED SYSTEMS

Abstract

This paper investigates the problem of finite-time stabilization for nonlinear switched systems under arbitrary switching laws. Such systems play an important role in control theory and are widely encountered in many engineering applications, including robotics, chemical reactors, mechatronic systems, and virtual interaction systems. While classical stability theory focuses on the asymptotic convergence of system states to an equilibrium point, finite-time stability guarantees that the system states reach the origin within a finite time interval. This property improves the response speed of the system and enhances its robustness against external disturbances. The paper considers a mathematical model of nonlinear systems with switching modes and analyzes stability conditions based on a common Lyapunov function. In addition, control design methods based on the adding a power integrator technique are presented to achieve finite-time stabilization. The proposed approaches ensure system stability even when the switching law is arbitrary. To verify the effectiveness of the developed methods, several numerical examples are investigated and simulated in the MATLAB environment. The obtained results demonstrate that the state variables converge to zero within a finite time. Numerical simulations confirm the effectiveness of the proposed control strategies and indicate their potential applicability in practical engineering systems.

Keywords: switched nonlinear systems, finite-time stabilization, Lyapunov function, adding a power integrator technique, control law.

Кіріспе

Соңғы жылдары коммутацияланатын сызықтық емес жүйелер басқару теориясының маңызды зерттеу бағыттарының біріне айналды. Мұндай жүйелер бірнеше динамикалық ішкі жүйелерден тұрады және динамикасы уақыт бойынша осы ішкі жүйелердің ауысуы арқылы сипатталады. Ауысу режимдері бар жүйелер робототехникада, химиялық реакторларды басқаруда, энергетикалық жүйелерде, мехатроникада және виртуалды өзара әрекеттесу құралығыларында кеңінен қолданылады. Осыған байланысты мұндай жүйелердің тұрақтылығы мен басқарылуын қамтамасыз ету мәселелері басқару теориясының маңызды ғылыми және қолданбалы міндеттерінің бірі болып табылады [1].

Классикалық басқару теориясында жүйелердің тұрақтылығы көбінесе асимптотикалық тұрақтылық тұрғысынан зерттеледі. Бұл жағдайда жүйе күйі тепе-теңдік нүктесіне уақыт шексіздікке ұмтылған кезде жақындайды. Алайда көптеген практикалық жүйелер үшін мұндай қасиет жеткіліксіз болуы мүмкін. Инженерлік қолданбаларда жүйе күйінің тепе-теңдік нүктесіне шекті уақыт аралығында жетуі талап етіледі. Осыған байланысты соңғы жылдары шекті уақыттағы тұрақтылық ұғымы кеңінен зерттелуде. Шекті уақыттағы тұрақтылық жүйенің күй айнымалылардың нөлдік тепе-теңдікке нақты бір уақыт шектеулі уақыт аралығында жинақталуын қамтамасыз етеді және мұндай басқару жүйелері жоғары жылдамдықпен жинақталу, сыртқы әсерлерге төзімділік және басқару сапасының жақсаруы сияқты бірқатар маңызды артықшылықтарға ие [2].

Сызықтық емес жүйелер үшін шекті уақыттағы тұрақтандыру мәселесі көптеген ғылыми еңбектерде қарастырылған. Әсіресе төменгі үшбұрышты құрылымдағы жүйелер үшін кері қадамдау әдісіне негізделген басқару тәсілдері кеңінен қолданылып келеді. Сонымен қатар, шекті уақыттағы тұрақтылықты қамтамасыз ету үшін дәрежелік интеграторды қосу әдісі ұсынылып, оның тиімділігі бірқатар жұмыстарда дәлелденген [3]. Дегенмен ауысу режимдері бар сызықтық емсе сызықтық емес жүйелер үшін шекті уақыттағы тұрақтандыру мәселесі әлдеқайда күрделі болып табылады. Себебі, жүйе динамикасы ауысу сигналдарының әсерінен

уақыт бойынша өзгеріп отырады. Сондықтан мұндай жүйелер үшін барлық қосалқы режимдерге ортақ Ляпунов функциясын құру және басқару заңын әртүрлі режимдер үшін бірдей түрде анықтау маңызды ғылыми мәселе болып табылады.

Осы жұмыста ауысу заңдары еркін болатын сызықтық емес жүйелердің белгілі бір класы үшін шекті уақыттағы тұрақтандыру мәселесі қарастырылады [4]. Жүйенің төменгі үшбұрышты құрылымы пайдалана отырып, дәрежелік интеграторды қосу әдісіне негізделген басқару тәсілдері талданады. Сонымен қатар, ұсынылған әдістердің тиімділігін көрсету мақсатында бірнеше сандық мысалдар қарастырылып, жүйенің динамикалық қасиеттері сандық моделдеу арқылы зерттелді.

Осылайша, зерттеудің негізгі мақсаты – ауысу режимдері бар сызықтық емес жүйелер үшін шекті уақыттағы тұрақтандыру әдістерін талдау және олардың тиімділігін сандық моделдеу арқылы көрсету.

Зерттеу әдіснамасы

Зерттеу шарттару мен әдістері

Бұл жұмыста ауысу режимдері бар сызықтық емес жүйелердің белгілі бір класы қарастырылады. Зерттелген жүйе төменгі үшбұрышты құрылымға ие және бірнеше ішкі жүйелерден тұратын ауысу жүйесі ретінде сипаталады. Жүйенің жалпы динамикалық моделі келесі түрде беріледі:

$$\begin{aligned}
\dot{x}_1 &= x_2 + \eta_{1,\sigma(t)}(x,u,t),\\
\dot{x}_2 &= x_3 + \eta_{2,\sigma(t)}(x,u,t),\\
&\dots\\
\dot{x}_n &= d(t)u + \eta_{n,\sigma(t)}(x,u,t),
\end{aligned}$$

Мұндағы $x \in R^n$ - жүйенің күй векторы, u - басқару сигналы, ал $\sigma(t)$ - уақыт бойынша өзгеретін ауысу заңы. Жүйе бірнеше ішкі жүйелерден тұрады және олардың арасында ауысу сигналы арқылы режим өзгерісі жүзеге асады.

Зерттеу барысында келесі шарттар қабылданды: Жүйедегі бейсызық функциялар үздіксіз дифференциаланатын функциялар болып табылады және олар белгілі бір шектеулерді қанағаттандырады. Сонымен қатар басқару каналына әсер ететін $d(t)$ функциясы нөлден үлкен және белгілі бір шектеулі аралықты өзгереді. Жүйені шекті уақыт ішінде тұрақтандыру үшін дәрежелік интеграторды қосу әдісі қолданылады [6]. Бұл әдіс рекурсивті басқару алгоритмін құруға мүмкіндік береді және жүйе үшін ортақ Ляпунов функциясын анықтауға негізделген [7]. Әрбір қадамда виртуалды басқару функциялары енгізіліп, жүйенің күй айнымалылары үшін жаңа координаталық түрлендірулер анықталады. Жүйенің тұрақтылығын талдау үшін Ляпуновтың тікелей әдісі қолданылады. Ляпунов функциясының туындысы теріс анық болатындығы көрсетіліп, жүйенің шекті уақыттағы тұрақтылығы дәлелденеді.

Ұсынылған әдістің тиімділігін тексеру мақсатынды сандық моделдеу жүргізілді. Моделдеу MATLAB бағдарламалық ортасында жүзеге асырылып, ауысу режимдері бар жүйелер үшін күй айнымалыларының уақыт бойынша өзгерісі зерттеледі.

Зерттеу нәтижелері

Зерттеу барысында ауысу режимдері бар сызықтық емес жүйелерді шекті уақыт ішінде тұрақтандыру мәселесі қарастырылады. Ұсынылған әдіс дәрежелік интеграторларды қосу тәсіліне негізделіп, жүйе үшін ортақ Ляпунов функциясын құруға мүмкіндік береді. Осы тәсілдің көмегімен ауысу заңдары еркін болатын жағдайда да жүйені шекті уақыт ішінде тұрақтандыратын басқару заңы алынды.

Теориялық талдау нәтижесінде жүйе үшін күй бойынша кері байланысқа негізделген басқару заңы құрылды. Сонымен қатар, барлық ішкі жүйелер үшін ортақ Ляпунов функциясы анықталды. Оның туындысы теріс екендігі көрсетілді. Бұл жүйенің жабық контурында шекті уақыттағы тұрақтылықтың орындалатынын дәлелдейді [8].

Ұсынылған әдістің тиімділігін тексеру мақсатында сандық моделдеу жүргізілді. Алдымен екі режимді CSTR жүйесі қарастырылды. Жүйенің бастапқы шарттары $x_1 = -2.5$ және $x_2 = 3.8$ деп алынды. Моделдеу нәтижелері жүйе күй айнымалыларының уақыт бойынша нөлге шекті уақытта жинақталатынын көрсетті. 1-3 суреттерде жүйе күй айнымалыларының өзгеруі және ауысу сигналының динамикасы бейнеленген. Нәтижелер ұсынылған шекті уақыттағы басқару әдісінің асимптотикалық басқару әдісіне қарағанда жылдамырақ жинақталуды қамтамасыз ететінін көрсетеді.

Екінші мысал ретінде виртуалды ортасы ауысатын тактильдік дисплей жүйесі қарастырылды. Бұл жүйе бірнеше режимнен тұратын ауысу жүйесі ретінде моделденді. Сандық моделдеу нәтижелері 4-5 суреттерді көрсетілген. Алынған нәтижелер жүйе күй айнымалыларының шекті уақыт ішінде нөлге жинақталатынын және ұсынылған басқару алгоритмінің ауысу сигналдарына тәуелсіз жұмыс істейтінін көрсетті.

Жалпы алғанда, жүргізілген сандық тәжірибелер ұсынылған әдістің тиімділігін растады. Ұсынылған басқару заңы ауысу режимдері бар сызықтық емес жүйелердің шекті уақыттағы тұрақтылығын қамтамасыз етеді және күй айнымалыларының жылдам жинақталуына мүмкіндік береді [9].

Дискуссия

Жүргізілген зерттеу нәтижелері ауысу режимдері бар сызықтық емес жүйелерді шекті уақыт ішінде тұрақтандыру мәселесін тиімді шешуге мүмкіндік беретін басқару тәсілін ұсынуға мүмкіндік берді. Ұсынылған әдіс дәрежелік интеграторды қосу тәсіліне негізделген және жүйенің барлық ішкі жүйелері үшін ортақ Ляпунов функциясына құруға мүмкіндік береді. Бұл тәсілдің маңызды ерекшелігі – басқару заңының ауысу сигналынан тәуелсіз құрылуы және жүйенің еркін ауысу жаңдайында тұрақтылығын қамтамасыз етуі.

Алынған нәтижелер көрсеткендей, ұсынылған басқару алгоритмі жүйе күй айнымалыларының шекті уақыт ішінде тепе-теңдік күйге жинақталуын қамтамасыз етеді. Бұл қасиет асимптотикалық тұрақтылықпен салыстырғанда анағұрлым тиімді, себебі, жүйе күйі нақты бір шектеулі уақыт ішінде нөлге жетеді. Сонымен қатар, шекті уақыттағы басқару жүйенің динамикалық қасиеттерін жақсартып, сыртқы әсерлерге және белгісіздіктерге төзімділігін арттырады. Сандық моделдеу нәтижелері ұсынылған әдістің практикалық тұрғыдан тиімді екенін көрсетті. CSTR жүйесі мен тактильді дисплей жүйесіне жүргізілген моделдеу нәтижелері жүйе күй айнымалыларының ауысу режимдеріне қарамастан шекті уақыт ішінде нөлге жинақталатынын көрсетті. Сонымен қатар, ұсынылған шекті уақыттағы басқару әдісі асимптотикалық тұрақтандыру әдістерімен салыстырғанда жүйе траекторияларының жылдамырақ жинақталуын қамтамасыз ететіні байқалды.

Дегенмен ұсынылған әдістің белгілі бір шектеулері де бар. Атап айтқанда, зерттеу барысында жүйелердің төменгі үшбұрышты құрылымы қарастырылды. Бұл әдісті құрылымы күрделірек немесе параметрлік белгісіздіктері бар ауысу жүйелеріне қолдану қосымша зерттеулерді талап етеді. Жалпы алғанда алынған нәтижелер ауысу режимдері бар сызықтық емес жүйелер үшін шекті уақыттағы тұрақтандыру әдістерінің тиімді екенін көрсетеді және ұсынылған тәсіл басқару теориясында әрі қарай дамытуды қажет ететін перспективалы бағыттардың бірі болып табылады.

Ауысу режимдері бар сызықтық месе жүйелерді шекті уақыт ішінде тұрақтандыру әдістері

Бұл бөлімде ауысу режимдері бар сызықтық емес жүйелерді еркін ауысулар жағдайындағы шекті уақыт ішінде тұрақтандыруға арналған негізгі әдістер математикалық формулалар

негізінде талданды. Негізгі назар ортақ Ляпунов функциясына негізделген тәсілдерге және дәрежелік интеграторды қосу әдісіне аударылады.

Ауысу режимдері бар сызықтық емес жүйенің жалпы математикалық моделі

Қарастырылатын ауысу режимдері бар сызықтық емес жүйелер төменгі үшбұрышты құрылымға ие және келесі түрде сипатталады:

$$\begin{aligned}
\dot{x}_1 &= x_2 + \eta_{1,\sigma(t)}(x),\\
\dot{x}_2 &= x_3 + \eta_{2,\sigma(t)}(x),\\
&\vdots\\
\dot{x}_{n-1} &= x_n + \eta_{n-1,\sigma(t)}(x),\\
\dot{x}_n &= d(t)u + \eta_{n,\sigma(t)}(x),
\end{aligned}$$

Мұндағы $[x_1, \dots, x_n]^T \in \mathbb{R}^n$ - күй векторы, u - басқару сигналы, $\sigma(t) \in \{1, \dots, m\}$ - еркін ауысу заңы, $d(t)$ - басқару каналындағы белгісіз, бірақ оң және шектелген параметр.

Бұл модель ауысу режимдері бар сызықтық емес жүйелердің кең класын қамтиды және көптеген инженерлік жүйелерге тән құрылымды сипаттайды.

Шекті уақыттағы тұрақтылықтың Ляпунов шарттары. Шекті уақыттаңы тұрақтылықты зерттеу үшін біз Ляпунов функциясына негізделген әдісті қолданылады. Егер оң анықталған $V(x)$ функциясы үшін

$$\dot{V}(x) + cV^{\alpha}(x) \le 0, \quad 0 < \alpha < 1, c > 0,$$

Шарты орындалса, онда жүйе күйлері тепе-теңдік күйге шекті уақыт ішінде жақындайды. Ауысу режимдері бар жүйелер үшін барлық ішкі жүйелерге ортақ Ляпунов функциясының болуы еркін ауысулар жағдайында да тұрақтылықты қамтамасыз етеді.

Теорема. Қарастырылатын жүйе үшін дәрежелік интеграторды қосу әдісіне негізделген басқару заңы қолданылған жағдайда, егер барлық $i = 1, \dots, n$ үшін $\eta_{i,\sigma(t)}(x)$ функциялары шектелген болса және $d(t) > 0$ шарты орындалса, онда жүйе еркін ауысулар жағдайында глобальды шекті уақыт ішінде тұрақты болады.

Дәлелдеу. Жүйе үшін ортақ Ляпунов функциясын $V(x)$ деп қарастырайық. Дәрежелік интеграторды қосу әдісін қолдану арқылы басқару заңы таңдалады, соның нәтижесінде жабық контурында Ляпунов функциясының туындысы келесі бағалауды қанағаттандырады:

$$\dot{V}(x) \le -cV^{\alpha}(x), \quad 0 < \alpha < 1$$

Бұл теңсіздік шекті уақыттағы тұрақтылықтың жеткілікті шарты болып табылады [11]. Сонымен қатар, Ляпунов функциясының барлық ішкі жүйелер үшін ортақ болуы жүйенің ауысу заңына тәуелсіз тұрақтылығын қамтамасыз етеді. Сол себепті жүйе күйлері нөлдік тепе-теңдік күйге шекті уақыт ішінде жақындайды [12].

Кері қадамдау әдісі және оның шектеулері

Төменгі үшбұрышты құрылымдағы ауысу режимдері бар сызықтық емес жүйелер үшін кері қадамдау әдісі кеңінен қолданылады. Бұл тәсілде жүйе рекурсивті түрде қарастырылып, әрбір қадамда виртуалды басқару енгізіледі. Нәтижесінде Ляпунов функциясының туындысы теріс анық болады. Алайда, классикалық кері қадамдау әдісі көбінесе асимптотикалық тұрақтылықты ғана қамтамасыз етеді, яғни жүйе күйі нөлдік тепе-теңдік күйге тек уақыт

шексіздікке ұмтылған кезде ғана жақындайды. Сондықтан шекті уақыттағы тұрақтандыруды қамтамасыз ету үшін кері қадамдау әдісін жетілдіру қажет

Дәрежелік интеграторды қосу әдісі

Шекті уақыттағы тұрақтандыруды жүзеге асыру үшін дәрежелік интеграторды қосу әдісі қолданылады. Бұл тәсіл кері қадамдау әдісінің құрылымын сақтай отырып, Липшиц шартын қанағаттандырмайтын басқару заңдарын құруға мүмкіндік береді [10].

Әрбір қадамда виртуалды басқару келесі түрде таңдалады:

$$x_{i+1}^{*} = -|\chi_i|^{\rho_i}\varphi_i(x), \quad 0 < \rho_i < 1,$$

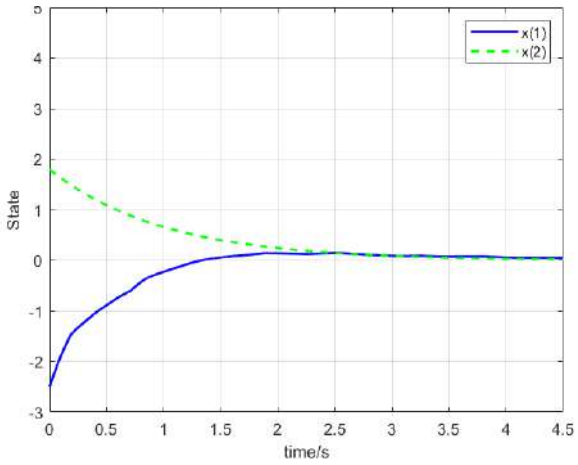
Мұндағы χ_i - кординаталық түрлендіруден кейінгі айнымалы, $\varphi_i(x)$ - оң анықталған, үздіксіз функция. Осы рекурсивті құрылым арқылы Ляпунов функциясының туындысы дәрежелік түрде теріс анық болады, яғни:

$$\dot{V}(x) \le -k\sum_{i=1}^{n}|\chi_i|^{\beta}, \quad 0 < \beta < 1,$$

Бұл жүйе күйлерінің шекті уақыт ішінде нөлге жинақталуын қамтамасыз етеді.

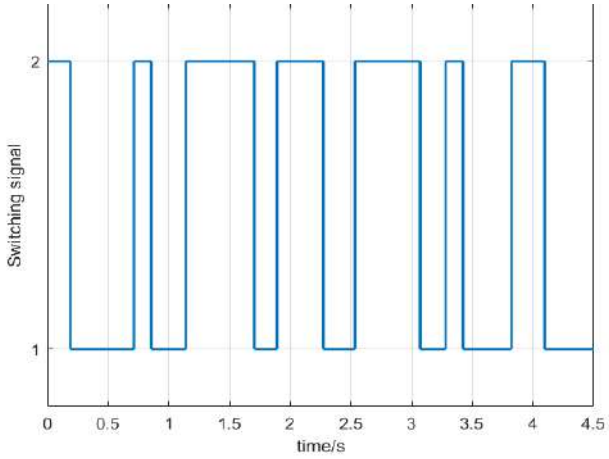
Әдістерді салыстырмалы талдау. Жоғарыда қарастырылған әдістерді салыстыратын болсақ, кері қадамдау әдісі ауысу режимдері бар сызықтық емес жүйелер үшін асимптотикалық тұрақтылықты тиімді қамтамасыз етеді, ал дәрежелік интеграторды қосу әдісі жүйе күйлерінің шекті уақыт ішінде нөлге жақындауын жүзеге асырады. Ортақ Ляпунов функциясымен біріктірілген дәрежелік интеграторды қосу әдісі еркін ауысулар жағдайында жүйені глобальды шекті уақыт ішінде тұрақтандыруды қамтамасыз ететін ең тиімді тәсілдердің бірі болып табылады [13, 14].

Алдымен екі режимі CSTR жүйесі қарастырылады. Жүйенің бастапқы шарттары $x_1(0) = -2.5$ және $x_2(0) = 3.8$ деп алынды. Ұсынылған басқару заңы қолданылған жағдайда жүйе күй айнымалыларының динамикасы 1 суретте көрсетілген.

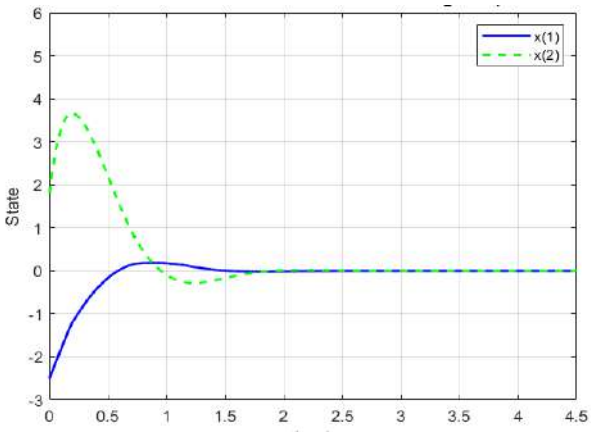


Сурет 1. Жүйе күй айнымалыларының уақыт бойынша өзгеруі

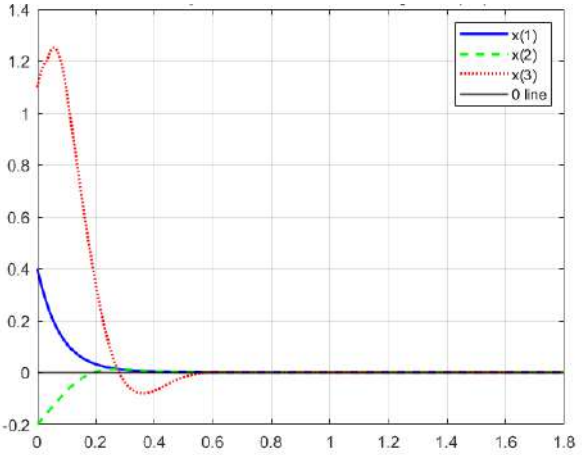
Екінші мысал ретінде виртуалды ортасы ауысатын тактильдік дисплей жүйесі қарастырылады. Ұсынылған басқару алгоритмі қолданылған кезде жүйе күй айнымалыларының динамикасы 4 суретте көрсетілген. Бұл бөлімде біз шекті уақыттағы тұрақтандыру әдістерінің тиімділігін сандық моделдеу арқылы көрсетілген. Әдебиеттерде ұсынылған ауысатын сызықтық емес жүйелерге арналған мысалдар қайта есептеліп, алынған нәтижелер теориялық қорытындылармен салыстырылды. Сандық моделдеу Matlab ортасында жүргізілді.



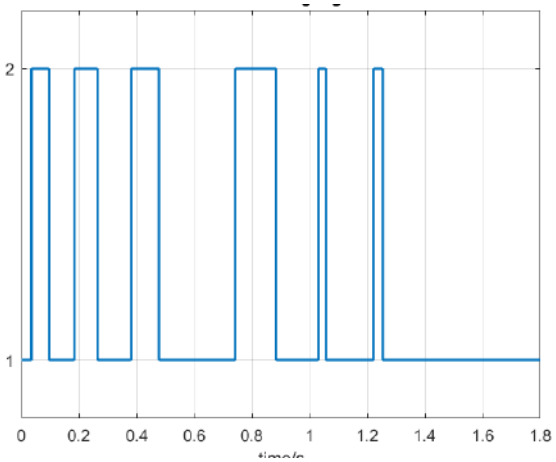
Сурет 2. Ауысу сигналы



Сурет 3. Асимптотикалық басқару кезіндегі жүйе күй айнымалылары



Сурет 4. Ауысу жүйесінің күй айнымалылары



Сурет 5. Ауысу сигналы

Екі режимді CSTR жүйесі. Алдымен үздіксіз араласатын реактор (CSTR – continuous-tank reactor) жүйесін қарастырайық. Координаталық түрлендіруден кейін бұл жүйе екі режимді ауысатын сызықтық емес жүйе ретінде келесі түрде жазылады:

$$\dot{x}_1 = x_2 + a_{\sigma(t)} x_1,$$
$$\dot{x}_2 = u,$$

Мұндағы ауысу заңы – $\sigma(t) \in \{1,2\}$, ал параметрлер $a_1 = 0.5, \quad a_2 = 2.$

Шекті уақытта тұрақтандыруды қамтамасыз ету үшін дәрежелік интеграторды қосу әдісіне негізделген басқару заңы құрылады. Бұл басқару заңы Липшиц шартын қанағаттандырмайды және ауысу сигналына тәуелсіз түрде анықталады. Сандық моделдеу барысында бастапқы шарттар әртүрлі мәндерде таңдалып, ауысу сигналы кездейсоқ түрде өзгертілді. Нәтижелер көрсетілгендей, жүйе күйлері $x_1(t)$ және $x_2(t)$ шекті уақыт ішінде нөлге жинақталады.

Графиктерді талдау. Сурет 1-де $x_1(t)$ және $x_2(t)$ күй айнымалыларының уақыт бойынша өзгеруі көрсетілген. Графиктерден жүйенің күйі шекті уақыт ішінде тепе-теңдікке жететіні айқын байқалады. Ауысу сигналының жиі өзгеруіне қарамастан, траекториялар тұрақты түрде нөлге бағытталған, бұл ортақ Ляпунов функциясының тиімділігін растайды.

Тактильді дисплей жүйесінің сандық талдауы. Келесі мысал ретінде ауысатын виртуалды орталармен жұмыс істейтін тактильді дисплей жүйесін қарастырамыз. Бұл жүйе бірнеше режимнен тұратын ауысатын сызықты емес динамикамен сипатталады:

$$\dot{x}_1 = -12x_1 + x_2,$$
$$\dot{x}_2 = -\alpha_{\sigma(t)}x_2 + x_3,$$
$$\dot{x}_3 = \beta_{\sigma(t)}x_2 - \gamma_{\sigma(t)}x_3 + d(t)u,$$

Мұндағы параметрлер режимге байланысты өзгереді, ал $d(t)$ - оң және шектелген функция. Бұл жүйе үшін де дәрежелік интеграторды қосу әдісіне негізделген шекті уақыттағы басқару заңы құрылады. Аталған басқару заңы барлық режимдер үшін ортақ құрылымға ие.

Сурет 2-де тактильді дисплей жүйесінің күй айнымалыларының уақыт бойынша өзгеруі көрсетілген. Нәтижелер жүйе күй айнымалыларының қысқа уақыт ішінде нөлдік тепе-теңдік күйге жинақталатынын айқын көрсетеді. Сонымен қатар, ауысу сигналының әсеріне қарамастан, шекті уақыттағы жинақталу қасиеті сақталатыны байқалады.

Шекті уақыттағы асимптотикалық басқаруды салыстыру. Сандық нәтижелер шекті уақыттағы басқарудың асимптотикалық басқарумен салыстырғандағы артықшылықтарын айқын көрсетеді. Асимптотикалық басқару кезінде күй айнымалылары нөлге бітіндеп, шексіз уақыт аралығында жақындайды. Ал шекті уақыттағы басқару жағдайында жүйе күйлері нақты бір шекті уақыттан кейін дәл нөлге жетеді. Сурет 3-те шекті уақыттағы және асимптотикалық басқару жағдайындағы күй траекториялары салыстырмалы түрде көрсетілген. Шекті уақыттағы басқару жүйенің жинақталу жылдамдығын едәуір арттыратыны байқалады.

Жүргізілген сандық моделдеу нәтижелері шекті уақыт уақыттағы. Тұрақтандыру әдістерінің тиімділігін толық растайды. Ауыспалы сызықтық емес жүйелер үшін ортақ Ляпунов функциясы және дәрежелік интеграторды қосу әдісіне негізделген басқару заңдары еркін ауысу жағдайында да жүйе күйлерінің шекті уақыт ішінде жинақталуын қамтамасыз етеді. Бұл нәтижелер шекті уақыттағы басқарудың практикалық қолданбалар үшін жоғары әлеуетке ие екенін көрсетеді.

Қорытынды

Бұл жұмыста ауысу режимдері бар сызықтық емес жүйелердің белгілі бір класы үшін шекті уақыттағы тұрақтандыру мәселесі қарастырылды. Жүйені тұрақтандыру үшін дәрежелік интеграторды қосу әдісіне негізделген басқару тәсілі қолданылды. Сонымен қатар, барлық ішкі жүйелер үшін ортақ Ляпунов функциясын құру арқылы жүйенің шекті уақыттағы тұрақтылығын дәлелденді. Ұсынылған әдіс ауысу сигналдарына тәуелсіз басқару заңын құрға мүмкіндік береді және жүйенің еркін ауысулар жағдайында тұрақтылығын қамтамасыз етеді. Жүргізілген сандық моделдеу нәтижелері ұсынылған басқару тәсілінің тиімді екенін көрсетті. Атап айтқанда, жүйе күй айнымалылары шекті уақыт ішінде нөлдік тепе-теңдік күйге жақындайды және жүйенің динамикалық қасиеттері жақсарады. Жалпы алғанда, алынған нәтижелер ауысу режимдері бар сызықтық емес жүйелер үшін шекті уақыттағы тұрақтандыру әдістерінің тиімді екенін көрсетеді. Ұсынылған тәсіл басқару теориясындағы осы бағыттағы зерттеулерді одан әрі дамытуға мүмкіндік береді.

Пайдаланылған әдебиеттер тізімі

[1] Bhat S. P., Bernstein D. S. Finite-time stability of continuous autonomous systems // SIAM Journal on Control and Optimization. – 2000. – Vol. 38, No. 3. – P. 751–766.

[2] Liberzon D. Switching in Systems and Control. – Boston: Birkhäuser, 2003. – 233 p.

[3] Lin W., Qian C. Adding one power integrator: A tool for global stabilization of high-order lower triangular systems // Systems & Control Letters. – 2000. – Vol. 39. – P. 339–351.

[4] Hong Y. Finite-time stabilization and stabilizability of a class of controllable systems // Systems & Control Letters. – 2002. – Vol. 46. – P. 231–236.

[5] Huang X., Lin W., Yang B. Global finite-time stabilization of a class of uncertain nonlinear systems // Automatica. – 2005. – Vol. 41. – P. 881–888.

[6] Wu J. L. Stabilizing controller design for switched nonlinear systems in strict-feedback form // Automatica. – 2009. – Vol. 45. – P. 1092–1096.

[7] Ma R., Zhao J. Backstepping design for global stabilization of switched nonlinear systems in lower triangular form under arbitrary switchings // Automatica. – 2010. – Vol. 46. – P. 1819–1823.

[8] Wang X., Hong Y., Jiang Z. Finite-time input-to-state stability of switched nonlinear systems // Proceedings of the Chinese Control Conference. – 2008. – P. 479–483.

[9] Orlov Y. Finite-time stability and robust control synthesis of uncertain switched systems // SIAM Journal on Control and Optimization. – 2004. – Vol. 43. – P. 1253–1271.

[10] Ding S., Li S., Zheng W. X. Nonsmooth stabilization of a class of nonlinear cascaded systems // Automatica. – 2012. – Vol. 48. – P. 2597–2606.

[11] Chen G., Yang Y., Li J. Finite-time stability of hybrid dynamical systems // IET Control Theory & Applications. – 2012. – Vol. 6. – P. 8–13.

[12] Khoo S., Yin J., Man Z., Yu X. Finite-time stabilization of stochastic nonlinear systems in strict-feedback form // Automatica. – 2013. – Vol. 49. – P. 1403–1410.

[13] Liang Y. J., Ma R., Wang M., Fu J. Global finite-time stabilisation of a class of switched nonlinear systems // International Journal of Systems Science. – 2014.

[14] Sun Y., Wang L. On stability of a class of switched nonlinear systems // Automatica. – 2013. – Vol. 49. – P. 305–307.

[15] Branicky M. S. Multiple Lyapunov functions and other analysis tools for switched and hybrid systems // IEEE Transactions on Automatic Control. – 1998. – Vol. 43. – P. 475–482.

References

[1] Bhat S. P., Bernstein D. S. Finite-time stability of continuous autonomous systems // SIAM Journal on Control and Optimization. – 2000. – Vol. 38, No. 3. – P. 751–766.

[2] Liberzon D. Switching in Systems and Control. – Boston: Birkhäuser, 2003. – 233 p.

[3] Lin W., Qian C. Adding one power integrator: A tool for global stabilization of high-order lower triangular systems // Systems & Control Letters. – 2000. – Vol. 39. – P. 339–351.

[4] Hong Y. Finite-time stabilization and stabilizability of a class of controllable systems // Systems & Control Letters. – 2002. – Vol. 46. – P. 231–236.

[5] Huang X., Lin W., Yang B. Global finite-time stabilization of a class of uncertain nonlinear systems // Automatica. – 2005. – Vol. 41. – P. 881–888.

[6] Wu J. L. Stabilizing controller design for switched nonlinear systems in strict-feedback form // Automatica. – 2009. – Vol. 45. – P. 1092–1096.

[7] Ma R., Zhao J. Backstepping design for global stabilization of switched nonlinear systems in lower triangular form under arbitrary switchings // Automatica. – 2010. – Vol. 46. – P. 1819–1823.

[8] Wang X., Hong Y., Jiang Z. Finite-time input-to-state stability of switched nonlinear systems // Proceedings of the Chinese Control Conference. – 2008. – P. 479–483.

[9] Orlov Y. Finite-time stability and robust control synthesis of uncertain switched systems // SIAM Journal on Control and Optimization. – 2004. – Vol. 43. – P. 1253–1271.

[10] Ding S., Li S., Zheng W. X. Nonsmooth stabilization of a class of nonlinear cascaded systems // Automatica. – 2012. – Vol. 48. – P. 2597–2606.

[11] Chen G., Yang Y., Li J. Finite-time stability of hybrid dynamical systems // IET Control Theory & Applications. – 2012. – Vol. 6. – P. 8–13.

[12] Khoo S., Yin J., Man Z., Yu X. Finite-time stabilization of stochastic nonlinear systems in strict-feedback form // Automatica. – 2013. – Vol. 49. – P. 1403–1410.

[13] Liang Y. J., Ma R., Wang M., Fu J. Global finite-time stabilisation of a class of switched nonlinear systems // International Journal of Systems Science. – 2014.

[14] Sun Y., Wang L. On stability of a class of switched nonlinear systems // Automatica. – 2013. – Vol. 49. – P. 305–307.

[15] Branicky M. S. Multiple Lyapunov functions and other analysis tools for switched and hybrid systems // IEEE Transactions on Automatic Control. – 1998. – Vol. 43. – P. 475–482.

ҒТАХР 27.39.19

https://doi.org/10.51889/2959-5894.2026.95.3.004

Н.С. Жаңабергенова[1,2]*, А.М. Темирханова[3]

[1]Л.Н. Гумилев атындағы Еуразия ұлттық университеті,
[2]Қазақ ұлттық спорт университеті, Астана қ., Қазақстан
[3]Л.Н. Гумилев атындағы Еуразия ұлттық университеті, Астана қ., Қазақстан
**e-mail:* zhanabergenova.ns@gmail.com

КВАЗИСЫЗЫҚТЫ ИНТЕГРАЛДЫҚ ХАРДИ ТИПТІ ТЕҢСІЗДІКТЕР

Аңдатпа

Квазисызықты Харди типті операторларға өзек енгізу маңызды бағыт болып табылады. Дегенмен, өзектің жалпы түрі үшін нәтижелер алу анықталмаған мәселе, сондықтан зерттеушілер белгілі бір өзек класстары үшін Харди типті операторларын зерттейді. Біз [21] жұмыста басталған зерттеулерді жалпылай отырып, O_1^- және O_1^+ кластарына жататын квазисызықты интегралдық Харди типті теңсіздіктердің $0 < r < \infty, 1 < p \leq q < \infty$ параметрлік диапазондарындағы орындалу критерийлері алынды. Яғни, салмақты Лебег кеңістіктеріндегі өзекті Харди операторы қатысқан квазисызықты операторлардың шенелімділігі үшін қажетті және жеткілікті шарттары анықталды. Бұл жалпыланған бисызықты операторларды зерттеуге мүмкіндік береді. Атап айтқанда, өзегі бар квазисызықты теңсіздіктердің нәтижелері бисызықты теңсіздіктерді бағалау үшін балама дәлелдемелер ұсынады. Бұл зерттеу функционалдық талдаудың интегралдық операторлар теориясына дамытады және дифференциалдық теңдеулердің шешімдерін табуда қолданыс табуы мүмкін.

Түйін сөздер: Харди теңсіздік, салмақты Лебег кеңістігі, салмақты функция, квазисызықты оператор.

Н.С. Жанабергенова[1,2], А.М. Темирханова[3]
[1]Евразийский национальный университет им. Л.Н. Гумилева, г.Астана,
[2]Казахский национальный университет спорта, г.Астана, Казахстан
[3]Евразийский национальный университет им. Л.Н. Гумилева, г.Астана, Казахстан

КВАЗИЛИНЕЙНЫЕ ИНТЕГРАЛЬНЫЕ НЕРАВЕНСТВА ТИПА ХАРДИ

Аннотация

Введение ядра в квазилинейные операторы типа Харди является важным направлением. Однако получение результатов для общего типа ядра является неопределенной проблемой, поэтому исследователи изучают операторы типа Харди для определенных классов ядер. Обобщая исследования, начатые в [21], мы получили критерии справедливости квазилинейных интегральных неравенств типа Харди, принадлежащих классам O_1^- и O_1^+ в диапазонах параметров $0 < r < \infty, 1 < p \leq q < \infty$. То есть были определены необходимые и достаточные условия справедливости квазилинейных операторов, содержащих ядерный оператор Харди в весовых пространствах Лебега. Это позволяет нам изучать обобщенные билинейные операторы. В частности, результаты квазилинейных неравенств с ядрами дают альтернативные доказательства для оценки билинейных неравенств. Это исследование вносит вклад в теорию интегральных операторов в функциональном анализе и может быть использовано при нахождении решений дифференциальных уравнений.

Ключевые слова: неравенство Харди, весовая пространства Лебега, весовая функция, квазилинейный оператор.

N. Zhangabergenova[1,2], A. Temirkhanova[3]
[1]L.N. Gumilyov Eurasian National University, Astana, Kazakhstan
[2]Kazakh National University of Sports, Astana, Kazakhstan
[2]L.N. Gumilyov Eurasian National University, Astana, Kazakhstan

QUASILINEAR INTEGRAL HARDY-TYPE INEQUALITIES

Abstract

The introduction of a kernel into quasilinear Hardy-type operators is an important direction. However, obtaining results for the general type of kernel is an undefined problem, so researchers study Hardy-type operators for certain kernel classes. Generalizing the research started in [21], we obtained the criteria for the validity of quasilinear integral Hardy-like inequalities belonging to the classes O_1^- and O_1^+ in the parameter ranges $0 < r < \infty, 1 < p \le q < \infty$. That is, necessary and sufficient conditions for the validity of quasilinear operators involving the kernel Hardy operator in weighted Lebesgue spaces were determined. This allows us to study generalized bilinear operators. In particular, the results of quasilinear inequalities with kernels provide alternative method for estimating bilinear inequalities. This study contributes to the theory of integral operators in functional analysis and can be used in finding solutions to differential equations.

Keywords: Hardy inequality, weighted Lebesgue space, weight function, quasilinear operator.

Кіріспе

Негізгі ережелер

Бұл жұмыста өзегі бар квазисызықты оператор қатысқан интегралдық Харди типті теңсіздіктің орындалуының қажетті және жеткілікті шарттары параметрдің кейбір жағдайында алынды. Дәлелдеу барысында локализация әдісі қолданылады. Квазисызықты теңсіздіктер үшін алынған критерийлерді бисызықты Харди теңсіздіктерді сипаттау кезінде қолдануға болады. Сонымен қатар, нәтижелер функционалдық анализ теориясы, операторлар теориясы, гармоникалық анализ, дифференциалдық теңдеулер және спектралды операторлар теориясында қолданыс табуы мүмкін.

Классикалық Харди теңсіздігі, алғаш рет 20 ғасырдың басында Г.Х. Харди енгізген еді. Жылдар өте келе, бұл теңсіздіктер әртүрлі бағыттарда жалпыланды, көпөлшемді және дискреттік жағдайларға, сондай-ақ квазисызықты және бисызықты теңсіздіктер сияқты күрделі теңсіздіктерге дейін кеңейді.

Интегралдық Харди операторы

$$Hf(x) = \int_0^x f(t)dt \tag{1}$$

және оның түйіндес жағдайының салмақты Лебег кеңістігінде шенелімділік критерийлері [1-3] және [4] сияқты еңбектерде алынған. Кейінгі жылдары (1) түріндегі $K(\cdot,\cdot)$ теріс емес өзекті операторлар белсенді зерттеле бастады. Дегенмен, мұндай оператордың жалпы түрі үшін шенелімділік критерийі анықталмаған проблема болып табылады, нәтижесінде зерттеушілер нақты шартты қанағаттандыратын өзектер класын енгізді. 1989-1990 жылдары Р. Ойнаров [5], сондай-ақ 1991 жылы оған тәуелсіз түрде американдық математиктер Р. Керман мен С. Блум [6] келесі шартты қанағаттандыратын өзекпен берілген (1) операторын зерттеді: барлық $0 < t \le \tau \le x < \infty$ айнымалылары үшін

$$d^{-1}(K(x,\tau) + K(\tau,t)) \le K(x,t) \le d(K(x,\tau) + K(\tau,t)) \tag{2}$$

шарт орындалатындай $d > 0$ табылады. Ал [7-9] жұмыстарда жалпыланған $O_n^{\pm}, n \ge 1$ өзектер класы таныстырылып, осы класқа жататын өзектер қатысқан интегралдық салмақты теңсіздіктер талданды. Сонымен қатар, (1) түріндегі операторлардың салмақты бағалауы Морри тәріздес кеңістікте де анықталды [10, 11]. [10] жұмыста (1) операторының көпөлшемді жағдайының Лебег кеңістігінен локальды Морри тәріздес кеңістікке шенелімділігі алынды

және бұл шенелімділікті сипаттайтын теңсіздік квазисызықты операторлар үшін салмақты теңсіздікке эквивалентті екенін көрсетті. Осы зерттеуден кейін интегралдық квазисызықты Харди типті операторлардың салмақты бағалаулары маңызды рөл атқара бастады, квазисызықты операторларға деген қызығушылық артты (мысалы, [12-16] қараңыз).

Бұл мақалада біз (1) интегралдық Харди операторының кеңейтілген түрі өзек қатысқан квазисызықты операторды қарастырамыз. Қарастырылып отырған мәселе келесідей тұжырымдалады: Айталық $0 < r, q, p < \infty$, $\frac{1}{p} + \frac{1}{p'} = 1$ және $I = (0, \infty)$. w, u және v - I аралығында өлшенетін және локальды интегралданатын теріс емес функциялар болсын, яғни оларды салмақты функциялар дейміз.

$f \geq 0$ үшін келесідей салмақты теңсіздіктерді қарастырамыз

$$\left(\int_0^\infty u(x) \left(\int_0^x \left(\int_0^t K(t,s) f(s) ds \right)^r w(t) dt \right)^{\frac{q}{r}} dx \right)^{\frac{1}{q}} \leq C \left(\int_0^\infty v(x) f^p(x) dx \right)^{\frac{1}{p}}, \qquad (3)$$

$$\left(\int_0^\infty u(x) \left(\int_x^\infty \left(\int_t^\infty K(s,t) f(s) ds \right)^r w(t) dt \right)^{\frac{q}{r}} dx \right)^{\frac{1}{q}} \leq C \left(\int_0^\infty v(x) f^p(x) dx \right)^{\frac{1}{p}} \qquad (4)$$

мұндағы $K(\cdot,\cdot) \geq 0$ өзектері $O_1^{\pm}$ кластарына жатады. Осы кластардың анықтамаларын берейік.

Анықтама 1. $K(\cdot,\cdot)$ теріс емес өлшемді функциясы $\{(t,s), 0 < s \leq t < \infty\}$ жиынында анықталған және екінші айнымалы бойынша өспейтін болсын. $K(t,s)$ функциясы O_1^- класына жатады дейміз, егер $U(\cdot)$ және $Q(\cdot,\cdot)$ теріс емес өлшемді функциялары және $d \geq 1$ тұрақтысы табылып, барлық $t, \tau, s: 0 < s \leq \tau \leq t < \infty$ үшін

$$d^{-1}\big(K(t,\tau) + U(t)Q(\tau,s)\big) \leq K(t,s) \leq d\big(K(t,\tau) + U(t)Q(\tau,s)\big) \qquad (5)$$

орындалса.

Анықтама 2. $K(\cdot,\cdot)$ теріс емес өлшемді функциясы $\{(t,s), 0 < s \leq t < \infty\}$ жиынында анықталған және бірінші айнымалы бойынша кемімейтін болсын. $K(t,s)$ функциясы O_1^+ класына жатады дейміз, егер $V(\cdot)$ және $R(\cdot,\cdot)$ теріс емес өлшемді функциялары және $d \geq 1$ тұрақтысы табылып, барлық $t, \tau, s: 0 < s \leq \tau \leq t < \infty$ үшін

$$d^{-1}\big(R(t,\tau)V(s) + K(\tau,s)\big) \leq K(t,s) \leq d\big(R(t,\tau)V(s) + K(\tau,s)\big) \qquad (6)$$

орындалса.

[17–21] жұмыстарында (3) және (4) теңсіздіктерінің орындалу критерийлері $K(\cdot,\cdot) \geq 0$ өзектері (2) шартты қанағаттандырған кезде табылған. Бұл жұмыстарда дәлелдеу кезінде u салмақ функциясына тәуелді көмекші функцияны қолданады. [22] жұмысында бұл мәселе басқаша әдіспен зерттелген, яғни параметрдің $1 < r < q < \infty$ және $1 < p < \infty$ жағдайлары үшін көмекші функциясыз айқын критерийлер алынған. Ал [23] мақалада параметрдің $0 < r < \infty, 0 < q < p < \infty$ жағдайында

$$\left(\int_0^\infty u(x) \left(\int_0^x \left(\int_t^\infty K(s,t) f(s) ds \right)^r w(t) dt \right)^{\frac{q}{r}} dx \right)^{\frac{1}{q}} \leq C \left(\int_0^\infty v(x) f^p(x) dx \right)^{\frac{1}{p}},$$

$$\left(\int_0^\infty u(x)\left(\int_x^\infty\left(\int_0^t K(t,s)f(s)ds\right)^r w(t)dt\right)^{\frac{q}{r}}dx\right)^{\frac{1}{q}} \le C\left(\int_0^\infty v(x)f^p(x)dx\right)^{\frac{1}{p}}$$

түріндегі теңсіздіктердің орындалу шарттары өзектері O_1^- және O_1^+ классына жатқан кезде алынған. Жұмыстың мақсаты [21] жұмыста қолданған әдіспен (5) немесе (6) шарттарды қанағаттандыратын $K(\cdot,\cdot)\ge 0$ өзектері бар (3) және (4) квазисызықты теңсіздіктерді параметрдің $0<r<\infty, 1<p\le q<\infty$ жағдайында сипаттау.

Зерттеу әдіснамасы

Негізгі теоремаларды дәлелдеу үшін алдын ала белгілі екі салмақты Харди типті теңсіздіктердің тұжырымдары мен нәтижелері, матрицаның қасиеттері, Минковский теңсіздігі және басқа да классикалық теңсіздіктер қолданылады.

$L_{p,v}(I)$ арқылы I аралығында анықталған барлық f өлшемді функциялардан тұратын және

$$\|f\|_{p,v,I} \equiv \|f\|_{p,v} = \begin{cases}\left(\int_I v(t)|f(t)|^p dt\right)^{\frac{1}{p}}, & 0<p<\infty,\\ \operatorname*{esssup}_{t\in I} v(t)|f(t)|, & p=\infty\end{cases}$$

функционал ақырлы болатын кеңістікті белгілейік. [21] жұмыста келтірілген бүркеуші тізбектің анықтамасын берейік:

Анықтама 3. $\{x_k\}\subset I$ өспелі тізбек I аралығын бүркеуші тізбек деп аталады, егер $k\to-\infty$ кезде $x_k\to 0$ және $k\to\infty$ кезде $x_k\to\infty$ үшін $\bigcup_k[x_k,x_{k+1})=I$ немесе $\{x_k\}=\{x_k\}_{k=n}^m$ және $x_{n-1}=0, x_{m+1}=\infty$.

Сонымен қатар, төмендегі Харди типті операторлардың нормалары қажет болады:

$$H_{r,p}^-(\alpha,\beta)=\sup_{f\ge 0}\frac{\left(\int_\alpha^\beta\left(\int_t^\beta K(s,t)f(s)ds\right)^r w(t)dt\right)^{\frac{1}{r}}}{\|f\|_{p,v,(\alpha,\beta)}},$$

$$H_{r,p}^+(\alpha,\beta)=\sup_{f\ge 0}\frac{\left(\int_\alpha^\beta\left(\int_\alpha^t K(t,s)f(s)ds\right)^r w(t)dt\right)^{\frac{1}{r}}}{\|f\|_{p,v,(\alpha,\beta)}}$$

мұндағы $0\le\alpha<\beta\le\infty$.

Бағалау процесінде келесі элементар теңсіздікті қолданылады:

Егер $a_i>0, i=1,2,\dots,k$, онда

$$\sum_{i=1}^k a_i^\alpha \le\left(\sum_{i=1}^k a_i\right)^\alpha, \quad \alpha\ge 1. \tag{7}$$

Келісім: $E\ll F$ белгісі p,q және r параметрлеріне тәуелді C тұрақтысымен $E\le CF$ білдіреді. Сонымен қатар, $E\approx F$ дегеніміз $E\ll F\ll E$ екіжақты бағалауы.

Ескерту: Анықтама 1 және Анықтама 2-дегі $Q(\cdot,\cdot)$ және $R(\cdot,\cdot)$ функциялары (2) шартты қанағаттандырады, яғни екінші айнымалысы бойынша өспейтін және бірінші айнымалысы бойынша кемімейтін функциялар.

Зерттеу нәтижелері

Келесі шамаларды енгізейік:

$$A_1^+ = \sup_{z>0}\left(\int_0^z u(x)dx\right)^{\frac{1}{q}} H_{r,p}^-(z,\infty),$$

$$A_2^+ = \sup_{z>0}\left(\int_0^z u(x)\left(\int_x^z w(t)dt\right)^{\frac{q}{r}}dx\right)^{\frac{1}{q}}\left(\int_z^\infty K^{p'}(s,z)v^{1-p'}(s)ds\right)^{\frac{1}{p'}},$$

$$A_3^+ = \sup_{z>0}\left(\int_0^z u(x)\left(\int_x^z Q^r(z,t)w(t)dt\right)^{\frac{q}{r}}dx\right)^{\frac{1}{q}}\left(\int_z^\infty U^{p'}(s)v^{1-p'}(s)ds\right)^{\frac{1}{p'}},$$

$$A_4^+ = \sup_{z>0}\left(\int_0^z u(x)\left(\int_x^z V^r(t)w(t)dt\right)^{\frac{q}{r}}dx\right)^{\frac{1}{q}}\left(\int_z^\infty R^{p'}(s,z)v^{1-p'}(s)ds\right)^{\frac{1}{p'}},$$

$$A_5^+ = \sup_{z>0}\left(\int_0^z u(x)\left(\int_x^z K^r(z,t)w(t)dt\right)^{\frac{q}{r}}dx\right)^{\frac{1}{q}}\left(\int_z^\infty v^{1-p'}(s)ds\right)^{\frac{1}{p'}}.$$

Теорема 1. Айталық $0 < r < \infty, 1 < p \le q < \infty$ және $K(\cdot,\cdot)$ өзегі O_1^- класына жатсын. Онда (4) салмақты теңсіздігі орындалады сонда тек сонда, егер $A^+ = \max\{A_1^+, A_2^+, A_3^+\} < \infty$. Сонымен қатар, $A^+ \approx C$, мұндағы C - (4) теңсіздігінің оптималды тұрақтысы.

Теорема 2. Айталық $0 < r < \infty, 1 < p \le q < \infty$ және $K(\cdot,\cdot)$ өзегі O_1^+ класына жатсын. Онда (4) салмақты теңсіздігі орындалады сонда тек сонда, егер $\tilde{A}^+ = \max\{A_1^+, A_4^+, A_5^+\} < \infty$. Сонымен қатар, $\tilde{A}^+ \approx C$, мұндағы C - (4) теңсіздігінің оптималды тұрақтысы.

Теорема 1 тұжырымының дәлелдеуі. Жеткіліктілігі. Айталық $A^+ < \infty$ болсын. (4) салмақты теңсіздік орындалатынын көрсетеміз. $\int_0^\infty f(s)ds < \infty$ болатындай $0 \le f$ болсын.

$$\int_{x_k}^\infty\left(\int_t^\infty K(s,t)f(s)ds\right)^r w(t)dt = 2^{-r(k-1)}$$

орындалатындай $\{x_k\}$ – бүркеуші тізбек болсын. (4) теңсіздігінің сол жағын J деп белгілеп, оны келесідей бағалаймыз

$$J = \left(\sum_k \int_{x_k}^{x_{k+1}} u(x)\left(\int_x^\infty\left(\int_t^\infty K(s,t)f(s)ds\right)^r w(t)dt\right)^{\frac{q}{r}}dx\right)^{\frac{1}{q}}$$

$$\le \left(\sum_k 2^{-q(k-1)}\int_{x_k}^{x_{k+1}} u(x)dx\right)^{\frac{1}{q}} \le 4^q\left(\sum_k 2^{-q(k+1)}\int_{x_k}^{x_{k+1}} u(x)dx\right)^{\frac{1}{q}}. \qquad (8)$$

Әрі қарай r параметрін екі жағдайға бөліп, келесі өрнектерді қарастырайық:

$1 \le r < \infty$ үшін

$$2^{-(k+1)} = 2^{-k} - 2^{-(k+1)}$$
$$= \left(\int_{x_{k+1}}^{\infty}\left(\int_{t}^{\infty} K(s,t)f(s)ds\right)^{r} w(t)dt\right)^{\frac{1}{r}} - \left(\int_{x_{k+2}}^{\infty}\left(\int_{t}^{\infty} K(s,t)f(s)ds\right)^{r} w(t)dt\right)^{\frac{1}{r}}$$
$$\leq \left(\int_{x_{k+1}}^{x_{k+2}}\left(\int_{t}^{\infty} K(s,t)f(s)ds\right)^{r} w(t)dt\right)^{\frac{1}{r}} \leq \left(\int_{x_{k+1}}^{x_{k+2}}\left(\int_{t}^{x_{k+2}} K(s,t)f(s)ds\right)^{r} w(t)dt\right)^{\frac{1}{r}}$$
$$+\left(\int_{x_{k+1}}^{x_{k+2}}\left(\int_{x_{k+2}}^{\infty} K(s,t)f(s)ds\right)^{r} w(t)dt\right)^{\frac{1}{r}}. \quad (9)$$

$0 < r < 1$ кезде

$$2^{-r(k+1)} = \frac{\left(2^{-rk} - 2^{-r(k+1)}\right)}{2^{r} - 1}$$
$$= \frac{1}{2^{r}-1}\left(\int_{x_{k+1}}^{\infty}\left(\int_{t}^{\infty} K(s,t)f(s)ds\right)^{r} w(t)dt - \int_{x_{k+2}}^{\infty}\left(\int_{t}^{\infty} K(s,t)f(s)ds\right)^{r} w(t)dt\right)$$
$$= \frac{1}{2^{r}-1}\left(\int_{x_{k+1}}^{x_{k+2}}\left(\int_{t}^{\infty} K(s,t)f(s)ds\right)^{r} w(t)dt\right)$$
$$\ll \int_{x_{k+1}}^{x_{k+2}}\left(\int_{t}^{x_{k+2}} K(s,t)f(s)ds\right)^{r} w(t)dt + \int_{x_{k+1}}^{x_{k+2}}\left(\int_{x_{k+2}}^{\infty} K(s,t)f(s)ds\right)^{r} w(t)dt. \quad (10)$$

(8) теңсіздігіне (9) және (10) өрнектерін қойып, келесіні аламыз:

$$J \ll \left(\sum_{k}\int_{x_k}^{x_{k+1}} u(x)\left(\int_{x_{k+1}}^{x_{k+2}}\left(\int_{t}^{x_{k+2}} K(s,t)f(s)ds\right)^{r} w(t)dt\right)^{\frac{q}{r}} dx\right)^{\frac{1}{q}}$$
$$+\left(\sum_{k}\int_{x_k}^{x_{k+1}} u(x)\left(\int_{x_{k+1}}^{x_{k+2}}\left(\int_{x_{k+2}}^{\infty} K(s,t)f(s)ds\right)^{r} w(t)dt\right)^{\frac{q}{r}} dx\right)^{\frac{1}{q}} = J_1 + J_2. \quad (11)$$

J_1 бағалау үшін $H_{r,p}^{-}(x_{k+1}, x_{k+2})$ шамасының анықтамасын қолданамыз және $\frac{q}{p} \geq 1$ шартын ескере отырып, (7) теңсіздігі бойынша келесі қорытындыға келеміз:

$$J_1 = \left(\sum_k \int_{x_k}^{x_{k+1}} u(x) \left(\int_{x_{k+1}}^{x_{k+2}} \left(\int_t^{x_{k+2}} K(s,t) f(s) ds \right)^r w(t) dt \right)^{\frac{q}{r}} dx \right)^{\frac{1}{q}}$$

$$\leq \left(\sum_k \int_{x_k}^{x_{k+1}} u(x) dx \left(J_{r,p}^-(x_{k+1}, x_{k+2}) \right)^q \|f\|_{p,v,(x_{k+1},x_{k+2})}^q \right)^{\frac{1}{q}}$$

$$\leq \sup_k \left(\int_0^{x_{k+1}} u(x) dx \right)^{\frac{1}{q}} J_{r,p}^-(x_{k+1}, \infty) \left(\sum_k \left(\|f\|_{p,v,(x_{k+1},x_{k+2})}^p \right)^{\frac{q}{p}} \right)^{\frac{1}{q}}$$

$$\leq \sup_{z>0} \left(\int_0^z u(x) dx \right)^{\frac{1}{q}} H_{r,p}^-(\mathrm{z}, \infty) \left(\sum_k \|f\|_{p,v,(x_{k+1},x_{k+2})}^p \right)^{\frac{1}{p}} \leq A_1^+ \|f\|_{p,v}. \quad (12)$$

Ал J_2 бағалау үшін өзекке қойылған шартты қолданып, төмендегідей екі шама аламыз:

$$J_2 \leq \left(\sum_k \int_{x_k}^{x_{k+1}} u(x) dx \left(\int_{x_{k+1}}^{x_{k+2}} w(t) dt \left(\int_{x_{k+2}}^{\infty} K(s, x_{k+2}) f(s) ds \right)^r \right.\right.$$

$$\left.\left. + \int_{x_{k+1}}^{x_{k+2}} Q^r(x_{k+2}, \mathrm{t}) w(t) dt \left(\int_{x_{k+2}}^{\infty} U(s) f(s) ds \right)^r \right)^{\frac{q}{r}} \right)^{\frac{1}{q}}$$

$$\ll \left(\sum_k \int_{x_k}^{x_{k+1}} u(x) dx \left(\int_{x_{k+1}}^{x_{k+2}} w(t) dt \right)^{\frac{q}{r}} \left(\int_{x_{k+2}}^{\infty} K(s, x_{k+2}) f(s) ds \right)^q \right)^{\frac{1}{q}}$$

$$+ \left(\sum_k \int_{x_k}^{x_{k+1}} u(x) dx \left(\int_{x_{k+1}}^{x_{k+2}} Q^r(x_{k+2}, t) w(t) dt \right)^{\frac{q}{r}} \left(\int_{x_{k+2}}^{\infty} U(s) f(s) ds \right)^q \right)^{\frac{1}{q}} = J_{21} + J_{22}. \quad (13)$$

J_{21} келесі түрде жазайық:

$$J_{21} = \left(\int_0^{\infty} \left(\int_x^{\infty} K(s,x) f(s) ds \right)^q d\mu(x) \right)^{\frac{1}{q}},$$

бұл жерде

$$d\mu(x) = \sum_k \int_{x_k}^{x_{k+1}} u(y) \left(\int_{x_{k+1}}^{x_{k+2}} w(t) dt \right)^{\frac{q}{r}} \delta(x - x_{k+2}) dy$$

және $\delta(\cdot)$ - дельта-функция. J_{21} -дегі өзек (5) шартын қанағаттандырады сондықтан [7] жұмыста келтірілген өзегі бар интегралдық оператордың шенелімділік критерийін қолданып келесі теңсіздікті аламыз:

$$J_{21} \leq \max\{A_{21}, A_{22}\} \|f\|_{p,v}, \quad (14)$$

мұндағы A_{21}, A_{22} шамалары келесідей анықталған:

$$A_{21} = \sup_{z>0} \left(\int_0^z d\mu(x) \right)^{\frac{1}{q}} \left(\int_z^\infty K^{p'}(s,z) v^{1-p'}(s) ds \right)^{\frac{1}{p'}},$$

$$A_{22} = \sup_{z>0} \left(\int_0^z Q^q(z,x) d\mu(x) \right)^{\frac{1}{q}} \left(\int_z^\infty U^{p'}(s) v^{1-p'}(s) ds \right)^{\frac{1}{p'}}.$$

Келесі интегралды қарастырайық:

$$\int_0^z d\mu(x) = \sum_{x_{k+2} \le z} \int_{x_k}^{x_{k+1}} u(x) \left(\int_{x_{k+1}}^{x_{k+2}} w(t) dt \right)^{\frac{q}{r}} dx$$

$$\le \sum_{x_{k+2} \le z} \int_{x_k}^{x_{k+1}} u(x) \left(\int_x^z w(t) dt \right)^{\frac{q}{r}} dx \le \int_0^z u(x) \left(\int_x^z w(t) dt \right)^{\frac{q}{r}} dx,$$

бұл

$$A_{21} \le A_2^+. \tag{15}$$

дегенді береді. Енді келесі интегралды бағалайық:

$$\int_0^z Q^q(z,x) d\mu(x) = \sum_{x_{k+2} \le z} \int_{x_k}^{x_{k+1}} u(x) dx \left(\int_{x_{k+1}}^{x_{k+2}} w(t) dt \right)^{\frac{q}{r}} Q^q(z, x_{k+2})$$

$$\le \sum_{x_{k+2} \le z} \int_{x_k}^{x_{k+1}} u(x) dx \left(\int_{x_{k+1}}^{x_{k+2}} Q^r(z,t) w(t) dt \right)^{\frac{q}{r}}$$

$$\le \sum_{x_{k+2} \le z} \int_{x_k}^{x_{k+1}} u(x) \left(\int_x^z Q^r(z,t) w(t) dt \right)^{\frac{q}{r}} dx$$

$$\le \int_0^z u(x) \left(\int_x^z Q^r(z,t) w(t) dt \right)^{\frac{q}{r}} dx$$

осыдан

$$A_{22} \le A_3^+. \tag{16}$$

(14), (15) және (16) теңсіздіктер негізінде

$$J_{21} \le \max\{A_2^+, A_3^+\} \|f\|_{p,v} \tag{17}$$

шығады. Осы жолмен J_{22} шамасын бағалайық:

$$J_{22} = \left(\int_0^\infty \left(\int_x^\infty U(s) f(s) ds \right)^q d\tau(x) \right)^{\frac{1}{q}},$$

мұндағы

$$d\tau(x) = \sum_k \int_{x_k}^{x_{k+1}} u(y) \left(\int_{x_{k+1}}^{x_{k+2}} Q^r(x_{k+2}, \mathrm{t}) w(t) dt \right)^{\frac{q}{r}} \delta(x - x_{k+2}) dy$$

[4] дереккөзінде келтірілген Харди теңсіздігіне қатысты нәтижелерді қолдана отырып, келесі қорытындыға келеміз:

$$J_{22} \le \sup_{z>0}\left(\int_0^z d\tau(x)\right)^{\frac{1}{q}}\left(\int_z^\infty U^{p'}(s)v^{1-p'}(s)ds\right)^{\frac{1}{p'}}\|f\|_{p,v}.$$

Бірінші интеграл келесідей бағаланады:

$$\int_0^z d\tau(x) = \sum_{x_{k+2}\le z}\int_{x_k}^{x_{k+1}} u(x)dx\left(\int_{x_{k+1}}^{x_{k+2}} Q^r(x_{k+2},t)w(t)dt\right)^{\frac{q}{r}}$$
$$\le \sum_{x_{k+2}\le z}\int_{x_k}^{x_{k+1}} u(x)\left(\int_x^z Q^r(z,t)w(t)dt\right)^{\frac{q}{r}}dx$$
$$\le \int_0^z u(x)\left(\int_x^z Q^r(z,t)(t)dt\right)^{\frac{q}{r}}dx$$

және

$$J_{22} \le A_3^+\|f\|_{p,v}$$

болады. Осы және (17) бойынша $J_2 \ll \max\{A_2^+, A_3^+\}\|f\|_{p,v}$ болатыны шығады. Осылайша, (12) және (11) қатынастары барлық f теріс емес функция үшін $J \ll A^+\|f\|_{p,v}$ екенін береді, демек (4) теңсіздігі

$$C \ll A^+ \tag{18}$$

бағалауымен орындалады, мұндағы $C > 0$ - (4) теңсіздік орындалатындай оптималды тұрақты.

Қажеттілігі. (4) теңсіздігі оптималды C оң тұрақтысымен орындалсын. Ендігі кезекте $A^+ < \infty$ болатынын анықтаймыз. (4) теңсіздігінен $\forall z \in I$ үшін

$$C\|f\|_{p,v,(z,\infty)} \ge \left(\int_0^z u(x)dx\right)^{\frac{1}{q}}\left(\int_z^\infty\left(\int_t^\infty K(s,t)f(s)ds\right)^r w(t)dt\right)^{\frac{1}{r}}$$

аламыз. Онда

$$C \ge \left(\int_0^z u(x)dx\right)^{\frac{1}{q}}\sup_{f\ge 0}\frac{\left(\int_z^\infty\left(\int_t^\infty K(s,t)f(s)ds\right)^r w(t)dt\right)^{\frac{1}{r}}}{\|f\|_{p,v,(z,\infty)}} = \left(\int_0^z u(x)dx\right)^{\frac{1}{q}} H_{r,p}^-(z,\infty)$$

Соңғы теңсіздіктің екі жағынан z бойынша супремум алсақ, онда

$$C \ge A_1^+. \tag{19}$$

Айталық $f \ge 0$ және $z > 0$ үшін $f_z(\cdot) = \chi_{(z,\infty)}(\cdot)f(\cdot)$ болсын. Бұл жерде және өзге жерлерде $\chi_{(\alpha,\beta)}(\cdot)$ белгісі (α,β) интервалының сипаттамалық функциясын білдіреді. f_z функциясы үшін (4) теңсіздігінен келесі өрнекке келеміз:

$$C\|f_z\|_{p,v} \ge \left(\int_0^\infty u(x)\left(\int_x^\infty\left(\int_t^\infty K(s,t)f(s)ds\right)^r w(t)dt\right)^{\frac{q}{r}}dx\right)^{\frac{1}{q}}$$
$$\ge\left(\int_0^z u(x)\left(\int_x^z\left(\int_z^\infty K(s,t)f_z(s)ds\right)^r w(t)dt\right)^{\frac{q}{r}}dx\right)^{\frac{1}{q}}. \tag{20}$$

$f_z(\cdot) = \chi_{(z,\infty)}(\cdot)K^{p'-1}(\cdot,z)v^{1-p'}(\cdot)$ тексеру функциясын (20) теңсіздікке қойсақ, онда

$$C\left(\int_z^\infty K^{p'}(x,z)v^{1-p'}(x)dx\right)^{\frac{1}{p}}$$

$$\geq\left(\int_0^z u(x)\left(\int_x^z\left(\int_z^\infty K(s,t)K^{p'-1}(s,z)v^{1-p'}(s)ds\right)^r w(t)dt\right)^{\frac{q}{r}}dx\right)^{\frac{1}{q}}.$$

Әрі қарай $z \geq t$ үшін $K(s,t) \geq K(s,z)$ өзекке қойылған шартты ескеріп төмендегіні аламыз:

$$C\left(\int_z^\infty K^{p'}(x,z)v^{1-p'}(x)dx\right)^{\frac{1}{p}} \geq \left(\int_0^z u(x)\left(\int_x^z w(t)dt\right)^{\frac{q}{r}}dx\right)^{\frac{1}{q}}\left(\int_z^\infty K^{p'}(s,z)v^{1-p'}(s)ds\right)$$

осы қорытынды негізінде

$$C \geq A_2^+. \tag{21}$$

Енді $f_z(\cdot) = \chi_{(z,\infty)}(\cdot)U^{p'-1}(\cdot)v^{1-p'}(\cdot)$ тексеру функциясын алайық. Осы функцияны (20)-ға қойып және $s \geq z$ үшін $K(s,t) \geq Q(z,t)U(s)$ өзек шартын ескерсек, онда

$$C\left(\int_{\mathrm{z}}^\infty U^{p'}(x)v^{1-p'}(x)dx\right)^{\frac{1}{p}}$$

$$\geq\left(\int_0^{\mathrm{z}} u(x)\left(\int_x^z\left(\int_{\mathrm{z}}^\infty K(s,t)U^{p'-1}(s)v^{1-p'}(s)ds\right)^r w(t)dt\right)^{\frac{q}{r}}dx\right)^{\frac{1}{q}}$$

$$\geq\left(\int_0^{\mathrm{z}} u(x)\left(\int_x^z Q^r(z,t)w(t)dt\right)^{\frac{q}{r}}dx\right)^{\frac{1}{q}}\left(\int_{\mathrm{z}}^\infty U^{p'}(s)v^{1-p'}(s)ds\right)$$

аламыз. Осылайша

$$C \geq A_3^+. \tag{22}$$

(19), (21) және (22) теңсіздіктерінен $A^+ < \infty$ шенелген және $C \geq A^+$ болатыны шығады. Бұл (18) өрнегімен $C \approx A^+$ береді. Теорема 1 тұжырымы дәлелденді.

Теорема 2 тұжырымының дәлелдеуі. Жеткіліктілігі. Бұл теореманың жеткіліктілігін дәлелдеу Теорема 1-дің жеткіліктілігі жолымен жүреді. Демек (11) теңсіздіктегі $J \ll J_1 + J_2$ бағалауын құрамыз. (12)-ге ұқсас түрде $J_1 \ll A_1^+\|f\|_{p,v}$ қорытындысына келеміз. Соңында J_2 интегралын бағалау үшін өзектің (6) қасиетін пайдаланамыз.

$$J_2 \ll \left(\sum_k \int_{x_k}^{x_{k+1}} u(x)dx\left(\int_{x_{k+1}}^{x_{k+2}} V^r(t)w(t)dt\right)^{\frac{q}{r}}\left(\int_{x_{k+2}}^\infty R(s,x_{k+2})f(s)ds\right)^q\right)^{\frac{1}{q}} +$$

$$\left(\sum_k \int_{x_k}^{x_{k+1}} u(x)dx\left(\int_{x_{k+1}}^{x_{k+2}} K^r(x_{k+2},t)w(t)dt\right)^{\frac{q}{r}}\left(\int_{x_{k+2}}^\infty f(s)ds\right)^q\right)^{\frac{1}{q}} = \tilde{J}_{21} + \tilde{J}_{22}. \tag{23}$$

Айталық

$$\tilde{J}_{21} = \left(\int_0^{\infty} \left(\int_x^{\infty} R(s,x) f(s) ds \right)^q d\vartheta(x) \right)^{\frac{1}{q}},$$

мұндағы

$$d\vartheta(x) = \sum_k \int_{x_k}^{x_{k+1}} u(y) \left(\int_{x_{k+1}}^{x_{k+2}} V^r(t) w(t) dt \right)^{\frac{q}{r}} \delta(x - x_{k+2}) dy.$$

(6) шартымен берілген өзегі бар интегралдық операторлың шенелімділік шарты, яғни [7] жұмысындағы Теорема 2-ге сүйене отырып, мынаны аламыз:

$$\tilde{J}_{21} \le \max\{\dot{A}_{21}, \dot{A}_{22}\} \|f\|_{p,v}, \tag{24}$$

мұндағы

$$\dot{A}_{21} = \sup_{z>0} \left(\int_0^{\mathrm{z}} d\vartheta(x) \right)^{\frac{1}{q}} \left(\int_{\mathrm{z}}^{\infty} R^{p'}(s,z) v^{1-p'}(s) ds \right)^{\frac{1}{p'}},$$

$$\dot{A}_{22} = \sup_{z>0} \left(\int_0^{\mathrm{z}} R^q(z,x) d\vartheta(x) \right)^{\frac{1}{q}} \left(\int_{\mathrm{z}}^{\infty} v^{1-p'}(s) ds \right)^{\frac{1}{p'}}.$$

Интеграл

$$\int_0^{\mathrm{z}} d\vartheta(x) = \sum_{x_{k+2} \le z} \int_{x_k}^{x_{k+1}} u(x) \left(\int_{x_{k+1}}^{x_{k+2}} V^r(t)\, w(t) dt \right)^{\frac{q}{r}} dx$$

$$\le \sum_{x_{k+2} \le z} \int_{x_k}^{x_{k+1}} u(x) \left(\int_x^z V^r(t) w(t) dt \right)^{\frac{q}{r}} dx \le \int_0^{\mathrm{z}} u(x) \left(\int_x^z V^r(t) w(t) dt \right)^{\frac{q}{r}} dx$$

болғандықтан

$$\dot{A}_{21} \le A_4^+. \tag{25}$$

Сонымен қатар,

$$\int_0^{\mathrm{z}} R^q(z,x) d\vartheta(x) = \sum_{x_{k+2} \le z} \int_{x_k}^{x_{k+1}} u(x) dx \left(\int_{x_{k+1}}^{x_{k+2}} V^r(t) w(t) dt \right)^{\frac{q}{r}} R^q(z, x_{k+2})$$

$$\le \int_0^{\mathrm{z}} u(x) \left(\int_x^z K^r(z,t) w(t) dt \right)^{\frac{q}{r}} dx$$

Болғандықтан

$$\dot{A}_{22} \le A_5^+. \tag{26}$$

(24), (25) және (26) шарттары

$$\tilde{J}_{21} \le \max\{A_4^+, A_5^+\} \|f\|_{p,v} \tag{27}$$

береді. Дәл осылай $\tilde{J}_{22}$ бағалаймыз.

$$\tilde{J}_{22} = \left(\int_0^{\infty} \left(\int_x^{\infty} f(s) ds \right)^q d\varphi(x) \right)^{\frac{1}{q}},$$

мұнда

$$d\varphi(x)=\sum_{k}\int_{x_k}^{x_{k+1}}u(y)\left(\int_{x_{k+1}}^{x_{k+2}}K^r(x_{k+2},\mathrm{t})w(t)dt\right)^{\frac{q}{r}}\delta(x-x_{k+2})dy.$$

Харди операторының шенелімділік шарты бойынша ([4] қараңыз)

$$\tilde{J}_{22}\ll\sup_{z>0}\left(\int_0^z d\varphi(x)\right)^{\frac{1}{q}}\left(\int_{\mathrm{z}}^{\infty}v^{1-p'}(s)ds\right)^{\frac{1}{p'}}\|f\|_{p,v}$$

қатынасына келеміз. Алдыңғы тәсілге сәйкес

$$\int_0^{\mathrm{z}}d\varphi(x)=\sum_{x_{k+2}\le z}\int_{x_k}^{x_{k+1}}u(x)dx\left(\int_{x_{k+1}}^{x_{k+2}}K^r(x_{k+2},\mathrm{t})w(t)dt\right)^{\frac{q}{r}}$$
$$\le\int_0^{\mathrm{z}}u(x)\left(\int_x^z K^r(\mathrm{z},\mathrm{t})w(t)dt\right)^{\frac{q}{r}}dx$$

болғандықтан

$$\tilde{J}_{22}\le A_5^+\|f\|_{p,v}.$$

Осы және (27) бағалауы бойынша $J_2\ll\max\{A_4^+,A_5^+\}\|f\|_{p,v}$ болады. Демек, (12) және (11) негізінде $J\ll\tilde{A}^+\|f\|_{p,v}$, яғни

$$C\ll\tilde{A}^+. \tag{28}$$

Қажеттілігі. Бұл теореманың қажеттілік бөлімінің дәлелі Теорема 1-дегідей басталады. Атап айтқанда, $A_1^+\le C$ болатыны (19)-ға ұқсас түрде шығарылады.

Кейін $f_z(\cdot)=\chi_{(z,\infty)}(\cdot)R^{p'-1}(\cdot,z)v^{1-p'}(\cdot)$ тексеру функциясын (20) -ға қойып, қажетті бағалауларды жүргіземіз

$$C\left(\int_{\mathrm{z}}^{\infty}R^{p'}(x,z)v^{1-p'}(x)dx\right)^{\frac{1}{p}}$$
$$\ge\left(\int_0^{\mathrm{z}}u(x)\left(\int_x^z\left(\int_{\mathrm{z}}^{\infty}K(s,t)R^{p'-1}(s,z)v^{1-p'}(s)ds\right)^r w(t)dt\right)^{\frac{q}{r}}dx\right)^{\frac{1}{q}}$$

((6) шарт бойынша $z\ge t$ үшін $K(s,t)\ge R(s,z)V(t)$)

$$\gg\left(\int_0^{\mathrm{z}}u(x)\left(\int_x^z V^r(t)w(t)dt\right)^{\frac{q}{r}}dx\right)^{\frac{1}{q}}\left(\int_{\mathrm{z}}^{\infty}R^{p'}(s,z)v^{1-p'}(s)ds\right),$$

демек,

$$C\ge A_4^+. \tag{29}$$

Ендігі кезекте $f_z(\cdot)=\chi_{(z,\infty)}(\cdot)v^{1-p'}(\cdot)$ деп алып, (20) -ға қойсақ, онда

$$C\left(\int_z^{\infty} v^{1-p'}(x)dx\right)^{\frac{1}{p}}$$

$$\geq\left(\int_0^z u(x)\left(\int_x^z\left(\int_z^{\infty} K(s,t)v^{1-p'}(s)ds\right)^r w(t)dt\right)^{\frac{q}{r}}dx\right)^{\frac{1}{q}}$$

((6) шарт бойынша $s\geq z$ үшін $K(s,t)\gg K(z,t)$)

$$\gg\left(\int_0^z u(x)\left(\int_x^z K^r(z,t)w(t)dt\right)^{\frac{q}{r}}dx\right)^{\frac{1}{q}}\left(\int_z^{\infty} v^{1-p'}(s)ds\right).$$

Осы қорытынды негізінде

$$C\geq A_5^+. \tag{30}$$

(19), (29) және (30) теңсіздіктеріне сүйене отырып, $\tilde{A}^+<\infty$ және $C\geq\tilde{A}^+$ болатынын көреміз. Бұл (28) қатынасымен бірге $C\approx\tilde{A}^+$ эквиваленттілігін береді. Осылайша, Теорема 2- нің дәлелі аяқталады.

Мынадай шамаларды енгізейік

$$A_1^-=\sup_{z>0}\left(\int_z^{\infty} u(x)dx\right)^{\frac{1}{q}}H_{r,p}^+(0,z),$$

$$A_2^-=\sup_{z>0}\left(\int_z^{\infty} u(x)\left(\int_z^x K^r(t,z)w(t)dt\right)^{\frac{q}{r}}dx\right)^{\frac{1}{q}}\left(\int_0^z v^{1-p'}(s)ds\right)^{\frac{1}{p'}},$$

$$A_3^-=\sup_{z>0}\left(\int_z^{\infty} u(x)\left(\int_z^x U^r(t)w(t)dt\right)^{\frac{q}{r}}dx\right)^{\frac{1}{q}}\left(\int_0^z Q^{p'}(z,s)v^{1-p'}(s)ds\right)^{\frac{1}{p'}},$$

$$A_4^-=\sup_{z>0}\left(\int_z^{\infty} u(x)\left(\int_z^x R^r(t,z)w(t)dt\right)^{\frac{q}{r}}dx\right)^{\frac{1}{q}}\left(\int_0^z V^{p'}(s)v^{1-p'}(s)ds\right)^{\frac{1}{p'}},$$

$$A_5^-=\sup_{z>0}\left(\int_z^{\infty} u(x)\left(\int_z^x w(t)dt\right)^{\frac{q}{r}}dx\right)^{\frac{1}{q}}\left(\int_0^z K^{p'}(z,s)v^{1-p'}(s)ds\right)^{\frac{1}{p'}}.$$

Теорема 3. Айталық $0<r<\infty, 1<p\leq q<\infty$ және $K(\cdot,\cdot)$ өзегі O_1^- классына жатсын. Онда (3) салмақты теңсіздігі орындалады сонда тек сонда, егер $A^-=\max\{A_1^-,A_2^-,A_3^-\}<\infty$. Сонымен қатар, $A^-\approx C$, мұндағы C - (3) теңсіздігінің оптималды тұрақтысы.

Теорема 4. Айталық $0<r<\infty, 1<p\leq q<\infty$ және $K(\cdot,\cdot)$ өзегі O_1^+ классына жатсын. Онда (3) салмақты теңсіздігі орындалады сонда тек сонда, егер $\tilde{A}^-=\max\{A_1^-,A_4^-,A_5^-\}<\infty$. Сонымен қатар, $\tilde{A}^-\approx C$, мұндағы C - (3) теңсіздігінің оптималды тұрақтысы.

Теорема 3 және Теорема 4 тұжырымдары Теорема 1 және Теорема 2 -де қолданылған әдістерге ұқсас түрде дәлелденеді.

Дискуссия

Әдетте Харди типті операторлар алдымен өзексіз жағдайда зерттеледі кейін белгілі бір шарттарды қанағаттандыратын өзекпен қарастырылады, яғни өзексіз жағдайға қарағанда кеңірек нәтижелер алынады. [21] жұмыста біз қарастырған (3) және (4) теңсіздіктерді (2) шартты қанағаттандыратын өзегі бар квазисызықты операторлар үшін қарастырған болатын. Айта кету керек, бұл жұмыста (3) және (4) теңсіздіктеріне қатысатын өзектер (2) шартынан әлсіз шарттарды қанағаттандырады. O_1^- және O_1^+ кластарының (2) шартын қанағаттандыратын өзектер классынан кең екенін көру оңай, өйткені (2) шарты (5) шартының $U(t) \equiv 1, Q(\tau, s) \equiv K(\tau, s)$ және (6) шартының $V(s) \equiv 1, R(t, \tau) \equiv K(t, \tau)$ жағдайындағы дербес жағдайы болып табылады, яғни (2) шартын қанағаттандыратын өзектер $O_1^- \cap O_1^+$ классына жатады. Бұл жұмыста [21] жұмыста басталған зерттеулерді жалпылай отырып, операторлар теориясына жаңа бағыт аштық. Әсіресе, өзектің жалпы кластары үшін алынған критерийлер тиімді қолданыс табады.

Қорытынды

Бұл жұмыста O_1^- және O_1^+ кластарына жататын өзекті операторды қамтитын (3) және (4) интегралдық теңсіздіктерінің орындалуының қажетті және жеткілікті шарттары алынды. Алынған критерийлер жаңа болып табылады. Болашақтағы зерттеулер p, q және r параметрлерінің басқа жағдайларында O_1^- және O_1^+ кластарына жататын өзегі бар квазисызықты интегралдық операторды қарастырумен байланысты болады.

Алғыс

Бұл мақала Қазақстан Республикасы Ғылым және жоғары білім министрлігінің қаржылық қолдауымен, AP23488579 гранттық зерттеу аясында дайындалды**.**

Пайдаланылған дереккөздер тізімі

[1] Bradley J.S. Hardy inequalities with mixed norms //Canad Math. Bull –1978.– Vol. 21–P. 405-408. https://doi.org/10.4153/CMB-1978-071-7

[2] Sinnamon G. Weighted Hardy and Opial-type inequalities //J.Math.Anal. Appl.–1991.–Vol. 160.–P. 434-445. https://doi.org/10.1016/0022-247X(91)90316-R

[3] Sinnamon G., Stepanov V.D. The weighted Hardy inequality: new proofs and the case $p=1$ //J. London Math. Soc.–1996.–Vol. 54.–P. 89-101. https://doi.org/10.1112/jlms/54.1.89

[4] Kufner A., Maligranda L., Persson L.-E. The Hardy inequality. About its history and some related results. Pilsen: Vydavatelský servis, 2007.

[5] Ойнаров Р. Двусторонние оценки нормы для некоторых классов интегральных операторов //Тр. Матиматического института РАН.–1993.–Т. 204.– С. 240-250.

https://www.mathnet.ru/tm1271

[6] Bloom S., Kerman P. Weighted norm inequlities for operators of Hardy type //Proc. Amer. Math. Soc.– 1991.–Vol. 113, №1.–P. 135-141. https://doi.org/10.1090/S0002-9939-1991-1059623-6

[7] Oinarov R. Boundedness and compactness of Volterra type integral operators //Siberian Math. J.– 2007.– Vol. 48, №5.– P. 884-896. https://doi.org/10.1007/s11202-007-0091-4

[8] Oinarov R. Boundedness and compactness in weighted Lebesgue spaces of integral operators with variable integration limits //Siberian Math. J.–2011.–Vol. 52, № 6.–P. 1042-1055. https://doi.org/10.1134/S0037446611060097

[9] Arendarenko L.S., Oinarov R., Persson L-E. On the boundedness of some classes of integral operators in weighted Lebesgue spaces //Eurasian Math. J.–2012.–Vol. 3, №1.–P. 5-17. https://www.mathnet.ru/eng/emj70

[10] Burenkov V.I., Oinarov R. Necessary and sufficient conditions for boundedness of the Hardytype operator from a weighted Lebesgue space to a Morrey-type space //Math. Inequal. Appl.–2013.–Vol. 16, № 1.– P. 1-19. dx.doi.org/10.7153/mia-16-01

[11] Kalybay A. On boundedness of the conjugate multidimensional Hardy operator from a Lebesgue space to a local Morrey-type space //Int. J. Math. Anal.–2014.–Vol. 8, № 11.– P. 539-553. http://dx.doi.org/10.12988/ijma.2014.423

[12] Gogatishvili A., Mustafayev R. Persson L-E. Some new iterated Hardy-type inequalities: the case $\theta = 1$ //J. Inequal. Appl.–2013.– Vol. 2013, №515.– P. 1-29. https://doi.org/10.1186/1029-242X-2013-515

[13] Prokhorov D.V., Stepanov V.D. On weighted Hardy inequalities in mixed norms // Proceedings of the Steklov Institute of Mathematics–2013.–Vol. 283.–P. 149-164. https://doi.org/10.1134/S0081543813080117

[14] Bernardis A.L., Salvador P.O. Some new iterated Hardy-type inequalities and applications //J. Math. Ineq.– 2017.– Vol. 11, №2.–P. 577-594. doi:10.7153/jmi-11-47

[15] Stepanov V.D., Shambilova G.E. On weighted iterated Hardy-type operators //Analysis Math.–2018.–Vol. 44, №2.– P. 273-283. https://doi.org/10.1007/s10476-018-0211-3

[16] Krepela M. Integral conditions for Hardy-type operators involving suprema //Collectanea Mathematica. – 2017. – Vol. 68. – P. 21-50. https://doi.org/10.1007/s13348-016-0170-6

[17] Prokhorov D.V. On the boundedness of a class of sublinear operators //Doklady Mathematics –2015.–Vol.92.–P. 602-605. https://doi.org/10.1134/S1064562415050294

[18] Prokhorov D.V. On a Class of Weighted Inequalities Containing Quasilinear Operators //Proceedings of the Steklov Institute of Mathematics–2016.–Vol. 293.–P. 272-287. https://doi.org/10.1134/S0081543816040192

[19] Prokhorov D.V., Stepanov V.D. Weighted estimates for a class of sublinear operators //Doklady Math.–2013.– Vol. 88, №3.–P. 721-723. https://doi.org/10.1134/S1064562413060264

[20] Prokhorov D.V., Stepanov V.D. Weighted inequalities for quasilinear integral operators on the semi-axis and applications to Lorentz spaces //Sbornik: Mathematics–2016.–Vol.207, №8.–P. 1159-1186. https://doi.org/10.1070/SM8535

[21] Kalybay A. Weighted estimates for a class of quasilinear integral operators //Siberian Math. J.–2019 –Vol. 60, №2.–P. 291-303. https://doi.org/10.1134/S0037446619020095

[22] Stepanov V.D., Shambilova G.E. On iterated and bilinear integral Hardy-type operators //Math. Inequal. Appl.–2019.–Vol. 22, №4.–P. 1505-1533. dx.doi.org/10.7153/mia-2019-22-105

[23] Kalybay A., Oinarov R. Weighted estimates containing quasilinear operators //Springer Proceedings in Mathematics & Statistics–2021.–Vol. 357.–P. 297-318.
https://doi.org/10.1007/978-3-030-77493-6_17

References

[1] Bradley J.S. Hardy inequalities with mixed norms //Canad Math. Bull.–1978.–Vol. 21.–P. 405-408. https://doi.org/10.4153/CMB-1978-071-7

[2] Sinnamon G. Weighted Hardy and Opial-type inequalities //J.Math.Anal. Appl.–1991.– Vol. 160.–P. 434-445. https://doi.org/10.1016/0022-247X(91)90316-R

[3] Sinnamon G., Stepanov V.D. The weighted Hardy inequality: new proofs and the case $p = 1$ //J. London Math. Soc.–1996.–Vol. 54.–P. 89-101. https://doi.org/10.1112/jlms/54.1.89

[4] Kufner A., Maligranda L., Persson L.-E. The Hardy inequality. About its history and some related results. Pilsen: Vydavatelský servis, 2007.

[5] Oinarov R. Dvustoronnie ocenki normy dlja nekotoryh klassov integral'nyh operatorov //Tr. Matimaticheskogo instituta RAN.–1993.–Vol. 204.–P. 240-250. https://www.mathnet.ru/tm1271 (in Russian)

[6] Bloom S., Kerman P. Weighted norm inequlities for operators of Hardy type //Proc. Amer. Math. Soc.–1991.–Vol. 113, №1.–P. 135-141. https://doi.org/10.1090/S0002-9939-1991-1059623-6

[7] Oinarov R. Boundedness and compactness of Volterra type integral operators //Siberian Math. J.–2007.– Vol. 48, №5.–P. 884-896. https://doi.org/10.1007/s11202-007-0091-4

[8] Oinarov R. Boundedness and compactness in weighted Lebesgue spaces of integral operators with variable integration limits //Siberian Math. J.–2011.–Vol. 52, № 6.–P. 1042-1055. https://doi.org/10.1134/S0037446611060097

[9] Arendarenko L.S., Oinarov R., Persson L-E. On the boundedness of some classes of integral operators in weighted Lebesgue spaces //Eurasian Math. J.–2012.–Vol. 3, №1.–P. 5-17. https://www.mathnet.ru/eng/emj70

[10] Burenkov V.I., Oinarov R. Necessary and sufficient conditions for boundedness of the Hardytype operator from a weighted Lebesgue space to a Morrey-type space //Math. Inequal. Appl.–2013.–Vol. 16, № 1.–P. 1-19. dx.doi.org/10.7153/mia-16-01

[11] Kalybay A. On boundedness of the conjugate multidimensional Hardy operator from a Lebesgue space to a local Morrey-type space //Int. J. Math. Anal.–2014.–Vol. 8, № 11.–P. 539-553. http://dx.doi.org/10.12988/ijma.2014.423

[12] Gogatishvili A., Mustafayev R. Persson L-E. Some new iterated Hardy-type inequalities: the case $\theta = 1$ //J. Inequal. Appl.–2013.–Vol. 2013, №515.–P. 1-29. https://doi.org/10.1186/1029-242X-2013-515

[13] Prokhorov D.V., Stepanov V.D. On weighted Hardy inequalities in mixed norms //Proceedings of the Steklov Institute of Mathematics–2013.–Vol. 283.–P. 149-164. https://doi.org/10.1134/S0081543813080117

[14] Bernardis A.L., Salvador P.O. Some new iterated Hardy-type inequalities and applications //J. Math. Ineq.–2017.–Vol. 11, №2.– P. 577-594. doi:10.7153/jmi-11-47

[15] Stepanov V.D., Shambilova G.E. On weighted iterated Hardy-type operators //Analysis Math.–2018.–Vol. 44, №2.–P. 273-283. https://doi.org/10.1007/s10476-018-0211-3

[16] Krepela M. Integral conditions for Hardy-type operators involving suprema //Collectanea Mathematica.–2017.–Vol. 68.–P. 21-50. https://doi.org/10.1007/s13348-016-0170-6

[17] Prokhorov D.V. On the boundedness of a class of sublinear operators //Doklady Mathematics –2015–Vol.92.–P. 602-605. https://doi.org/10.1134/S1064562415050294

[18] Prokhorov D.V. On a Class of Weighted Inequalities Containing Quasilinear Operators //Proceedings of the Steklov Institute of Mathematics–2016–Vol. 293.–P. 272-287. https://doi.org/10.1134/S0081543816040192

[19] Prokhorov D.V., Stepanov V.D. Weighted estimates for a class of sublinear operators //Doklady Math.–2013.–Vol. 88, №3.–P. 721-723. https://doi.org/10.1134/S1064562413060264

[20] Prokhorov D.V., Stepanov V.D. Weighted inequalities for quasilinear integral operators on the semi-axis and applications to Lorentz spaces //Sbornik: Mathematics–2016.–Vol.207, №8.–P. 1159-1186. https://doi.org/10.1070/SM8535

[21] Kalybay A. Weighted estimates for a class of quasilinear integral operators //Siberian Math. J.–2019–Vol. 60, №2.–P. 291-303. https://doi.org/10.1134/S0037446619020095

[22] Stepanov V.D., Shambilova G.E. On iterated and bilinear integral Hardy-type operators //Math. Inequal. Appl.–2019.–Vol. 22, №4.–P. 1505-1533. dx.doi.org/10.7153/mia-2019-22-105

[23] Kalybay A., Oinarov R. Weighted estimates containing quasilinear operators //Springer Proceedings in Mathematics & Statistics–2021.–Vol. 357.–P. 297-318. https://doi.org/10.1007/978-3-030-77493-6_17

МРНТИ 27.25.19

https://doi.org/10.51889/2959-5894.2026.95.3.005

А.Е. Жетписбаева[1], А.А. Джумабаева[1*]

[1]Евразийский национальный университет имени Л.Н. Гумилева, г. Астана, Казахстан
[*]*e-mail: akniet-1978@mail.ru*

НЕРАВЕНСТВО ТИПА УЛЬЯНОВА ДЛЯ СМЕШАННЫХ МОДУЛЕЙ ГЛАДКОСТИ ОБОБЩЁННЫХ ПРОИЗВОДНЫХ ЛИУВИЛЛЯ-ВЕЙЛЯ

Аннотация

В статье доказывается неравенство типа Ульянова для смешанного модуля гладкости обобщённой производной Лиувилля-Вейля функции двух переменных в смешанных метриках. Производная определяется преобразованием двумерного ряда Фурье с помощью положительной обобщённо монотонной последовательности класса GM^2. Основная оценка связывает модуль гладкости производной в метрике $L_{q_1q_2}$ с интегральным оператором от dyadic-функционала, построенного по угловым наилучшим приближениям исходной функции и разностям множителя. Доказательство основано на смешанном неравенстве Ульянова, обратной оценке модуля через угловые приближения и двумерном преобразовании Абеля для GM^2-множителей. Рассматриваются также частные случаи обычных производных Вейля и логарифмически модифицированных множителей.

Ключевые слова: смешанный модуль гладкости, смешанная метрика, угловое приближение, неравенство Ульянова, производная Лиувилля-Вейля, GM^2-множитель, ряд Фурье.

А.Е.Жетписбаева[1], А.А.Джумабаева[2*]
Л.Н. Гумилев атындағы Еуразия Ұлттық Университеті, Астана, Қазақстан[1,2]

ЛИУВИЛЛЬ-ВЕЙЛЬДІҢ ЖАЛПЫЛАНҒАН ТУЫНДЫЛАРЫНЫҢ АРАЛАС ТЕГІСТІК МОДУЛЬДЕРІ ҮШІН УЛЬЯНОВ ТИПТІ ТЕҢСІЗДІК

Аңдатпа

Мақалада екі айнымалы функцияның Лиувилль-Вейль мағынасындағы жалпыланған туындысының аралас тегістік модулі үшін Ульянов типті теңсіздік дәлелденеді. Жалпыланған туынды екі өлшемді Фурье қатарының коэффициенттерін GM^2 класындағы оң жалпыланған монотонды тізбекпен арқылы анықталады. Негізгі нәтиже туындының $L_{q_1q_2}$метрикасындағы аралас модулін бастапқы функцияның бұрыштық ең жақсы жуықтаулары және көбейткіштің айырымдары арқылы құрылған dyadic-функционалдың интегралдық өрнегімен бағалайды. Дәлелдеу аралас Ульянов теңсіздігіне, модульді бұрыштық жуықтаулар арқылы бағалауға және GM^2-көбейткіштер үшін екі өлшемді Абель түрлендіруіне негізделеді. Вейль туындылары және логарифмдік көбейткіштер мысал ретінде қарастырылады.

Түйін сөздер: аралас тегістік модулі, аралас метрика, бұрыштық жуықтау, Ульянов теңсіздігі, Лиувилль-Вейль туындысы, GM^2-көбейткіш, Фурье қатары.

A.E. Zhetpisbayeva [1], A.A. Jumabayeva [2*]
Eurasian National University named after L.N. Gumilyov, Astana, Kazakhstan[1,2]

ULYANOV-TYPE INEQUALITY FOR MIXED MODULI OF SMOOTHNESS OF GENERALIZED LIOUVILLE-WEYL DERIVATIVES

Abstract

The paper proves an Ulyanov-type inequality for the mixed modulus of smoothness of a generalized Liouville-Weyl derivative of a function of two variables in mixed metrics. The derivative is defined by transforming the two-dimensional Fourier series by a positive general monotone sequence of the class GM^2 and by phase parameters. The main estimate relates the mixed modulus of the derivative in the metric $L_{q_1q_2}$ to an integral operator involving a dyadic functional constructed from angular best approximations of the original function and differences of the multiplier. The proof combines the mixed Ulyanov inequality, an inverse estimate of the modulus through angular approximations, and a two-dimensional Abel transformation for

GM^2multipliers. Particular cases include the classical Weyl derivatives and logarithmically modified multipliers.

Keywords: mixed modulus of smoothness, mixed metric, angular approximation, Ulyanov inequality, Liouville-Weyl derivative, GM^2 multiplier, Fourier series.

Введение

Неравенства типа Ульянова занимают важное место в теории приближений и гармоническом анализе. В классической форме они оценивают модуль гладкости функции в более сильной метрике через интеграл от модуля гладкости в более слабой метрике. Одномерные варианты таких оценок восходят к работе П. Л. Ульянова и далее развивались в работах В.И. Коляды, Б.В. Симонова, С.Ю. Тихонова, W. Trebels и других авторов; см. обзорные ссылки в [1-11].

Для функций многих переменных особый интерес представляют смешанные модули гладкости. Они отражают совместное действие разностей по различным переменным и тесно связаны с приближением тригонометрическими полиномами по углу. Для смешанных модулей положительного порядка и смешанных метрик $L_{p_1p_2}$ соответствующие неравенства Ульянова были установлены в [12]. В этой работе, в частности, доказаны оценки между смешанными модулями в разных метриках и оценки модулей гладкости обычных производных Вейля.

С другой стороны, в работах [7, 9] изучались обобщённые производные Лиувилля-Вейля, задаваемые не обязательно степенным множителем Фурье, а более общим множителем λ. В одномерном случае такие множители естественно описываются классом обобщённо монотонных последовательностей GM. В двумерном случае для приближения двумерным углом возникает класс GM^2, в котором контролируются две одномерные вариации и смешанная вариация множителя.

Цель настоящей статьи – объединить эти два направления: получить неравенство типа Ульянова для смешанного модуля гладкости обобщённой производной Лиувилля-Вейля с GM^2-множителем. Основной результат имеет вид

$$\omega_{\alpha_1,\alpha_2}\left(f^{(\lambda,\beta_1,\beta_2)};\delta_1,\delta_2\right)_{q_1q_2} \lesssim \left(\int_0^{\delta_2}\left(\int_0^{\delta_1}\left[t_1^{-\sigma_1}t_2^{-\sigma_2}\Phi_{\lambda,\alpha+\sigma}(f;t_1,t_2)\right]^{q_1^*}\frac{dt_1}{t_1}\right)^{q_2^*/q_1^*}\frac{dt_2}{t_2}\right)^{1/q_2^*},$$

где $\sigma_i = 1/p_i - 1/q_i$, а $\Phi_{\lambda,s}$ – dyadic-функционал, построенный по угловым приближениям функции f и разностям множителя λ.

В статье подробно доказываются две технические леммы. Первая лемма оценивает смешанный модуль гладкости через угловые наилучшие приближения. Вторая лемма оценивает угловые приближения обобщённой производной через угловые приближения исходной функции. Последняя лемма является двумерной версией известных оценок для преобразованных рядов Фурье с обобщённо монотонными множителями и доказывается с помощью двумерного преобразования Абеля.

Смешанные пространства, модули гладкости и угловые приближения

Пусть $\mathbb{T} = [0,2\pi)$, $\mathbb{T}^2 = \mathbb{T}\times\mathbb{T}$. Все функции считаются 2π-периодическими по каждой переменной.

Определение 1 (Смешанная метрика). Пусть $1 \le p_1, p_2 \le \infty$. Через $L_{p_1p_2} = L_{p_1p_2}(\mathbb{T}^2)$ обозначим пространство измеримых функций $f(x_1,x_2)$, для которых конечна смешанная норма $\| f \|_{p_1p_2} = \left(\int_0^{2\pi}\left(\int_0^{2\pi}|f(x_1,x_2)|^{p_1}\,dx_1\right)^{p_2/p_1}dx_2\right)^{1/p_2}$, с обычными изменениями при $p_1 = \infty$ или $p_2 = \infty$.

Определение 2 (Подпространство нулевых средних). Через $L^0_{p_1p_2}$ обозначим множество функций $f \in L_{p_1p_2}$ таких, что $\int_0^{2\pi} f(x_1,x_2)\,dx_1 = 0$ для почти всех x_2, и $\int_0^{2\pi} f(x_1,x_2)\,dx_2 = 0$ для почти всех x_1.

Для $\alpha > 0$ положим

$$\Delta_h^\alpha f(x) = \sum_{\nu=0}^{\infty} (-1)^\nu \binom{\alpha}{\nu} f(x + (\alpha-\nu)h),$$

где

$$\binom{\alpha}{0} = 1, \qquad \binom{\alpha}{\nu} = \frac{\alpha(\alpha-1)\cdots(\alpha-\nu+1)}{\nu!}, \quad \nu \geq 1.$$

Разности по переменным x_1 и x_2 будем обозначать через $\Delta^{\alpha_1}_{h_1,1}$ и $\Delta^{\alpha_2}_{h_2,2}$.

Определение 3 Для $\alpha_1, \alpha_2 > 0$ смешанный модуль гладкости функции $f \in L_{p_1p_2}$ определяется равенством

$$\omega_{\alpha_1,\alpha_2}(f;\delta_1,\delta_2)_{p_1p_2} = \sup_{|h_i|\leq\delta_i,\ i=1,2} \left\|\Delta^{\alpha_1}_{h_1,1}\Delta^{\alpha_2}_{h_2,2} f\right\|_{p_1p_2}.$$

Определение 4 Пусть $m_1, m_2 \in \mathbb{N} \cup \{0\}$. Угловым наилучшим приближением функции $f \in L_{p_1p_2}$ называется величина

$$Y_{m_1,m_2}(f)_{p_1p_2} = \inf_{T_{m_1,\infty}, T_{\infty,m_2}} \| f - T_{m_1,\infty} - T_{\infty,m_2} \|_{p_1p_2},$$

где $T_{m_1,\infty}$ является тригонометрическим полиномом порядка не выше m_1 по x_1, а T_{∞,m_2} является тригонометрическим полиномом порядка не выше m_2 по x_2.

Запись $A \lesssim B$ будет означать, что $A \leq CB$ с постоянной $C > 0$, не зависящей от рассматриваемой функции и основных параметров. Если одновременно $A \lesssim B$ и $B \lesssim A$, пишем $A \asymp B$.

Обобщённая производная Лиувилля-Вейля и класс GM^2

Пусть $f \in L^0_{p_1p_2}$ имеет ряд Фурье

$$\sigma(f) = \sum_{k_1=1}^{\infty}\sum_{k_2=1}^{\infty} (a_{k_1k_2}\cos k_1x_1\cos k_2x_2 + b_{k_1k_2}\cos k_1x_1\sin k_2x_2 + c_{k_1k_2}\sin k_1x_1\cos k_2x_2 + d_{k_1k_2}\sin k_1x_1\sin k_2x_2).$$

Определение 5 (Обобщённая производная Лиувилля-Вейля). Пусть $\lambda = \{\lambda_{k_1,k_2}\}_{k_1,k_2\geq 1}$ – положительная последовательность, а $\beta_1, \beta_2 \in \mathbb{R}$.

Рассмотрим преобразованный ряд

$$\sigma(f,\lambda,\beta_1,\beta_2) = \sum_{k_1=1}^{\infty}\sum_{k_2=1}^{\infty} \lambda_{k_1,k_2} \Big(a_{k_1k_2}\cos(k_1x_1 + \beta_1\pi/2)\cos(k_2x_2 + \beta_2\pi/2) + b_{k_1k_2}\cos(k_1x_1 + \beta_1\pi/2)\sin(k_2x_2 + \beta_2\pi/2) + c_{k_1k_2}\sin(k_1x_1 + \beta_1\pi/2)\cos(k_2x_2 + \beta_2\pi/2) + d_{k_1k_2}\sin(k_1x_1 + \beta_1\pi/2)\sin(k_2x_2 + \beta_2\pi/2)\Big).$$

Если этот ряд является рядом Фурье некоторой функции $g \in L^0_{p_1p_2}$, то g называется $(\lambda,\beta_1,\beta_2)$-производной, или обобщённой производной Лиувилля-Вейля, функции f и обозначается $g = f^{(\lambda,\beta_1,\beta_2)}$.

Если $\lambda_{k_1,k_2} = k_1^{\rho_1} k_2^{\rho_2}$ и $\beta_i = \rho_i$, то $f^{(\lambda,\beta_1,\beta_2)}$ совпадает с обычной смешанной производной Вейля $f^{(\rho_1,\rho_2)}$.

Определение 6 Положительная последовательность $\lambda = \{\lambda_{k_1,k_2}\}_{k_1,k_2 \geq 1}$ принадлежит классу GM^2, если существует постоянная $C > 0$ такая, что для всех $n_1, n_2 \in \mathbb{N}$ выполнены условия

$\sum_{k_1=n_1}^{2n_1} |\lambda_{k_1,n_2} - \lambda_{k_1+1,n_2}| \leq C\lambda_{n_1,n_2}$, $\sum_{k_2=n_2}^{2n_2} |\lambda_{n_1,k_2} - \lambda_{n_1,k_2+1}| \leq C\lambda_{n_1,n_2}$, и $\sum_{k_1=n_1}^{2n_1} \sum_{k_2=n_2}^{2n_2} |\lambda_{k_1,k_2} - \lambda_{k_1+1,k_2} - \lambda_{k_1,k_2+1} + \lambda_{k_1+1,k_2+1}| \leq C\lambda_{n_1,n_2}$.

Это двумерный аналог класса GM, введённого и подробно исследованного в связи с тригонометрическими рядами и преобразованиями Фурье; см. [8, 15, 16].

Методология исследования

Сначала приведём смешанное неравенство Ульянова, которое будет использоваться как известный результат.

Лемма 1 (Смешанное неравенство Ульянова). Пусть $h \in L^0_{p_1p_2}$, $\alpha_i > 0$, $\delta_i \in (0,1)$, $i = 1,2$, и для каждого i выполнено одно из условий $1 < p_i < q_i < \infty$ или $p_i = 1,\ q_i = \infty$.

Положим $\sigma_i = \frac{1}{p_i} - \frac{1}{q_i}$, $q_i^* = \begin{cases} q_i, & q_i < \infty, \\ 1, & q_i = \infty. \end{cases}$

Тогда

$$\omega_{\alpha_1,\alpha_2}(h;\delta_1,\delta_2)_{q_1q_2} \lesssim \left(\int_0^{\delta_2} \left(\int_0^{\delta_1} \left[t_1^{-\sigma_1} t_2^{-\sigma_2} \omega_{\alpha_1+\sigma_1,\alpha_2+\sigma_2}(h;t_1,t_2)_{p_1p_2} \right]^{q_1^*} \frac{dt_1}{t_1} \right)^{q_2^*/q_1^*} \frac{dt_2}{t_2} \right)^{1/q_2^*}.$$

Этот результат доказан в [12]. В случае $q_i = \infty$ соответствующая внешняя степень понимается как указано в определении q_i^*.

Далее докажем техническую лемму, которая переводит смешанный модуль в диадический-сумму угловых приближений.

Лемма 2 ([13,14]). Если $0 \leq m_i \leq M_i$, $i = 1,2$, то $Y_{M_1,M_2}(h)_{p_1p_2} \leq Y_{m_1,m_2}(h)_{p_1p_2}$.

Лемма 3 (Модуль через угловые приближения). Пусть $s_1, s_2 > 0$ и $h \in L^0_{p_1p_2}$. Если $2^{-n_i-1} < t_i \leq 2^{-n_i}$, $i = 1,2$, то

$$\omega_{s_1,s_2}(h;t_1,t_2)_{p_1p_2} \lesssim 2^{-n_1s_1} 2^{-n_2s_2} \sum_{m_1=0}^{n_1+1} \sum_{m_2=0}^{n_2+1} 2^{m_1s_1}\, 2^{m_2s_2} Y_{2^{m_1}-1,2^{m_2}-1}(h)_{p_1p_2}.$$

Доказательство. Докажем утверждение по шагам.

По обратной теореме для смешанных модулей гладкости и угловых приближений, см. [9, 12], для любых $N_1, N_2 \in \mathbb{N}$ имеем

$$\omega_{s_1,s_2}\left(h;\frac{1}{N_1},\frac{1}{N_2}\right)_{p_1p_2} \lesssim N_1^{-s_1} N_2^{-s_2} \sum_{v_1=1}^{N_1+1} \sum_{v_2=1}^{N_2+1} v_1^{s_1-1}\, v_2^{s_2-1} Y_{v_1-1,v_2-1}(h)_{p_1p_2}.$$

Положим $N_i = 2^{n_i}$. Из условия $t_i \leq 2^{-n_i} = 1/N_i$ и монотонности смешанного модуля по каждому аргументу следует

$$\omega_{s_1,s_2}(h;t_1,t_2)_{p_1p_2} \leq \omega_{s_1,s_2}(h;2^{-n_1},2^{-n_2})_{p_1p_2}.$$

Поэтому

$$\omega_{s_1,s_2}(h;t_1,t_2)_{p_1p_2} \lesssim 2^{-n_1s_1} 2^{-n_2s_2} S,$$

где

$$S = \sum_{v_1=1}^{2^{n_1}+1} \sum_{v_2=1}^{2^{n_2}+1} v_1^{s_1-1} v_2^{s_2-1} Y_{v_1-1,v_2-1}(h)_{p_1p_2}.$$

Разобьём множества индексов на диадический -блоки. Для $m \geq 1$ положим

$$B_m = \{v \in \mathbb{N}: 2^{m-1} < v \leq 2^m\}, \qquad B_0 = \{1\}.$$

Тогда $\{1, \dots, 2^{n_i}+1\} \subset \bigcup_{m_i=0}^{n_i+1} B_{m_i}$.

Следовательно,

$$S \leq \sum_{m_1=0}^{n_1+1} \sum_{m_2=0}^{n_2+1} \sum_{v_1 \in B_{m_1}} \sum_{v_2 \in B_{m_2}} v_1^{s_1-1} v_2^{s_2-1} Y_{v_1-1,v_2-1}(h)_{p_1p_2}.$$

Для $s_i > 0$ справедлива оценка $\sum_{v_i \in B_{m_i}} v_i^{s_i-1} \lesssim 2^{m_i s_i}, \qquad m_i = 0,1, \dots, n_i+1.$

Действительно, при $m_i \geq 1$ это следует из сравнения суммы с интегралом $\int_{2^{m_i-1}}^{2^{m_i}} x^{s_i-1}\, dx$, а при $m_i = 0$ очевидно.

Теперь используем монотонность Y_{m_1,m_2} в правильном направлении. Если $v_i \in B_{m_i}$ и $m_i \geq 1$, то $v_i - 1 \geq 2^{m_i-1} - 1$. Поэтому, полагая $r_i = (m_i - 1)_+ = \max\{m_i - 1, 0\}$, получаем по лемме 2

$$Y_{v_1-1,v_2-1}(h)_{p_1p_2} \leq Y_{2^{r_1}-1,2^{r_2}-1}(h)_{p_1p_2}.$$

Следовательно,

$$S \lesssim \sum_{m_1=0}^{n_1+1} \sum_{m_2=0}^{n_2+1} 2^{m_1 s_1}\, 2^{m_2 s_2} Y_{2^{(m_1-1)_+}-1,2^{(m_2-1)_+}-1}(h)_{p_1p_2}.$$

Сдвиг индексов не меняет порядок суммы, поскольку каждому r_i соответствует не более двух значений m_i, а $2^{m_i s_i} \leq C(s_i) 2^{r_i s_i}$. Поэтому

$$S \lesssim \sum_{r_1=0}^{n_1+1} \sum_{r_2=0}^{n_2+1} 2^{r_1 s_1}\, 2^{r_2 s_2} Y_{2^{r_1}-1,2^{r_2}-1}(h)_{p_1p_2}.$$

Подставляя эту оценку в результат шага 2, получаем требуемое неравенство.

Угловые приближения обобщённой производной

Для $m \geq 0$ положим

$$N_m = 2^m.$$

Будем использовать обозначения

$$\Delta_1 \lambda_{\nu_1,m_2} = \lambda_{N_{\nu_1},N_{m_2}} - \lambda_{N_{\nu_1+1},N_{m_2}},$$
$$\Delta_2 \lambda_{m_1,\nu_2} = \lambda_{N_{m_1},N_{\nu_2}} - \lambda_{N_{m_1},N_{\nu_2+1}},$$

и

$$\Delta_{12} \lambda_{\nu_1,\nu_2} = \lambda_{N_{\nu_1},N_{\nu_2}} - \lambda_{N_{\nu_1+1},N_{\nu_2}} - \lambda_{N_{\nu_1},N_{\nu_2+1}} + \lambda_{N_{\nu_1+1},N_{\nu_2+1}}.$$

Определение 7 Для $m_1, m_2 \geq 0$ положим

$$\begin{aligned} B_\lambda(m_1, m_2; f) = \quad & \lambda_{N_{m_1}, N_{m_2}} Y_{N_{m_1}-1, N_{m_2}-1}(f)_{p_1 p_2} \\ & + \sum_{\nu_1=m_1}^{\infty} \left|\Delta_1 \lambda_{\nu_1, m_2}\right| Y_{N_{\nu_1}-1, N_{m_2}-1}(f)_{p_1 p_2} \\ & + \sum_{\nu_2=m_2}^{\infty} \left|\Delta_2 \lambda_{m_1, \nu_2}\right| Y_{N_{m_1}-1, N_{\nu_2}-1}(f)_{p_1 p_2} \\ & + \sum_{\nu_1=m_1}^{\infty} \sum_{\nu_2=m_2}^{\infty} \left|\Delta_{12} \lambda_{\nu_1, \nu_2}\right| Y_{N_{\nu_1}-1, N_{\nu_2}-1}(f)_{p_1 p_2}. \end{aligned}$$

Следующая лемма является двумерным вариантом оценки наилучших приближений обобщённой производной.

Лемма 4 ([9]).

Пусть $1 < p_1, p_2 < \infty$, $f \in L^0_{p_1 p_2}$, $\lambda \in GM^2$, $\beta_1, \beta_2 \in \mathbb{R}$, и пусть $g = f^{(\lambda, \beta_1, \beta_2)} \in L^0_{p_1 p_2}$.

Тогда для всех $m_1, m_2 \geq 0$ выполнена оценка

$$Y_{N_{m_1}-1, N_{m_2}-1}(g)_{p_1 p_2} \lesssim B_\lambda(m_1, m_2; f).$$

Определение 8 (Диадический -функционал $\Phi_{\lambda,s}$).

Пусть $s = (s_1, s_2)$, $s_1, s_2 > 0$. Если $2^{-n_i - 1} < t_i \leq 2^{-n_i}, \quad i = 1,2$, то положим

$$\Phi_{\lambda,s}(f; t_1, t_2) = 2^{-n_1 s_1} 2^{-n_2 s_2} \sum_{m_1=0}^{n_1+1} \sum_{m_2=0}^{n_2+1} 2^{m_1 s_1}\, 2^{m_2 s_2} B_\lambda(m_1, m_2; f).$$

Лемма 5 Пусть $s_1, s_2 > 0$, $1 < p_1, p_2 < \infty$, $f \in L^0_{p_1 p_2}$, $\lambda \in GM^2$, $\beta_1, \beta_2 \in \mathbb{R}$, и пусть $g = f^{(\lambda, \beta_1, \beta_2)} \in L^0_{p_1 p_2}$. Тогда при $2^{-n_i - 1} < t_i \leq 2^{-n_i}$, $i = 1,2$, выполнено

$$\omega_{s_1, s_2}(g; t_1, t_2)_{p_1 p_2} \lesssim \Phi_{\lambda,s}(f; t_1, t_2).$$

Доказательство. Применяем лемму 3 к функции $h = g$:

$$\omega_{s_1, s_2}(g; t_1, t_2)_{p_1 p_2} \lesssim 2^{-n_1 s_1} 2^{-n_2 s_2} \sum_{m_1=0}^{n_1+1} \sum_{m_2=0}^{n_2+1} 2^{m_1 s_1}\, 2^{m_2 s_2} Y_{N_{m_1}-1, N_{m_2}-1}(g)_{p_1 p_2}.$$

По лемме 4

$$Y_{N_{m_1}-1, N_{m_2}-1}(g)_{p_1 p_2} \lesssim B_\lambda(m_1, m_2; f).$$

Подставляя эту оценку в предыдущую сумму, получаем

$$\omega_{s_1, s_2}(g; t_1, t_2)_{p_1 p_2} \lesssim 2^{-n_1 s_1} 2^{-n_2 s_2} \sum_{m_1=0}^{n_1+1} \sum_{m_2=0}^{n_2+1} 2^{m_1 s_1}\, 2^{m_2 s_2} B_\lambda(m_1, m_2; f).$$

Правая часть по определению равна $\Phi_{\lambda,s}(f; t_1, t_2)$. □

Результаты исследования

Перейдем к формулировке главного результата данного раздела. Следующая теорема устанавливает основное неравенство типа Ульянова, которое будет играть ключевую роль в дальнейших оценках.

Теорема 1 Пусть $1 < p_i < q_i < \infty,\ i = 1,2$, и $\sigma_i = \frac{1}{p_i} - \frac{1}{q_i},\ s_i = \alpha_i + \sigma_i,\ \alpha_i > 0$. Пусть $f \in L^0_{p_1p_2}$, $\lambda \in GM^2$, $\beta_1, \beta_2 \in \mathbb{R}$, и пусть существует $g = f^{(\lambda,\beta_1,\beta_2)} \in L^0_{p_1p_2}$.

Тогда для любых $\delta_1, \delta_2 \in (0,1)$ справедливо неравенство

$$\omega_{\alpha_1,\alpha_2}\left(f^{(\lambda,\beta_1,\beta_2)};\delta_1,\delta_2\right)_{q_1q_2} \lesssim \left(\int_0^{\delta_2}\left(\int_0^{\delta_1}\left[t_1^{-\sigma_1}t_2^{-\sigma_2}\Phi_{\lambda,s}(f;t_1,t_2)\right]^{q_1}\frac{dt_1}{t_1}\right)^{q_2/q_1}\frac{dt_2}{t_2}\right)^{1/q_2}.$$

Доказательство. Положим $g = f^{(\lambda,\beta_1,\beta_2)}$. По смешанному неравенству Ульянова, лемма 1, имеем

$$\omega_{\alpha_1,\alpha_2}(g;\delta_1,\delta_2)_{q_1q_2} \lesssim \left(\int_0^{\delta_2}\left(\int_0^{\delta_1}\left[t_1^{-\sigma_1}t_2^{-\sigma_2}\omega_{\alpha_1+\sigma_1,\alpha_2+\sigma_2}(g;t_1,t_2)_{p_1p_2}\right]^{q_1}\frac{dt_1}{t_1}\right)^{q_2/q_1}\frac{dt_2}{t_2}\right)^{1/q_2}.$$

Для каждого $t_i \in (0,1)$ выберем n_i так, что $2^{-n_i-1} < t_i \le 2^{-n_i}$. Применяя лемму 5 с $s_i = \alpha_i + \sigma_i$,

Получаем

$$\omega_{\alpha_1+\sigma_1,\alpha_2+\sigma_2}(g;t_1,t_2)_{p_1p_2} \lesssim \Phi_{\lambda,s}(f;t_1,t_2).$$

Подстановка этой оценки в предыдущее интегральное неравенство даёт

$$\omega_{\alpha_1,\alpha_2}(g;\delta_1,\delta_2)_{q_1q_2} \lesssim \left(\int_0^{\delta_2}\left(\int_0^{\delta_1}\left[t_1^{-\sigma_1}t_2^{-\sigma_2}\Phi_{\lambda,s}(f;t_1,t_2)\right]^{q_1}\frac{dt_1}{t_1}\right)^{q_2/q_1}\frac{dt_2}{t_2}\right)^{1/q_2}.$$

Так как $g = f^{(\lambda,\beta_1,\beta_2)}$, теорема доказана. □

Теорема 2 Пусть для каждого $i = 1,2$ выполнено одно из условий $1 < p_i < q_i < \infty$ или $p_i = 1,\ q_i = \infty$. Пусть, кроме того, для рассматриваемого множителя λ и фаз β_1, β_2 справедлива угловая оценка $Y_{N_{m_1}-1,N_{m_2}-1}\left(f^{(\lambda,\beta_1,\beta_2)}\right)_{p_1p_2} \lesssim B_\lambda(m_1,m_2;f)$. Тогда утверждение теоремы 1 остаётся верным в форме

$$\omega_{\alpha_1,\alpha_2}\left(f^{(\lambda,\beta_1,\beta_2)};\delta_1,\delta_2\right)_{q_1q_2} \lesssim \left(\int_0^{\delta_2}\left(\int_0^{\delta_1}\left[t_1^{-\sigma_1}t_2^{-\sigma_2}\Phi_{\lambda,s}(f;t_1,t_2)\right]^{q_1^*}\frac{dt_1}{t_1}\right)^{q_2^*/q_1^*}\frac{dt_2}{t_2}\right)^{1/q_2^*}.$$

Доказательство. Доказательство дословно повторяет доказательство теоремы 1, но вместо q_i используются q_i^*, а смешанное неравенство Ульянова берётся в полной форме леммы 1. Единственное дополнительное требование – наличие угловой оценки для обобщённой производной в соответствующей предельной метрике. □

Частные случаи и примеры

Обычная смешанная производная Вейля

Следствие 1. Пусть $\rho_1, \rho_2 \ge 0$ и $\lambda_{k_1,k_2} = k_1^{\rho_1}k_2^{\rho_2}$. Тогда при подходящем выборе β_i имеем $f^{(\lambda,\beta_1,\beta_2)} = f^{(\rho_1,\rho_2)}$. Кроме того,

$$\Phi_{\lambda,s}(f;t_1,t_2) \lesssim t_1^{-\rho_1} t_2^{-\rho_2} \omega_{s_1+\rho_1,s_2+\rho_2}(f;t_1,t_2)_{p_1p_2}.$$

Следовательно, из теоремы 1 следует

$$\omega_{\alpha_1,\alpha_2}\left(f^{(\rho_1,\rho_2)};\delta_1,\delta_2\right)_{q_1q_2}$$
$$\lesssim \left(\int_0^{\delta_2}\left(\int_0^{\delta_1}\left[t_1^{-\rho_1-\sigma_1}t_2^{-\rho_2-\sigma_2}\omega_{\alpha_1+\rho_1+\sigma_1,\alpha_2+\rho_2+\sigma_2}(f;t_1,t_2)_{p_1p_2}\right]^{q_1}\frac{dt_1}{t_1}\right)^{q_2/q_1}\frac{dt_2}{t_2}\right)^{1/q_2}.$$

Proof. Для $\lambda_{k_1,k_2} = k_1^{\rho_1}k_2^{\rho_2}$ dyadic-разности имеют порядок

$$\left|\Delta_1\lambda_{\nu_1,m_2}\right| \lesssim 2^{\nu_1\rho_1}2^{m_2\rho_2}, \qquad \left|\Delta_2\lambda_{m_1,\nu_2}\right| \lesssim 2^{m_1\rho_1}2^{\nu_2\rho_2},$$

и

$$\left|\Delta_{12}\lambda_{\nu_1,\nu_2}\right| \lesssim 2^{\nu_1\rho_1}2^{\nu_2\rho_2}.$$

Подставляя эти оценки в B_λ и применяя обратную оценку для смешанного модуля через угловые приближения порядка $(s_1+\rho_1, s_2+\rho_2)$, получаем

$$\Phi_{\lambda,s}(f;t_1,t_2) \lesssim t_1^{-\rho_1} t_2^{-\rho_2} \omega_{s_1+\rho_1,s_2+\rho_2}(f;t_1,t_2)_{p_1p_2}.$$

После этого достаточно применить теорему 1. □

Дискуссия

Главным результатом данного исследования является установление нового неравенства типа Ульянова для смешанного модуля гладкости обобщённых производных Лиувилля-Вейля функций двух переменных в смешанных метриках $L_{q_1q_2}$. Значимость полученной оценки (Теоремы 1 и 2) заключается в том, что она связывает дифференциально-разностные свойства обобщённой производной функции непосредственно с её наилучшими угловыми приближениями и структурными свойствами трансформирующего множителя через специальный диадический-функционал $\Phi_{\lambda,s}$. Теоретическая ценность работы заключается в преодолении ограничений, связанных с фиксированным (степенным) порядком дифференцирования. Введение в многомерный анализ классов последовательностей GM^2 (обобщённо монотонных последовательностей) позволяет описывать гладкость функций в терминах широкого спектра операторов дифференцирования. Это имеет фундаментальное значение для теории вложений функциональных пространств, так как даёт точный аппарат для оценки изменения гладкости при одновременном переходе в более сильную смешанную метрику и применении интегро-дифференциальных операторов общего вида.

Полученные результаты представляют собой естественное и существенное развитие классического направления теории приближений, заложенного П.Л. Ульяновым, и развитого для смешанных метрик и модулей гладкости в работах В.И. Коляды, М.К. Потапова, Б.В. Симонова и С.Ю. Тихонова. Настоящая работа открывает несколько многообещающих направлений для дальнейших изысканий в области гармонического анализа:

- Расширение размерности: Очевидной, но математически нетривиальной задачей является обобщение полученных оценок на функции n переменных (n > 2), что потребует определения и исследования многомерных классов монотонности GM^nи работы с кратными преобразованиями Абеля.

- Предельные случаи метрик: Перспективным видится детальное изучение граничных пространств, когда параметры p_i или q_i принимают значения 1 или ∞, а также рассмотрение аналогичных неравенств в пространствах Лоренца или Орлича со смешанной нормой.

- Обратные неравенства: Важным продолжением станет доказательство соответствующих обратных неравенств типа Ульянова, что позволит получить критерии совпадения или вложения классов функций, определяемых через обобщённые модули гладкости.

Заключение

Доказано неравенство типа Ульянова для смешанных модулей гладкости обобщённых производных Лиувилля-Вейля с двумерно обобщённо монотонными множителями. Новая часть оценки состоит в том, что вместо обычной производной Вейля используется общий множитель Фурье $\lambda \in GM^2$, а гладкость производной контролируется через функционал $\Phi_{\lambda,s}$, построенный по угловым приближениям исходной функции и разностям множителя. При степенном множителе $\lambda_{k_1,k_2} = k_1^{\rho_1} k_2^{\rho_2}$ получаем известную оценку для смешанной производной Вейля.

Благодарности

Данное исследование частично финансировалось Научным комитетом Министерства науки и высшего образования Республики Казахстан (грант № AP22686170).

Список использованных источников

[1] Z. Ditzian, S. Tikhonov, Ul'yanov and Nikol'skii-type inequalities, Journal of Approximation Theory, 133(1) (2005), 100–133. DOI:10.1016/j.jat.2004.12.008

[2] P. L. Ul'yanov, The imbedding of certain function classes H_p^ω, Izv. Akad. Nauk SSSR Ser. Mat., 32(3) (1968), 649–686. DOI: https://doi.org/10.1070/IM1968v002n03ABEH000650

[3] V. I. Kolyada, On the relations between the moduli of continuity in different metrics, Trudy Mat. Inst. Steklov, 181 (1988), 117–136.https://www.mathnet.ru/php/archive.phtml?wshow=paper&jrnid=tm&paperid=1937&option_lang=rus

[4] B. Simonov, S. Tikhonov, Sharp Ul'yanov-type inequalities using fractional smoothness, Journal of Approximation Theory, 162(9) (2010), 1654–1684. DOI: 10.1016/j.jat.2010.04.010

[5] S. Tikhonov, W. Trebels, Ulyanov inequalities and generalized Liouville derivatives, Proceedings of the Royal Society of Edinburgh Section A, 141(1) (2011), 205–224. DOI: 10.1017/S0308210509001048

A. Jumabayeva, Sharp Ul'yanov inequalities for generalized Liouville–Weyl derivatives, Analysis Mathematica, 43(2) (2017), 279–302. DOI: 10.1007/s10476-017-0308-0

B. Jumabayeva, Liouville–Weyl derivatives, best approximations, and moduli of smoothness, Acta Mathematica Hungarica, 145(2) (2015), 369–391. DOI: 10.1007/s10474-015-0485-x

[6] Jumabayeva, B. V. Simonov, Transformation of Fourier series by means of general monotone sequences, Mathematical Notes, 107(5) (2020), 674–692. DOI: 10.1134/S0001434620050041

[7] M. K. Potapov, B. V. Simonov, S. Yu. Tikhonov, Mixed moduli of smoothness in L_p, $1 < p < \infty$: a survey, Surveys in Approximation Theory, 8 (2013), 1–57. https://arxiv.org/abs/1304.3329

[8] M. K. Potapov, B. V. Simonov, Properties of mixed moduli of smoothness of positive order in a mixed metric, Vestnik Moskovskogo Universiteta. Seriya 1. Matematika. Mekhanika, 6 (2014), 31–40. DOI: 10.3103/S0027132214060059

[9] M. K. Potapov, B. V. Simonov, Inequalities of different metrics for trigonometric polynomials, Russian Mathematics, 63(1) (2019), 49–62. DOI: https://doi.org/10.26907/0021-3446-2019-1-49-62

[10] V. Simonov, A. A. Jumabayeva, Ulyanov inequalities for the mixed moduli of smoothness in mixed metrics, Georgian Mathematical Journal 32 (1), 149-161. DOI: 10.1515/gmj-2024-2034

[11] Yu. Kolomoitsev, S. Tikhonov, Properties of moduli of smoothness in $L_p(\mathbb{R}^d)$, Journal of Approximation Theory, 257 (2020), Article 105423. DOI:10.1016/j.jat.2020.105423

[12] Yu. Kolomoitsev, S. Tikhonov, Hardy–Littlewood and Ulyanov inequalities, Memoirs of the American Mathematical Society, 271(1325) (2021). DOI / Код: MEMO/271/1325 / AMS memo1325.pdf

[13] S. Tikhonov, Trigonometric series with general monotone coefficients, Journal of Mathematical Analysis and Applications, 326(1) (2007), 721–735. DOI: 10.1016/j.jmaa.2006.02.053

[14] E. Liflyand, S. Tikhonov, A concept of general monotonicity and applications, Mathematische Nachrichten, 284(8–9) (2011), 1083–1098. https://doi.org/10.1002/mana.200810262

IRSTI 49.33.29

https://doi.org/10.51889/2959-5894.2026.95.3.006

D. Marlambekov[1*], S. Mukhambetzhanov[1], S. Ibraheem[2]

[1]Al-Farabi Kazakh National University, Almaty, Kazakhstan
[2]Istanbul Technical University, Istanbul, Turkey
[*]*e-mail: marlambekov_duman@live.kaznu.kz*

PPO-BASED RESOURCE ALLOCATION FOR FIFTH-GENERATION NETWORK SLICING

Abstract

This paper presents a deep reinforcement learning approach to service-level-agreement (SLA) aware network slicing in 5G radio access networks. We formulate the dynamic allocation of Physical Resource Blocks (PRB) among three coexisting slices – enhanced Mobile Broadband (eMBB), Ultra-Reliable Low-Latency Communications (URLLC), and massive Machine-Type Communications (mMTC) – as a Markov Decision Process and develop a Gymnasium-compatible simulation environment in which per-slice SLA requirements follow the 3GPP TS 28.530 framework. A Proximal Policy Optimization (PPO) agent is trained with a composite reward that jointly rewards aggregate throughput, penalizes SLA violations, and encourages resource utilization. The agent is evaluated over 100 episodes against three classical baselines: Static Equal, Static Priority, and Greedy Proportional allocation. The PPO agent reduces the aggregate SLA violation rate to 2×10^{-5}, three to four orders of magnitude lower than the Static Priority baseline (0.46), while sustaining a mean aggregate throughput of 74.7 Mbit/s and a resource utilization of 98.6%. A throughput-versus-violation analysis confirms that the learned policy occupies an operating point that no single static rule attains. The complete environment, training pipeline, and evaluation scripts are open-source and reproducible using the Stable-Baselines3 library.

Keywords: 5G network slicing, reinforcement learning, proximal policy optimization, service-level agreement, PRB allocation, eMBB, URLLC, mMTC, radio resource management.

Д. Марламбеков[1], С. Мұхамбетжанов[1], Ш. Ибрахим[2]
[1] Әл-Фараби атындағы Қазақ ұлттық университеті, Алматы қ., Қазақстан
[2] Стамбул техникалық университеті, Стамбул қ., Түркия

РРО НЕГІЗІНДЕ БЕСІНШІ БУЫН ЖЕЛІЛІК СЛАЙСИНГІ ҮШІН РЕСУРСТАРДЫ БӨЛУ

Аңдатпа

Бұл жұмыста 5G радиоқолжетімділік желілеріндегі қызмет көрсету деңгейі туралы келісімдерді (SLA) ескеретін желілік слайсингке арналған тереңдетілген оқыту тәсілі ұсынылған. Үш қатар жұмыс істейтін слайс – кеңейтілген мобильді кеңжолақты байланыс (eMBB), аса сенімді және кідірісі төмен байланыс (URLLC) және жаппай машиналық типтегі байланыс (mMTC) – арасында физикалық ресурстық блоктарды (PRB) динамикалық бөлу міндеті Марков шешім қабылдау процесі ретінде тұжырымдалған. Әрбір слайсқа қойылатын SLA талаптары 3GPP TS 28.530 стандартына сәйкес келетін, Gymnasium ортасымен үйлесімді имитациялық орта әзірленді. Proximal Policy Optimization (PPO) агенті жиынтық өткізу қабілетін ынталандыратын, SLA бұзушылықтарын айыптайтын және ресурстарды тиімді пайдалануды көтермелейтін құрамдас марапат функциясы арқылы оқытылады. Агент 100 эпизод бойынша бағаланып, Static Equal, Static Priority және Greedy Proportional атты үш классикалық базалық стратегиямен салыстырылды. PPO агенті SLA бұзушылықтарының жиынтық жиілігін 2×10^{-5} деңгейіне дейін төмендетті, бұл Static Priority базалық стратегиясымен салыстырғанда (0,46) үш-төрт реттік шамаға аз. Сонымен қатар орташа жиынтық өткізу қабілеті 74,7 Мбит/с, ал ресурстарды пайдалану деңгейі 98,6% болды. Өткізу қабілеті мен бұзушылықтар арасындағы ымыраны талдау үйренген саясаттың бірде-бір статикалық стратегия қол жеткізе алмайтын жұмыс нүктесіне жететінін көрсетті. Толық имитациялық орта, оқыту үдерісі және бағалау скрипттері ашық түрде ұсынылған және Stable-Baselines3 кітапханасы арқылы қайта өндіріледі.

Түйін сөздер: 5G желілік слайсингі, күшейтпелі оқыту, Proximal Policy Optimization, қызмет көрсету деңгейі туралы келісім, PRB бөлу, eMBB, URLLC, mMTC, радиоресурстарды басқару.

Д. Марламбеков[1*], С. Мухамбетжанов[1], Ш. Ибрахим[2]
[1]Казахский национальный университет имени аль-Фараби, г.Алматы, Казахстан
[2]Стамбульский технический университет, г.Стамбул, Турция

РАСПРЕДЕЛЕНИЕ РЕСУРСОВ НА ОСНОВЕ PPO ДЛЯ СЕГМЕНТАЦИИ СЕТЕЙ ПЯТОГО ПОКОЛЕНИЯ

Аннотация

В данной работе представлен подход на основе глубокого обучения с подкреплением к сетевому слайсингу в сетях радиодоступа 5G с учётом соглашений об уровне обслуживания (SLA). Задача динамического распределения физических ресурсных блоков (PRB) между тремя одновременно функционирующими слайсами – усовершенствованной мобильной широкополосной связью (eMBB), сверхнадёжной связью с малой задержкой (URLLC) и массовой межмашинной связью (mMTC) – формулируется как марковский процесс принятия решений. Разработана совместимая с Gymnasium имитационная среда, в которой требования SLA для каждого слайса соответствуют стандарту 3GPP TS 28.530. Агент Proximal Policy Optimization (PPO) обучается с использованием составной функции вознаграждения, которая поощряет совокупную пропускную способность, штрафует нарушения SLA и стимулирует эффективное использование ресурсов. Агент оценивается в течение 100 эпизодов и сравнивается с тремя классическими базовыми стратегиями: Static Equal, Static Priority и Greedy Proportional. Агент PPO снижает совокупную частоту нарушений SLA до 2×10^{-5}, что на три-четыре порядка ниже по сравнению с базовой стратегией Static Priority (0,46), при этом обеспечивая среднюю совокупную пропускную способность 74,7 Мбит/с и коэффициент использования ресурсов 98,6%. Анализ компромисса между пропускной способностью и частотой нарушений показывает, что обученная политика достигает рабочей точки, недоступной ни одной отдельной статической стратегии. Полная имитационная среда, процесс обучения и скрипты оценки являются открытыми и воспроизводимыми с использованием библиотеки Stable-Baselines3.

Ключевые слова: сетевой слайсинг 5G, обучение с подкреплением, Proximal Policy Optimization, соглашение об уровне обслуживания, распределение PRB, eMBB, URLLC, mMTC, управление радиоресурсами.

Introduction

Main Provisions

A Gymnasium-compatible simulation environment was developed for SLA-aware PRB allocation across eMBB, URLLC, and mMTC slices, with per-slice contracts aligned to the 3GPP TS 28.530 management framework. A PPO agent trained with a composite throughput-SLA-utilization reward was benchmarked over 100 episodes against Static Equal, Static Priority, and Greedy Proportional baselines. The learned policy reduced the aggregate SLA violation rate to 2×10^{-5} – three to four orders of magnitude below the Static Priority baseline – while sustaining 74.7 Mbit/s of throughput and 98.6% resource utilization, occupying an operating point on the throughput-violation frontier that no static rule reaches. These results establish reinforcement learning with an explicit, decomposed SLA objective as a practical path toward adaptive, contract-compliant network slice management.

Network slicing is one of the defining capabilities of fifth-generation (5G) mobile systems. Mobile data traffic continues to grow and is carried by increasingly diverse services with very different performance requirements [1, 2]. Slicing allows a single physical infrastructure to be partitioned into multiple logical networks, each tailored to a distinct class of service with its own performance contract [3, 4]. The three canonical slice types standardized for 5G – enhanced Mobile Broadband (eMBB), Ultra-Reliable Low-Latency Communications (URLLC), and massive Machine-Type Communications (mMTC) – impose fundamentally different and often conflicting demands on the radio scheduler [5, 6]. eMBB traffic seeks high sustained data rates, URLLC traffic requires stringent reliability and deadline guarantees, and mMTC traffic must connect a very large number of low-rate devices. Allocating the finite pool of Physical Resource Blocks (PRB) at the base station so that every slice meets its service-level agreement (SLA) is therefore a continuous, multi-objective decision problem under uncertainty.

Traditional slicing controllers rely on fixed allocation policies. Equal partitioning divides PRBs uniformly across slices, static priority schemes reserve a guaranteed share for the most critical slice, and traffic-proportional rules distribute resources according to instantaneous load. While simple and predictable, these heuristics cannot adapt to the non-stationary statistics of mixed 5G traffic: a policy tuned for one load regime systematically violates SLAs in another [7, 8]. The combinatorial structure of the allocation, together with the stochastic arrival of eMBB, URLLC, and mMTC demand, makes closed-form optimal control intractable for online deployment [9].

Deep reinforcement learning (DRL) has emerged as a powerful data-driven alternative for radio resource management. Following the demonstration that neural networks can learn control policies directly from interaction [10], a substantial body of work has applied DRL to slicing. Li et al. introduced DRL for resource management across slices [7], Raza et al. applied reinforcement learning to slice admission in a flexible 5G RAN [11], and distributional and generative extensions further improved sample efficiency [12]. Recent studies address end-to-end slicing with multi-agent actor-critic methods [13], slice scaling and placement [14], admission control [15, 16], and constrained formulations that embed SLA limits directly into the learning objective [17]. Comprehensive surveys confirm that learned policies consistently dominate fixed heuristics across heterogeneous network conditions [8, 14]. Among policy-gradient methods, Proximal Policy Optimization (PPO) [18] has become the algorithm of choice for slicing control because its clipped surrogate objective yields stable training over the large action spaces typical of PRB allocation, as demonstrated in O-RAN xApp deployments [19] and reward-shaped orchestration frameworks [20].

Despite this progress, several gaps persist. First, many DRL slicing studies optimize an abstract utility and do not verify performance against explicit, standardized SLA contracts, making the operational relevance of reported gains difficult to assess [8, 21]. Second, the per-slice nature of SLA satisfaction – distinguishing a throughput floor for eMBB, a reliability target for URLLC, and a served-device ratio for mMTC – is frequently collapsed into a single scalar, obscuring where violations actually occur. Third, systematic benchmarking against multiple interpretable static baselines within one reproducible environment remains uncommon [22]. The present work targets these gaps. The central research question is: can a PPO agent learn a slicing policy that simultaneously keeps the per-slice SLA violation rate near zero and sustains high aggregate throughput and resource utilization, outperforming Static Equal, Static Priority, and Greedy Proportional allocation within a 3GPP-aligned simulation environment? Our contributions are fourfold. (i) We develop a Gymnasium-compatible single-cell slicing environment in which eMBB, URLLC, and mMTC SLA requirements follow the 3GPP TS 28.530 management framework [3], with a Shannon-based throughput model driven by CQI-to-spectral-efficiency mapping. (ii) We design a PPO agent with a composite reward that jointly rewards normalized throughput, penalizes aggregate SLA violation, and rewards utilization. (iii) We benchmark the agent against three deterministic baselines over 100 evaluation episodes, reporting per-slice violation rates and PRB distributions. (iv) We show through a throughput-versus-violation trade-off analysis that the learned policy reaches an operating region inaccessible to any single static rule. All code and results are open-source and reproducible with Stable-Baselines3 [23].

Research methodology

Network and Traffic Model

We consider the downlink of a single 5G base station that serves three slices simultaneously. The cell owns a fixed pool of $N_{PRB} = 100$ Physical Resource Blocks, each spanning $B_{sc} = 180$ kHz (twelve 15-kHz subcarriers). Decisions are taken once per Transmission Time Interval (TTI); an episode spans 500 TTIs. At each TTI the traffic model draws the instantaneous demand of every slice from independent Poisson processes: the number of active eMBB, URLLC, and mMTC users follows Poisson distributions with means of 20, 5, and 50, respectively, reflecting a broadband-moderate, critical-sparse, and massive-but-light demand profile. Each user is assigned a Channel Quality

Indicator (CQI) drawn from a slice-specific range – high for URLLC, moderate for eMBB, and low for mMTC – capturing the typical propagation conditions of each service class.

The spectral efficiency of a user with channel quality index q is obtained from the 3GPP CQI-to-SNR mapping [24] followed by the Shannon bound as shown in Equation 1:

$$SE(q) = log_2(1 + 10\text{^}(SNR(q)/10)) \quad (1)$$

where SNR(q) is the signal-to-noise ratio tabulated for CQI index q. The throughput (Equation 2) delivered to a slice that receives k PRBs whose users have mean spectral efficiency $S\bar{E}$ is then:

$$R = k \cdot B_sc \cdot S\bar{E} \quad [bit/s] \quad (2)$$

where B_sc = 180 kHz is the PRB bandwidth. The theoretical cell capacity used to normalize throughput in the reward, denoted R_max, is obtained by assigning all PRBs at the maximum CQI across the three slices.

Per-Slice SLA Requirements (3GPP TS 28.530)

Each slice carries an SLA contract expressed in the terms of the 3GPP TS 28.530 management framework [3]. The eMBB slice requires a minimum throughput; the URLLC slice requires a minimum reliability under a one-TTI deadline; the mMTC slice requires a minimum fraction of served devices. Formally, for the eMBB slice the violation indicator is shown in Equation 3:

$$v_{eMBB} = \mathbb{1}[R_{eMBB} < R_{min}], \;\; R_{min} = 10\ Mbit/s \quad (3)$$

For the URLLC slice, let n_p be the number of critical packets generated in the current TTI and s the number served within the deadline (Equation 4). The reliability and its violation indicator are:

$$\rho = 1 - \frac{max(0, n_p - s)}{n_p}, \;\; v_{URLLC} = \mathbb{1}[\rho < 0.999] \quad (4)$$

For the mMTC slice, where each PRB serves on average two low-rate devices, the served-device ratio (Equation 5) and its violation indicator are:

$$\sigma = min\left(1, \frac{2 \cdot k_{mMTC}}{n_{mMTC}}\right), \;\; v_{mMTC} = \mathbb{1}[\sigma < 0.80] \quad (5)$$

MDP Formulation

The slicing-control task is cast as a Markov Decision Process (S, A, P, R, γ). The state observed by the agent at each TTI is a 12-dimensional normalized vector which is shown in Equation 6:

$$s = (\ell_{eMBB}, \ell_{URLLC}, \ell_{mMTC}, k', v', \bar{q}_{eMBB}, \bar{q}_{URLLC}, \bar{q}_{mMTC}) \quad (6)$$

comprising the normalized per-slice load ℓ, the previous PRB allocation k′ and previous violation flags v′ (three components each), and the mean CQI of every slice. The action is a vector of requested PRB counts, one per slice (Equation 7):

$$a = (k_{eMBB}, k_{URLLC}, k_{mMTC}) \in \{0, \ldots, 100\}^3 \quad (7)$$

represented as a MultiDiscrete space and rescaled so that the total never exceeds N_PRB = 100. The composite reward as shown in Equation 8, couples the three design goals – throughput, SLA compliance, and utilization:

$$r = 0.4 \cdot \frac{R_{tot}}{R_{max}} - 0.4 \cdot \frac{v_{eMBB} + v_{URLLC} + v_{mMTC}}{3} + 0.2 \cdot \frac{k_{eMBB} + k_{URLLC} + k_{mMTC}}{N_{PRB}} \quad (8)$$

where R_tot is the aggregate delivered throughput and R_max the normalizing capacity. The discount factor is $\gamma = 0.99$. This reward deliberately balances a reward term (throughput), a penalty term (SLA violation), and a regularizer (utilization), so that the agent is driven toward high-rate yet contract-compliant allocations rather than extreme ones.

PPO Agent and Baseline Policies

The control policy is learned with Proximal Policy Optimization [18], an on-policy actor-critic algorithm whose clipped surrogate objective bounds the policy update at each step and thereby stabilizes training over the large MultiDiscrete action space of Equation (7). The policy and value networks are multilayer perceptrons with two hidden layers of 64 units and tanh activations. The agent is implemented with Stable-Baselines3 [23]. Table 1 lists the hyperparameters.

We compare the learned policy against three deterministic baselines. Static Equal partitions the 100 PRBs as evenly as possible across the three slices (33/33/34) at every TTI, independent of load or channel. Static Priority reserves a fixed, generous share for the throughput- and reliability-critical slices (eMBB and URLLC) and assigns the remainder to mMTC, reflecting an operator-priority rule. Greedy Proportional distributes PRBs in proportion to the instantaneous per-slice load, a demand-following heuristic.

Table 1. PPO hyperparameters

Hyperparameter	*Value*
Policy network	*MLP, two layers of 64 units, tanh*
Total timesteps	*1,000,000*
Parallel environments	*4*
Learning rate	3×10^{-4}
Rollout length (n_steps)	*2048*
Mini-batch size	*256*
Entropy coefficient	*0.01*
Discount factor (γ)	*0.99*
Random seeds	*0, 1, 2*

Experimental Setup

The PPO agent is trained for one million environment steps with four parallel environments and three independent random seeds (0–2) to quantify run-to-run variability. After training, every method – the PPO agent and the three baselines – is evaluated over 100 independent episodes of 500 TTIs each. For each method we record the mean aggregate throughput, the aggregate and per-slice SLA violation rates, the resource utilization, and the mean PRB allocation per slice. The environment, traffic model, SLA checker, training loop, and evaluation pipeline are released as open-source code; all reported numbers are reproducible from the accompanying repository.

Results of the Study

Figure 1 shows the PPO training reward as a function of timesteps, averaged over the three seeds with a 95% confidence band. The agent improves rapidly during the first 100,000 steps as it learns to read the per-slice load and CQI components of the observation, and then settles onto a stable plateau. The final mean reward across seeds is 134.9 ± 0.3, and the narrow confidence band throughout the steady-state phase indicates that learning is robust to the stochastic traffic and channel draws.

Table 2 summarizes the performance of all four methods over the 100 evaluation episodes. The PPO agent drives the aggregate SLA violation rate down to 2×10^{-5} while sustaining 74.7 Mbit/s of throughput and 98.6% utilization. The Static Priority baseline attains the highest raw throughput (79.5 Mbit/s) but at a catastrophic SLA violation rate of 0.46, because its rigid reservation starves the mMTC slice whenever device counts are high. Static Equal achieves zero violations but leaves throughput on the table (69.8 Mbit/s), since it cannot shift resources toward the slices that need them. Greedy Proportional follows demand and keeps violations low (5.3×10^{-3}) but still trails the learned policy on both throughput and reliability.

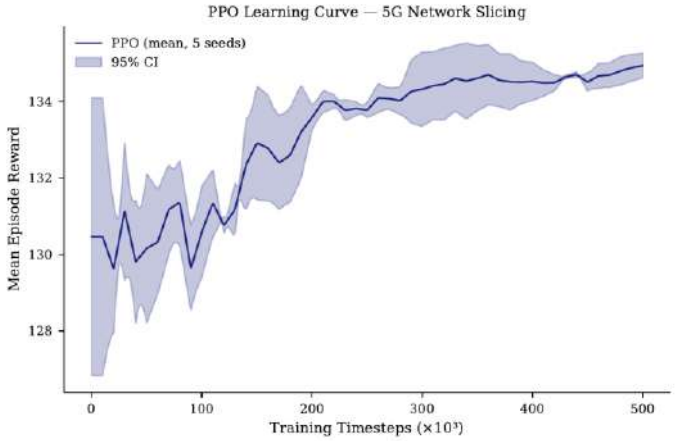


Figure 1. PPO training reward versus timesteps, averaged over three seeds (shaded area: 95% confidence interval)

Table 2. Aggregate performance over 100 evaluation episodes

Method	*Throughput (Mbit/s)*	*SLA violation rate*	*Utilization*	*PRB eMBB/URLLC/mMTC*
PPO (ours)	*74.7*	*2×10^{-5}*	*0.986*	*28.9 / 42.2 / 27.6*
Static Equal	*69.8*	*0.000*	*1.000*	*33.0 / 33.0 / 34.0*
Static Priority	*79.5*	*0.464*	*1.000*	*40.0 / 40.0 / 20.0*
Greedy Proportional	*66.8*	*5.3×10^{-3}*	*1.000*	*35.2 / 27.8 / 37.0*

Figure 2 breaks the violation rate down by slice. The static rules concentrate their failures in a single slice – Static Priority on mMTC, Greedy Proportional spread thinly across eMBB and URLLC – whereas the PPO agent keeps all three per-slice rates at or near zero, with only a negligible residual on mMTC. This is the practical meaning of SLA-aware control: not a lower average, but the elimination of the systematic per-slice failures that fixed rules incur.

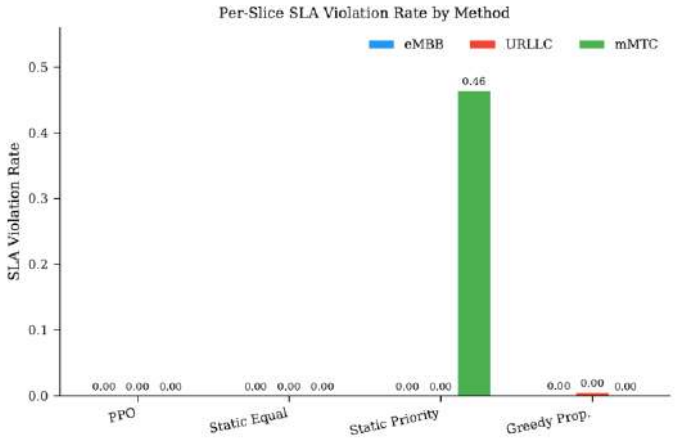


Figure 2. Per-slice SLA violation rate by method

Figure 3 plots throughput against the aggregate SLA violation rate, exposing the trade-off frontier directly. Static Priority sits far to the right (high throughput, high violation); Static Equal sits at the bottom (zero violation, modest throughput). The PPO agent occupies the desirable upper-left region – high throughput at near-zero violation – an operating point that none of the static policies can reach, confirming that the learned policy is not merely interpolating between baselines but discovering a qualitatively better allocation strategy.

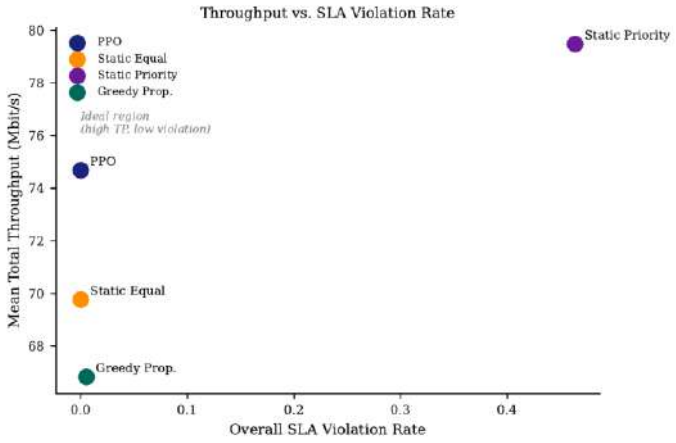


Figure 3. Throughput versus aggregate SLA violation trade-off across methods.

Figure 4 shows the mean PRB distribution chosen by each method.

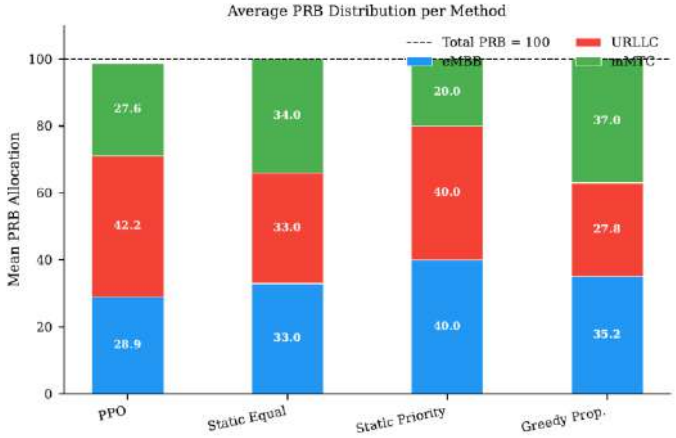


Figure 4.Mean PRB allocation per slice for each method

The PPO agent learns an asymmetric split that allocates the largest share to URLLC (≈42 PRBs) to protect its reliability deadline, while balancing eMBB and mMTC near the levels each needs to clear its SLA floor. This learned allocation differs markedly from the uniform split of Static Equal and the rigid reservation of Static Priority and explains how the agent simultaneously satisfies three heterogeneous contracts.

Discussion

The results support a clear conclusion: when SLA satisfaction is an explicit objective, a learned policy substantially outperforms interpretable static rules without sacrificing efficiency. The PPO agent matches the violation-free behavior of Static Equal while delivering roughly 7% more throughput, and it matches the high-throughput regime of Static Priority while reducing SLA violations by more than four orders of magnitude. The key enabler is the composite reward of Equation (8), which exposes all three objectives to the agent at once and lets it learn a load- and channel-dependent allocation rather than committing to a single fixed split. A limitation of the present study is its single-cell, single-base-station scope and its simulation-based traffic model; extending the approach to multi-cell coordination and to traces collected from real deployments is a natural direction for future work, as is the incorporation of hard constraints through constrained-RL formulations [17].

Conclusion

We presented a PPO-based, SLA-aware approach to 5G network slicing that dynamically allocates Physical Resource Blocks among eMBB, URLLC, and mMTC slices subject to per-slice contracts aligned with 3GPP TS 28.530. In a reproducible Gymnasium environment, the learned agent reduced the aggregate SLA violation rate to 2×10^{-5} while sustaining 74.7 Mbit/s of throughput and 98.6% utilization, dominating Static Equal, Static Priority, and Greedy Proportional baselines on the throughput-versus-violation frontier. The trade-off analysis confirmed that the policy reaches an operating point inaccessible to any single static rule. These findings indicate that reinforcement learning with an explicit, decomposed SLA objective is a practical path toward adaptive, contract-compliant slice management. Future work will extend the framework to multi-cell settings, real traffic traces, and constrained-RL formulations that encode SLA limits as hard constraints.

Acknowledgment

This research was funded by the Committee of Science of the Ministry of Science and Higher Education of the Republic of Kazakhstan (Grant No. BR24993211 ‘Development of Solutions for 5G Networks under the ORAN Concept and Efficient Routing and Load Balancing in SDN Environments for Kazakhstan’).

References

[1] Ericsson, "Ericsson Mobility Report," Ericsson AB, Stockholm, Sweden, Tech. Rep., Jun. 2024. [Online]. Available: https://www.ericsson.com/en/reports-and-papers/mobility-report

[2] ITU-R, "IMT traffic estimates for the years 2020 to 2030," Report ITU-R M.2370-0, International Telecommunication Union, Geneva, 2015. [Online]. Available: https://www.itu.int/pub/R-REP-M.2370

[3] 3GPP, "Management and orchestration; Concepts, use cases and requirements," 3GPP TS 28.530, v18.0.0, 2024. [Online]. Available: https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3273

[4] 3GPP, "System architecture for the 5G System (5GS)," 3GPP TS 23.501, v18.0.0, 2024. [Online]. Available: https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3144

[5] X. Foukas, G. Patounas, A. Elmokashfi, and M. K. Marina, "Network slicing in 5G: Survey and challenges," IEEE Commun. Mag., vol. 55, no. 5, pp. 94–100, 2017, doi: 10.1109/MCOM.2017.1600951.

[6] I. Afolabi, T. Taleb, K. Samdanis, A. Ksentini, and H. Flinck, "Network slicing and softwarization: A survey on principles, enabling technologies, and solutions," IEEE Commun. Surveys Tuts., vol. 20, no. 3, pp. 2429–2453, 2018, doi: 10.1109/COMST.2018.2815638.

[7] R. Li, Z. Zhao, Q. Sun, C.-L. I, C. Yang, X. Chen, M. Zhao, and H. Zhang, "Deep reinforcement learning for resource management in network slicing," IEEE Access, vol. 6, pp. 74429–74441, 2018, doi: 10.1109/ACCESS.2018.2881964.

[8] S. Ebrahimi, F. Bouali, and O. C. L. Haas, "Resource management from single-domain 5G to end-to-end 6G network slicing: A survey," IEEE Commun. Surveys Tuts., vol. 26, no. 4, pp. 2836–2866, 2024, doi: 10.1109/COMST.2024.3390613.

[9] N. Van Huynh, D. T. Hoang, D. N. Nguyen, and E. Dutkiewicz, "Optimal and fast real-time resource slicing with deep dueling neural networks," IEEE J. Sel. Areas Commun., vol. 37, no. 6, pp. 1455–1470, 2019, doi: 10.1109/JSAC.2019.2904371.

[10] V. Mnih et al., "Human-level control through deep reinforcement learning," Nature, vol. 518, no. 7540, pp. 529–533, 2015, doi: 10.1038/nature14236.

[11] M. R. Raza, C. Natalino, P. Öhlen, L. Wosinska, and P. Monti, "Reinforcement learning for slicing in a 5G flexible RAN," J. Lightw. Technol., vol. 37, no. 20, pp. 5161–5169, 2019, doi: 10.1109/JLT.2019.2924345.

[12] Y. Hua, R. Li, Z. Zhao, X. Chen, and H. Zhang, "GAN-powered deep distributional reinforcement learning for resource management in network slicing," IEEE J. Sel. Areas Commun., vol. 38, no. 2, pp. 334–349, 2020, doi: 10.1109/JSAC.2019.2959185.

[13] M. A. Ejaz, G. Wu, and T. Iqbal, "Dynamic and efficient resource allocation for 5G end-to-end network slicing: A multi-agent deep reinforcement learning approach," Int. J. Commun. Syst., vol. 37, no. 17, art. e5916, 2024, doi: 10.1002/dac.5916.

[14] N. Saha, M. Zangooei, M. Golkarifard, and R. Boutaba, "Deep reinforcement learning approaches to network slice scaling and placement: A survey," IEEE Commun. Mag., vol. 61, no. 2, pp. 82–87, 2023. [Online]. Available: https://ieeexplore.ieee.org/document/10004899

[15] M. Sulaiman, A. Moayyedi, M. Ahmadi, M. A. Salahuddin, R. Boutaba, and A. Saleh, "Coordinated slicing and admission control using multi-agent deep reinforcement learning," IEEE Trans. Netw. Serv. Manage., vol. 20, no. 2, pp. 1110–1124, 2023, doi: 10.1109/TNSM.2022.3222589.

[16] M. Khani, S. Jamali, M. K. Sohrabi, M. M. Sadr, and A. Ghaffari, "Slice admission control in 5G cloud radio access network using deep reinforcement learning: A survey," Int. J. Commun. Syst., vol. 37, no. 13, art. e5857, 2024, doi: 10.1002/dac.5857.

[17] M. Sulaiman, M. Ahmadi, M. A. Salahuddin, R. Boutaba, and A. Saleh, "Generalizable resource scaling of 5G slices using constrained reinforcement learning," in Proc. IEEE/IFIP Netw. Oper. Manage. Symp. (NOMS), Miami, FL, USA, 2023, pp. 1–9, doi: 10.1109/NOMS56928.2023.10154267.

[18] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "Proximal policy optimization algorithms," arXiv preprint arXiv:1707.06347, 2017, doi: 10.48550/arXiv.1707.06347.

[19] R. Barker, A. Ebrahimi Dorcheh, T. Seyfi, and F. Afghah, "REAL: Reinforcement learning-enabled xApps for experimental closed-loop optimization in O-RAN with OSC RIC and srsRAN," in Proc. IEEE Int. Conf. Commun. Workshops (ICC Workshops), 2025, doi: 10.1109/ICCWorkshops67674.2025.11162144.

[20] N. Shezi, B. Nleya, and B. Pule, "Slice-aware and computationally efficient resource orchestration for converged mmWave–PON O-RAN: A reward-shaped PPO approach for joint DBA and PRB allocation," Telecom, vol. 7, no. 3, art. 75, 2026, doi: 10.3390/telecom7030075.

[21] Y. Azimi, S. Yousefi, H. Kalbkhani, and T. Kunz, "Energy-efficient deep reinforcement learning assisted resource allocation for 5G-RAN slicing," IEEE Trans. Veh. Technol., vol. 71, no. 1, pp. 856–871, 2022, doi: 10.1109/TVT.2021.3128513.

[22] S. Choi, S. Choi, G. Lee, S.-G. Yoon, and S. Bahk, "Deep reinforcement learning for scalable dynamic bandwidth allocation in RAN slicing with highly mobile users," IEEE Trans. Veh. Technol., vol. 73, no. 1, pp. 576–590, 2024, doi: 10.1109/TVT.2023.3302416.

[23] A. Raffin, A. Hill, A. Gleave, A. Kanervisto, M. Ernestus, and N. Dormann, "Stable-Baselines3: Reliable reinforcement learning implementations," J. Mach. Learn. Res., vol. 22, no. 268, pp. 1–8, 2021. [Online]. Available: https://jmlr.org/papers/v22/20-1364.html

[24] 3GPP, "Evolved Universal Terrestrial Radio Access (E-UTRA); Physical layer procedures," 3GPP TS 36.213, v17.0.0, 2022. [Online]. Available: https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=2427

МРНТИ 41.03

https://doi.org/10.51889/2959-5894.2026.95.3.007

М.Дж. Минглибаев[1], М.Р. Сапарова[2*]

[1]Казахский национальный университет имени аль-Фараби, г.Алматы, Казахстан
[2]Институт механики и машиноведения имени У.А. Джолдасбекова, г.Алматы, Казахстан
**e-mail: moldir170788@gmail.com*

УРАВНЕНИЯ ДВИЖЕНИЯ В ОСКУЛИРУЮЩИХ ПЕРЕМЕННЫХ В КООРДИНАТАХ ЯКОБИ

Аннотация

В данной работе рассматривается проблема описания уравнений движения в координатах Якоби через оскулирующие элементы как эффективный метод решения задач многих тел. Составлены уравнения движения семипланетной экзопланетной системы TRAPPIST-1 с переменной массой и неизотропным излучением в условиях действия реактивных сил с использованием координат Якоби. В задачах многих тел использование традиционных гелиоцентрических координат значительно усложняет вычисления. Преимущество координат Якоби заключается в том, что движение каждого тела рассматривается относительно центра масс предшествующих тел, что позволяет адекватно описать иерархическую систему многих тел, удобную для использования каноническую теорию возмущения. Тела взаимодействуют согласно закону всемирного тяготения Ньютона, при этом наличие переменной массы обуславливает возникновение реактивных сил, усложняющих динамику системы. Основная цель исследования – построение системы дифференциальных уравнений в оскулирующих аналогах элементов Делоне, которые являются переменными «действие-угол» исходя из уравнений в координатах Якоби. Использование координат Якоби позволяет работать с одной силовой функции для всех тел, упрощает математическую структуру задачи. В то время как оскулирующие переменные позволяют наглядно интерпретировать геометрическую картину движения. На основе уравнений движения, полученных в координатах Якоби, разработана каноническая теория возмущений, и выведены уравнения возмущенного движения в аналогах элементов Делоне. Данный подход обеспечивает лучшую точность расчетов динамики многопланетных систем, благодаря свойствам «действие-угол» аналогов элементов Делоне. Полученные уравнения позволяют проводить дальнейшие исследования системы в барицентрической системе координат с применением теории возмущений.

Ключевые слова: оскулирующие переменные, система координат Якоби, элементы Делоне, каноническая теория возмущений.

М.Дж. Минглибаев[1], М.Р.Сапарова[2]
[1]әл-Фараби атындағы Қазақ Ұлттық университеті, Алматы, Қазақстан
[2]Ө.А.Жолдасбеков атындағы механика және машинатану институты, Алматы, Қазақстан

ЯКОБИ КООРДИНАТТАРЫНДАҒЫ ОСКУЛЯЦИЯЛАНАТЫН АЙНЫМАЛЫЛАРДАҒЫ ҚОЗҒАЛЫС ТЕҢДЕУЛЕРІ

Аңдатпа

Жұмыста көп денелі мәселелерді шешудің тиімді әдісі ретінде Якоби координаттарындағы қозғалыс теңдеулерін оскуляциялық элементтер арқылы сипаттау мәселесі қарастырылады. Айнымалы массалы, изотропты емес TRAPPIST-1 жеті планеталы экзопланеталық жүйенің реактивті күштер бар болған кездегі Якоби координатасын пайдалана отырып, оскуляция элементтері арқылы қозғалыс теңдеуі жазылған. Көп денелер есебінде әдеттегі гелиоцентрлік координаттарды пайдалану есептеуді қиындатады. Якоби координаттарының артықшылығы – әрбір дененің қозғалысы алдыңғы денелердің массалар центріне қатысты қарастырылады, бұл иерархиялық көп дене жүйесін адекватты сипаттауға мүмкіндік береді және канондық қозғалыстар теориясын қолдануға ыңғайлы. Денелер Ньютонның бүкіләлемдік тартылыс заңына сәйкес өзара тартылады. Массаның айнымалы болуы есебінен, есепті күрделендіретін реактивті күштер бар. Зерттеудің негізгі мақсаты – Якоби координаттарындағы теңдеулерге сүйене отырып, «әсер-бұрыш» айнымалылары болып табылатын Делоне элементтерінің

оскуляциялық аналогтарындағы дифференциалдық теңдеулер жүйесін құру. Якоби координаттарын қолдану барлық денелер үшін бір ғана күш функциясымен жұмыс істеуге мүмкіндік береді және есептің математикалық құрылымын қарапайымдыландырады. Ал оскуляциялық айнымалылар қозғалыстың геометриялық көрінісін көрнекі түрде түсіндіруге мүмкіндік береді.Якоби координаттарында алынған қозғалыс теңдеулері негізінде канондық қозғалыстар теориясы жасалып, Делоне элементтерінің аналогтарындағы қозғалыс теңдеулері қорытылып шығарылды. Бұл тәсіл Делоне элементтері аналогтарының «әсер-бұрыш» қасиеттерінің арқасында көп планеталы жүйелер динамикасын есептеудің жоғары дәлдігін қамтамасыз етеді. Алынған теңдеулер қозғалыстар теориясын қолдана отырып, жүйені барицентрлік координаттар жүйесінде әрі қарай зерттеуге мүмкіндік береді.

Түйін сөздер: оскуляцияланатын айнымалылар, Якоби координаттар жүйесі, Делоне элементтері, канондық ұйытқу теориясы.

M.Zh. Minglibaev[1], M.R. Saparova[2]
[1]Al-Farabi Kazakh National University, Almaty, Kazakhstan
[2]Institute of Mechanics and Machine Science named after U.A. Joldasbekov, Almaty, Kazakhstan

EQUATIONS OF MOTION IN OSCULATING VARIABLES WITHIN JACOBI COORDINATES

Abstract

This paper addresses the problem of describing the equations of motion in Jacobi coordinates using osculating elements as an effective method for solving many-body problems. The equations of motion for the seven-planet exoplanetary system TRAPPIST-1 with variable mass and non-isotropic radiation under the influence of reactive forces are formulated using Jacobi coordinates. In many-body problems, the use of traditional heliocentric coordinates significantly complicates calculations. The advantage of Jacobi coordinates lies in the fact that the motion of each body is considered relative to the center of mass of the preceding bodies. This allows for an adequate description of a hierarchical many-body system, which is convenient for applying the canonical perturbation theory. The bodies interact according to Newton's law of universal gravitation, while the presence of variable mass gives rise to reactive forces that complicate the system's dynamics. The primary objective of this research is to construct a system of differential equations in osculating analogues of Delaunay elements, which serve as «action-angle» variables derived from the equations in Jacobi coordinates. The utilization of Jacobi coordinates enables the use of a single force function for all bodies, simplifying the mathematical structure of the problem. Meanwhile, the osculating variables provide a visual and clear interpretation of the geometric picture of motion. Based on the equations of motion obtained in Jacobi coordinates, a canonical perturbation theory has been developed, and the equations of perturbed motion in analogues of Delaunay elements have been derived. Thanks to the action-angle properties of the Delaunay element analogues, this approach ensures better accuracy in calculating the dynamics of multi-planetary systems. The resulting equations enable further investigation of the system in the barycentric coordinate system using perturbation theory.

Keywords: osculating variables, Jacobi coordinate system, Delaunay elements, canonical perturbation theory.

Введение

Стремительное развитие современной астрофизики, особенно открытие экзопланетных систем (таких как TRAPPIST-1, Kepler и др.), требует новых методов исследования динамической эволюции нестационарных систем многих тел. Описание долгосрочной устойчивости планет и изменений их орбитальных элементов в таких системах остается одной из самых актуальных задач небесной механики. Представим обзор работ ряда ученых.

Прежде всего, представим краткий обзор задачи многих тел для случая постоянных масс. В работе Перминова и Кузнецова [1] предложен современный высокотехнологичный подход к построению аналитической теории, описывающей движение планет-гигантов Солнечной системы при условии постоянства их масс. Движение четырехпланетной системы исследовано в системе координат Якоби. Основной целью исследования является поиск эффективного способа долгосрочного прогнозирования сложных колебаний, вызванных взаимным притяжением планет, путем математического осреднения. Авторы использовали один из

наиболее мощных инструментов гамильтоновой механики – метод Хори–Депри, основанный на преобразованиях Ли. Этот метод позволяет исключить из уравнений движения короткопериодические возмущения, что значительно упрощает изучение вековой динамики планет и позволяет существенно увеличить шаг времени при численном интегрировании. В качестве математического аппарата были выбраны переменные Пуанкаре второй системы, а Гамильтониан был разложен в ряды Пуассона с точностью до третьей степени малого параметра.

Еще одна работа Перминова и Кузнецова [2], опубликованная в 2020 году, посвящена изучению орбитальной эволюции четырех планет-гигантов Солнечной системы (Юпитера, Сатурна, Урана и Нептуна) на сверхдлительных интервалах времени (до 10 миллиардов лет) в условиях постоянства их масс. Авторами разработана полуаналитическая осредненная теория движения, обладающая точностью третьего порядка относительно масс планет и шестого порядка – относительно эксцентриситетов и наклонений орбит. Для разложения Гамильтониана использовались система координат Якоби и вторая система элементов Пуанкаре. В ходе исследования полученные уравнения движения в осредненных элементах были численно проинтегрированы различными методами на интервале в 10 миллиардов лет. Результаты моделирования показали, что движение планет носит квазипериодический характер, а короткопериодические колебания орбитальных элементов сохраняют крайне малые значения на протяжении всего исследуемого процесса.

Работа испанского ученого Давида Арнаса [3] посвящена аналитическому исследованию влияния пертурбации J_2, возникающей вследствие полярного сжатия центрального небесного тела, на динамику орбит. Массы тел в системе рассматриваются как постоянные. Автор предлагает метод разложения всех переменных системы в степенные ряды по константе J_2, а также способ регуляризации времени, основанный на аргументе широты. Ключевой особенностью данной методики является возможность получения точных аналитических решений без необходимости специального контроля возмущенной частоты системы. Предложенный подход позволяет аппроксимировать динамику орбит с любым значением эксцентриситета в оскулирующих элементах, а также получать аналитические преобразования для перехода от оскулирующих элементов к осредненным для данных орбит.

В работе [4], посвященной прогнозированию движения системы из N небесных тел под действием сил взаимного тяготения, предложен метод получения точных решений с помощью нейронных сетей (NN). Ввиду отсутствия общего аналитического решения для задачи $N > 2$ тел, авторы применили подход, основанный на декомпозиции гамильтониана на аналитически решаемую двухтельную часть и интерактивную часть, аппроксимируемую нейронной сетью. Предложенный нейронный симплектический интегратор N тел обеспечил высокую точность вычислений для системы трех тел с постоянными массами на интервале 10^5 шагов, не уступая традиционным численным методам. Ключевым достижением этой работы является способность нейронной сети соблюдать физические законы сохранения, такие как энергия и момент импульса, в долгосрочной перспективе, а также наличие индуктивного смещения, позволяющего успешно прогнозировать эволюцию новых систем, не входивших в обучающую выборку.

В работах Холшевникова, Гребеникова и Кузнецова [5–8] исследована орбитальная эволюция системы трех тел «Солнце – Юпитер – Сатурн» при условии постоянства их масс. Гамильтониан, записанный в оскулирующих элементах, был разложен в ряды Пуассона по всем переменным. Усредненные уравнения движения для системы двух планет были численно проинтегрированы на временном интервале в 10 млрд лет. Установлено, что эксцентриситеты и наклонения орбит Юпитера и Сатурна остаются малыми ненулевыми величинами. В результате выявлено, что движение планет носит квазипериодический характер, а короткопериодические возмущения сохраняют малые значения на всем рассматриваемом интервале времени.

Ввиду отсутствия общего решения задачи Гюльдена-Мещерского относительно закономерностей изменения массы, динамика систем планет переменной массы исследовалась методами канонической теории возмущений. Изменение масс в нестационарных планетных системах оказывает прямое влияние на их динамическую эволюцию. В связи с этим движение нестационарных тел рассматривалось с применением специальной канонической теории возмущений [9]

Поскольку массы небесных тел изменяются с течением времени, исследование влияния таких изменений на орбитальные параметры в рамках задачи трех тел представляет значительный интерес. В работе [10] рассмотрен частный случай, когда массы двух тел изменяются изотропно по единому закону. Для решения уравнений движения использованы оскулирующие элементы апериодического квазиконического движения. Научная значимость исследования заключается в разработке алгоритма расчета долгопериодических, или вековых возмущений. Вычислив вековую часть функций возмущения, авторы вывели дифференциальные уравнения, определяющие эволюцию орбитальных элементов.

Главная особенность и научная ценность работы [11] заключаются в рассмотрении случая, когда массы тел изменяются с различной интенсивностью и неизотропно. Такое неравномерное изменение массы приводит к возникновению в системе реактивных сил, что значительно усложняет динамику движения тел. Принимая произвольный характер законов изменения масс, авторы с помощью методов теории возмущений вывели канонические уравнения вековых возмущений данной системы.

В научной работе [12] исследованы вопросы существования и устойчивости внеплоскостных точек равновесия экзопланет в системе двойных звезд с переменными массами, окруженных кластером материальных точек. Основной особенностью исследования является допущение о том, что массы всех тел в системе изменяются с одинаковой скоростью в соответствии с единым законом Мещерского. Автономные уравнения движения, полученные в рамках задачи Гюльдена – Мещерского, с помощью специальных преобразований были приведены к автономной системе с постоянными коэффициентами для последующего анализа. Авторы провели проверку устойчивости точек равновесия для автономных и неавтономных систем и установили, что во всех рассматриваемых случаях данные точки являются неустойчивыми.

В исследовании [13], проведенном при участии Они Леке (Oni Leke), анализируется устойчивость точек равновесия в ограниченной задаче трех тел с переменными массами при наличии тормозящей силы Пойнтинга – Робертсона (P-R). Движение основных тел описывается в рамках задачи Гюльдена – Мещерского, а изменение массы – единым законом Мещерского. Кроме того, при выводе динамических уравнений для автономных и неавтономных систем учитывались излучение двойной звезды, доплеровское смещение и перераспределение энергии. Анализ кривых нулевой скорости (ZVC) показал, что под воздействием радиационного давления область движения незначительно расширяется, тогда как с увеличением массы «запрещенные зоны» расширяются. При расчетах в оскулирующих элементах мы отслеживаем не "быстрые" переменные, а "медленно" меняющиеся параметры (форму орбиты). Это позволяет повысить точность вычислений и использовать более крупный шаг интегрирования. Применение данного метода, особенно в сочетании с гамильтоновой механикой и каноническими преобразованиями, является наиболее эффективным подходом для таких сложных систем, как TRAPPIST-1. Метод крайне удобен для использования канонической теории возмущений. Запись гамильтониана в координатах Якоби значительно упрощает разложение функции возмущения.

Методология исследования

Оскулирующие переменные в координатах Якоби. Используя уравнение движения, определенное в [14, формула 15] для неизотропной экзопланетной системы TPAPPIST-1 с переменной массой и семью планетами в присутствии реактивных сил,

$$\mu_i \ddot{\vec{r}}_i = grad_{\vec{r}_i} U - \mu_i \vec{\Phi}_i + \mu_i \vec{F}_i; \tag{1}$$

где $$\vec{\Phi}_i = \sum_{i=1}^{7} \left(2\dot{\nu}_i \dot{\vec{r}}_i + \ddot{\nu}_i \vec{r}_i \right); \qquad i = 1, 2, ..., 7 \tag{2}$$

перепишем в виде

$$\mu_i \ddot{\vec{r}}_i = grad_{\vec{r}_i} \left(f \frac{m_i \sigma_{i-1}}{r_i} \right) + b_i \vec{r}_i + grad_{\vec{r}_i} B_i, \tag{3}$$

где

$$B_i = U - V_i - f \frac{m_i \sigma_{i-1}}{r_i} - \frac{1}{2} b_i r_i^2 + A_i. \tag{4}$$

$$U = f \sum_{i,j} \frac{m_i m_j}{r_{ij}}, \; i \neq j \tag{5}$$

(2) $\rightarrow$ $\vec{\Phi}_i = V_i$;

$$V_i = \mu_i \left\{ \sum_{i=1}^{7} \left[\left(2\dot{\nu}_i \dot{x}_i + \ddot{\nu}_i x_i \right) x_j + \left(2\dot{\nu}_i \dot{y}_i + \ddot{\nu}_i y_i \right) y_j + \left(2\dot{\nu}_i \dot{z}_i + \ddot{\nu}_i z_i \right) z_j \right] \right\}, \tag{6}$$

$$b = \mu_i \frac{\ddot{\gamma}_i}{\gamma_i}, \quad \gamma_i = \gamma_i(t) = \frac{m_i(t_0)}{m_i(t)} \cdot \frac{\sigma_{i-1}(t_0)}{\sigma_{i-1}(t)} \cdot \frac{\mu_i(t)}{\mu_i(t_0)} = \frac{\sigma_i(t_0)}{\sigma_i(t)}, \tag{7}$$

$$A_i = \mu_i \sum_{i=1}^{7} \left(F_{ix} x_i + F_{iy} y_i + F_{iz} z_i \right), \text{- реактивная сила.} \tag{8}$$

В уравнениях (3) переходим к квазисферическим координатам $\rho_i, \varphi_i, \theta_i \; (i = 1, 2, ..., n)$

$$x_i = \gamma_i \rho_i \cos\varphi_i \cos\theta_i, \quad y_i = \gamma_i \rho_i \cos\varphi_i \cos\theta_i, \quad z_i = \gamma_i \rho_i \sin\varphi_i, \tag{9}$$

$$\gamma_i = \gamma_i(t) = \frac{m_i(t_0)}{m_i(t)} \cdot \frac{\sigma_{i-1}(t_0)}{\sigma_{i-1}(t)} \cdot \frac{\mu_i(t)}{\mu_i(t_0)} = \frac{\sigma_i(t_0)}{\sigma_i(t)}, \tag{10}$$

Так как

$$\sigma_i(t) = \sigma_i = m_0 + m_1 + m_2 + ... + m_i.$$

Тогда уравнения движения в канонических переменных [9]

$$\begin{aligned} \rho_i, \quad & P_{\rho_i} = \frac{\partial K_i}{\partial \dot{\rho}_i} = \mu_i \gamma_i^2 \dot{\rho}_i + \mu_i \gamma_i \dot{\gamma}_i \rho_i, \\ \varphi_i, \quad & P_{\varphi_i} = \frac{\partial K_i}{\partial \dot{\varphi}_i} = \mu_i \gamma_i^2 \rho_i^2 \dot{\varphi}_i, \\ \theta_i, \quad & P_{\theta_i} = \frac{\partial K_i}{\partial \dot{\theta}_i} = \mu_i \gamma_i^2 \rho_i^2 \cos^2 \varphi_i \dot{\theta}_i, \end{aligned} \tag{11}$$

можно написать в следующей форме

$$\dot{\rho}_i = \frac{\partial H_i}{\partial P_{\rho_i}}, \quad \dot{P}_{\rho_i} = -\frac{\partial H_i}{\partial \rho_i} + \frac{\dot{\mu}_i}{\mu_i} P_{\rho_i},$$
$$\dot{\varphi}_i = \frac{\partial H_i}{\partial P_{\varphi_i}}, \quad \dot{P}_{\varphi_i} = -\frac{\partial H_i}{\partial \varphi_i} + \frac{\dot{\mu}_i}{\mu_i} P_{\varphi_i}, \tag{12}$$
$$\dot{\theta}_i = \frac{\partial H_i}{\partial P_{\theta_i}}, \quad \dot{P}_{\theta_i} = -\frac{\partial H_i}{\partial \theta_i} + \frac{\dot{\mu}_i}{\mu_i} P_{\theta_i},$$

соответственно

$$H_i = \frac{1}{2\mu_i\gamma_i^2}\left[\left(P_{\rho_i} - \mu_i\gamma_i\dot{\gamma}_i\rho_i\right)^2 + \frac{P_{\varphi_i}^2}{\rho_i^2} + \frac{P_{\theta_i}^2}{\rho_i^2\cos^2\varphi_i}\right] - U_i^*, \tag{13}$$

$$U_i^* = f\frac{m_i\sigma_{i-1}}{\gamma_i\rho_i} + \frac{1}{2}\left(b_i\gamma_i^2 + \mu_i\dot{\gamma}_i^2\right)\rho_i^2 + B_i. \tag{14}$$

Вводя новые импульсы

$$P_{\rho_i} = \psi_i\tilde{P}_{\rho_i}, \quad P_{\varphi_i} = \psi_i\tilde{P}_{\varphi_i}, \quad P_{\theta_i} = \psi_i\tilde{P}_{\theta_i}, \tag{15}$$

$$\psi_i = \frac{\mu_i}{\mu_{i0}} = \frac{\mu_i(t)}{\mu_i(t_0)},$$

напишем уравнения движения в канонической форме

$$\dot{\rho}_i = \frac{\partial\tilde{H}_i}{\partial\tilde{P}_{\rho_i}} \qquad \dot{\varphi}_i = \frac{\partial\tilde{H}_i}{\partial\tilde{P}_{\varphi_i}} \qquad \dot{\theta}_i = \frac{\partial\tilde{H}_i}{\partial\tilde{P}_{\theta_i}},$$
$$\dot{\tilde{P}}_{\rho_i} = -\frac{\partial\tilde{H}_i}{\partial\rho_i} \qquad \dot{\tilde{P}}_{\varphi_i} = -\frac{\partial\tilde{H}_i}{\partial\varphi_i} \qquad \dot{\tilde{P}}_{\theta_i} = -\frac{\partial\tilde{H}_i}{\partial\theta_i}, \tag{16}$$

где

$$\tilde{H}_i = \tilde{H}_{i0} + \tilde{H}_{i1}, \tag{17}$$

В случае

$$\tilde{H}_i = \tilde{H}_{i0}, \tag{18}$$

$$\dot{\rho}_i = \frac{\partial\tilde{H}_{i0}}{\partial\tilde{P}_{\rho_i}} \qquad \dot{\varphi}_i = \frac{\partial\tilde{H}_{i0}}{\partial\tilde{P}_{\varphi_i}} \qquad \dot{\theta}_i = \frac{\partial\tilde{H}_{i0}}{\partial\tilde{P}_{\theta_i}},$$
$$\dot{\tilde{P}}_{\rho_i} = -\frac{\partial\tilde{H}_{i0}}{\partial\rho_i} \qquad \dot{\tilde{P}}_{\varphi_i} = -\frac{\partial\tilde{H}_{i0}}{\partial\varphi_i} \qquad \dot{\tilde{P}}_{\theta_i} = -\frac{\partial\tilde{H}_{i0}}{\partial\theta_i}, \tag{19}$$

Уравнение невозмущенного движение интегрируется методом Гамильтона-Якоби

$$\tilde{H}_{i0} = \frac{\psi_i}{2\mu_i\gamma_i^2}\left[\left(\tilde{P}_{\rho_i} - \frac{\mu_i\gamma_i\dot{\gamma}_i}{\psi_i}\rho_i\right)^2 + \frac{\tilde{P}_{\varphi_i}^2}{\rho_i^2} + \frac{\tilde{P}_{\theta_i}^2}{\rho_i^2\cos^2\varphi_i}\right] -$$
$$f\frac{m_i\sigma_{i-1}}{\psi_i\gamma_i\rho_i} - \frac{1}{\psi_i}\left(b_i\gamma_i^2 + \mu_i\dot{\gamma}_i^2\right)\rho_i^2, \tag{20}$$

$$\tilde{H}_{i1} = -\frac{1}{\psi_i}B_i. \tag{21}$$

При $B_i = 0$ уравнения (12) - (16) интегрируется методом Гамильтона-Якоби [15], причем постоянные интегрирования

$$\alpha_{1i}, \alpha_{2i}, \alpha_{3i}, \beta_{1i}, \beta_{2i}, \beta_{3i}, \tag{22}$$

Есть аналоги соответствующих элементов Якоби в классической задаче двух тел (точек) постоянной массы. Каждая система уравнений (12) - (16) определяет апериодическое движение по квазиконическому сечению

$$\rho_i = \frac{p_i}{1+e_i\cos\upsilon_i},\ \upsilon_i = u_i - \omega_i,\ \ p_i = a_i(1-e_i^2), \tag{23}$$

$$\dot\rho_i = \frac{1}{\mu_{i0}\gamma_i^2(t)}\cdot\frac{\beta_i}{\sqrt{p_i}}e_i\sin\upsilon_i, \qquad \dot u_i = \frac{1}{\mu_{i0}\gamma_i^2(t)}\cdot\frac{\beta_i\sqrt{p_i}}{\rho_i^2}, \tag{24}$$

$$\beta_i^2 = f\cdot\mu_i(t_0)m_i(t_0)\sigma_{i-1}(t_0), \tag{25}$$

$$tg\frac{\upsilon_i}{2} = \frac{\sqrt{1+e_i}}{\sqrt{1-e_i}}tg\frac{E_i}{2}, \quad e_i < 1, \tag{26}$$

$$E_i - e_i\sin E_i = M_i, \tag{27}$$

$$M_i = n_i\left[\phi_i(t) - \phi_i(\tau_i)\right],\quad n_i = \frac{\beta_i}{\mu_{i0}a_i^{3/2}}, \tag{28}$$

где $\phi_i(t)$ первообразная функции $\gamma_i^{-2}(t)$. В прямоугольных системах координат Якоби выражения координат и скорости промежуточных движений даются формулами

$$\begin{aligned} x_i &= \gamma_i\rho_i(\cos u_i\cos\Omega_i - \sin u_i\sin\Omega_i\cos i_i),\\ y_i &= \gamma_i\rho_i(\cos u_i\sin\Omega_i + \sin u_i\cos\Omega_i\cos i_i),\\ z_i &= \gamma_i\rho_i(\sin u_i\sin i_i),\quad r_i = \gamma_i\rho_i, \end{aligned} \tag{29}$$

$$\begin{aligned} \dot x_i &= \left(\frac{\dot\gamma_i}{\gamma_i}+\frac{\dot\rho_i}{\rho_i}\right)x_i + \gamma_i\rho_i\dot u_i(-\sin u_i\cos\Omega_i - \cos u_i\sin\Omega_i\cos i_i),\\ \dot y_i &= \left(\frac{\dot\gamma_i}{\gamma_i}+\frac{\dot\rho_i}{\rho_i}\right)y_i + \gamma_i\rho_i\dot u_i(-\sin u_i\sin\Omega_i + \cos u_i\cos\Omega_i\cos i_i),\\ \dot z_i &= \left(\frac{\dot\gamma_i}{\gamma_i}+\frac{\dot\rho_i}{\rho_i}\right)z_i + \gamma_i\rho_i\dot u_i(\cos u_i\sin i_i),\quad u_i = \upsilon_i + \omega_i. \end{aligned} \tag{30}$$

Величины

$$a_i, e_i, \omega_i, \Omega_i, i_i, \phi_i(\tau_i) \tag{31}$$

элементы орбиты – аналоги соответствующих кеплеровских элементов, причем

$$-2\alpha_{1i} = \frac{\beta_i^2}{\mu_{i0} a_i}, \quad a_{2i} = \beta_i \sqrt{p_i}, \quad a_{3i} = \beta_i \sqrt{p_i} \cos i_i,$$
$$\beta_{1i} = -\phi_i(\tau_i), \quad \beta_{2i} = \omega_i, \qquad \beta_{3i} = \Omega_i. \tag{32}$$

Уравнения (12) - (17) в системе переменных (18), как уравнения возмущенного движения имеют вид

$$\dot{\alpha}_{ki} = \frac{\partial \tilde{R}_i}{\partial \beta_{ki}}, \quad \dot{\beta}_{ki} = -\frac{\partial \tilde{R}_i}{\partial \alpha_{ki}}, \ k = 1,2,3, \tag{33}$$

где

$$\tilde{R}_i = \frac{1}{\psi_i} R_i(t, \alpha_{ki}, \beta_{ki}). \tag{34}$$

Введем аналоги элементов Делоне

$$L_i, \ G_i, \ H_i, \ l_i, \ g_i, \ h_i \tag{35}$$

посредством формул

$$-2\alpha_{1i} = \frac{\beta_i^4}{\mu_{i0} L_i^2}, \quad a_{2i} = G_i, \quad a_{3i} = H_i,$$
$$\beta_{1i} = \frac{l_i}{n_i} - \phi_i(t), \quad \beta_{2i} = g_i, \qquad \beta_{3i} = h_i. \tag{36}$$

Результат исследования

Уравнение возмущенного движения в аналогах элементов Делоне (31) имеют вид

$$\dot{L}_i = \frac{\partial R_i^*}{\partial l_i}, \quad \dot{G}_i = \frac{\partial R_i^*}{\partial g_i}, \quad \dot{H}_i = \frac{\partial R_i^*}{\partial h_i},$$
$$\dot{l}_i = -\frac{\partial R_i^*}{\partial L_i}, \quad \dot{g}_i = -\frac{\partial R_i^*}{\partial G_i}, \quad \dot{h}_i = -\frac{\partial R_i^*}{\partial H_i}, \tag{37}$$

где

$$R_i^* = \frac{1}{\gamma_i^2(t)} \cdot \frac{\beta_i^4}{2\mu_{i0} L_i^2} + \tilde{R}_i. \tag{38}$$

Дискуссия

Прежде всего необходимо подчеркнуть математическую и динамическую сложность моделирования многопланетных экзосистем, в частности, таких компактных семипланетных структур, как TRAPPIST-1. Применение в данной работе системы координат Якоби позволило существенно упростить аналитическую структуру задачи.

Однако специфической особенностью рассматриваемой системы TRAPPIST-1 является то, что массы составляющих её тел изменяются с течением времени (переменная масса), причем этот процесс носит не изотропный характер. Не изотропное рассеяние массы приводит к возникновению в системе дополнительных реактивных сил, что нарушает классические законы сохранения и выводит уравнения движения за рамки чисто ньютоновского гравитационного взаимодействия. Учет влияния этих реактивных сил при построении системы дифференциальных уравнений составляет основную научную новизну и главную трудность данного исследования.

Примененный в работе метод оскулирующих переменных позволяет получить не только чисто математическое решение уравнений движения, но и дает их глубокую

геометрическую и астрофизическую интерпретацию. Вводя оскулирующие элементы в качестве переменных, изменяющихся во времени, мы получаем возможность в динамике прослеживать эволюцию таких параметров, как эксцентриситет, большая полуось и пространственная ориентация орбит (наклонение, долгота восходящего узла). Это имеет важное значение для анализа долгосрочной орбитальной стабильности таких резонансных систем, как TRAPPIST-1.

Важнейшим теоретическим результатом работы следует считать построение канонической теории возмущений на основе полученных уравнений движения, а также вывод аналогов элементов Якоби и Делоне.

В конечном счете, возможность перехода к барицентрической системе координат позволяет рассматривать разработанную модель не просто как частное научное исследование системы TRAPPIST-1, а как универсальный математический аппарат, предназначенный для высокоточного расчета динамики любых многопланетных систем, находящихся на стадиях потери массы или аккреции.

Заключение

Исходя из уравнений движения задачи, полученных в координатах Якоби на базе соответствующих апериодических движений по квазиконическому сечению, строится каноническая теория возмущений. Получены уравнения возмущенного движения в аналогах элементов Якоби, Делоне. На основе полученных уравнений можно использовать теорию возмущений.

Список использованных источников

[1] Perminov A.S., Kuznetsov E.D. The Implementation of Hori–Deprit Method to the Construction Averaged Planetary Motion Theory by Means of Computer Algebra System Piranha // Mathematics in Computer Science. – 2020. – Vol. 14, № 2. – P. 305-316. https://doi.org/10.1007/s11786-019-00441-4

[2] Perminov A.S., Kuznetsov E.D. The orbital evolution of the Sun–Jupiter–Saturn–Uranus–Neptune system on long timescales// Astrophysics and Space Science. – 2020. – Vol. 365, № 144, - P.21. https://doi.org/10.1007/s10509-020-03855-w

[3] David Arnas. Analytic transformation from osculating to mean elements under J2 perturbation // Journal of Guidance, Control, and Dynamics. – 2023. - Vol.46, №11, P.2150-2158. https://doi.org/10.48550/arXiv.2212.08746

[4] Maxwell X. C., Simon P.Z., Damian P. Neural Symplectic Integrator with Hamiltonian Inductive Bias for the Gravitational N-body Problem // Instrumentation and Methods for Astrophysics.-2021.- P.7. https://doi.org/10.48550/arXiv.2111.15631

[5] Холшевников К.В., Греб А.В., Кузнецов Э.Д. Разложение гамильтониана в ряд Пуассона по всем элементам // Астрономический вестник. – 2001. – Т.35, №3. – С. 243-248. https://doi.org/10.1023/A:1010487107989

[6] Кузнецов Э.Д., Холшевников К.В. Разложение гамильтониана двупланетной задачи в ряд Пуассона по всем элементам: применение пуассоновского процессора // Астрономический вестник. – 2004. – Т. 38, № 2. – С. 147–154. https://doi.org/10.1023/B:SOLS.0000022825.93837.7d

[7] Кузнецов Э.Д., Холшевников К.В. Динамическая эволюция слабовозмущенной двупланетной системы на космогоническом интервале времени: система Солнце-Юпитер-Сатурн // Астрономический вестник. – 2006. – Т. 40, № 3. – С. 239-250. DOI: 10.1134/S0038094606030087

[8] Кузнецов Э.Д., Холшевников К.В. Орбитальная эволюция двупланетной системы Солнце – Юпитер – Сатурн // Вестник СПбГУ. – 2009. – Т. 1. – С. 139-150. https://cyberleninka.ru/article/n/orbitalnaya-evolyutsiya-dvuplanetnoy-sistemy-solntse-yupiter-saturn

[9] Минглибаев М.Дж. – Динамика гравитирующих тел с переменными массами и размерами. Поступательное и поступательно-вращательное движение / LAP LAMBERT Academic Publishing. - 2012. P.224.

[10] Prokopenya, A.N., Minglibayev, M. Zh. & Mayemerova, G.M. Symbolic calculations in studying the problem of three bodies with variable masses. Program Computer Soft. – 2014. – Vol. 40, №2. – P.79-85. https://doi.org/10.1134/S036176881402008X

[11] Minglibayev, M.Zh., Mayemerova, G.M., Imanova, Zh.U. Secular perturbations in the three-body problem with masses changing anisotropically at the different rates// Bulletin of the Karaganda University. Mathematics Series. - 2016.- Vol.84, №4. P.94–98. https://doi.org/10.31489/2016m4/94-98

[12] Oni Leke., Jagadish Singh. Out-of-plane equilibrium points of extra-solar planets in the central binaries PSR B1620-26 and Kepler-16 with clusters of material points and variable masses//New Astronomy.- 2023.- Vol.99. P.101958. https://doi.org/10.1016/j.newast.2022.101958

[13] Oni Leke., Tajudeen Oluwafemi Amuda. Instability of In-plane Equilibrium Points of the Restricted Three-Body Problem with P-R Drag and Variable Masses//Proceedings of the International Astronomical Union. - 2024.- Vol.20.-P. 105 – 113. DOI: https://doi.org/10.1017/S174392132500047X

[14] Минглибаев М.Дж., Сапарова М.Р., Уравнения движения экзопланетной системы TRAPPIST-1 с переменными массами в системе координат Якоби // Вестник Национальной инженерной академии Республики Казахстан.- 2026.-Т.99, №1. С.219-228. DOI: https://doi.org/10.47533/2026.1606-146X.1-18

[15] Минглибаев М.Дж. К канонической теории возмущений в небесной механике тел переменной массы.// Труды АФИ АН КазССР. – Алма-Ата: Ғылым, 1992. – Т.50. – С.71-78. https://elibrary.kaznu.kz/wp-content/uploads/2025/11/minglibaev-m.d.-teoriya-vozmushheniya-na-baze-aperiodicheskogo-dvizheniya-po-kvazikonicheskomu-secheniyu.pdf

References

[1] Perminov A. S., Kuznetsov E. D. The Implementation of Hori–Deprit Method to the Construction Averaged Planetary Motion Theory by Means of Computer Algebra System Piranha // Mathematics in Computer Science. – 2020. – Vol. 14, № 2. – P. 305-316. https://doi.org/10.1007/s11786-019-00441-4

[2] Perminov A. S., Kuznetsov E. D. The orbital evolution of the Sun–Jupiter–Saturn–Uranus–Neptune system on long timescales // Astrophysics and Space Science. – 2020. – Vol. 365, № 144. – P. 21. https://doi.org/10.1007/s10509-020-03855-w

[3] Arnas David. Analytic transformation from osculating to mean elements under J2 perturbation // Journal of Guidance, Control, and Dynamics. – 2023. – Vol. 46, № 11. – P. 2150-2158. https://doi.org/10.48550/arXiv.2212.08746

[4] Maxwell X. C., Simon P. Z., Damian P. Neural Symplectic Integrator with Hamiltonian Inductive Bias for the Gravitational N-body Problem // Instrumentation and Methods for Astrophysics. – 2021. – P.7. https://doi.org/10.48550/arXiv.2111.15631

[5] Kholshevnikov K. V., Greb A. V., Kuznetsov E. D. Razlozhenie gamil'toniana v ryad Puassona po vsem elementam [Expansion of the Hamiltonian in a Poisson series in terms of all elements] // Astronomicheskii vestnik. – 2001. – Vol. 35, № 3. – P. 243-248. https://doi.org/10.1023/A:1010487107989 – (In Russian).

[6] Kuznetsov E. D., Kholshevnikov K. V. Razlozhenie gamil'toniana dvuplanetnoi zadachi v ryad Puassona po vsem elementam: primenenie puassonovskogo processora [Expansion of the Hamiltonian of a two-planet problem into a Poisson series in all elements: application of a Poisson processor] // Astronomicheskii vestnik. – 2004. – Vol. 38, № 2. – P. 147–154. https://doi.org/10.1023/B:SOLS.0000022825.93837.7d – (In Russian).

[7] Kuznetsov E. D., Kholshevnikov K. V. Dinamicheskaya evolyuciya slabovozmushchennoi dvuplanetnoi sistemy na kosmogonicheskom intervale vremeni: sistema Solnce-Yupiter-Saturn [Dynamical evolution of a weakly perturbed two-planet system on a cosmogonic time interval: the Sun-Jupiter-Saturn system] // Astronomicheskii vestnik. – 2006. – Vol. 40, № 3. – P. 239-250. – DOI: 10.1134/S0038094606030087 – (In Russian).

[8] Kuznetsov E. D., Kholshevnikov K. V. Orbital'naya evolyuciya dvuplanetnoi sistemy Solnce – Yupiter – Saturn [Orbital evolution of the two-planet system Sun–Jupiter–Saturn] // Vestnik SPbGU. – 2009. – Vol. 1. – P. 139-150. https://cyberleninka.ru/article/n/orbitalnaya-evolyutsiya-dvuplanetnoy-sistemy-solntse-yupiter-saturn. – (In Russian).

[9] Minglibayev M.Zh. Dinamika gravitiruyushchih tel s peremennymi massami i razmerami. Postupatel'noe i postupatel'no-vraschatel'noe dvizhenie [Dynamics of gravitating bodies with variable

masses and sizes. Translational and translational-rotational motion]. – LAP LAMBERT Academic Publishing, 2012. – P.224. – (In Russian).

[10] Prokopenya A. N., Minglibayev M. Zh., Mayemerova G. M. Symbolic calculations in studying the problem of three bodies with variable masses // Programming and Computer Software. – 2014. – Vol. 40, № 2. – P. 79-85. DOI: 10.1134/S036176881402008X

[11] Minglibayev M. Zh., Mayemerova G. M., Imanova Zh. U. Secular perturbations in the three-body problem with masses changing anisotropically at the different rates // Bulletin of the Karaganda University. Mathematics Series. – 2016. – Vol. 84, № 4. – P. 94–98. DOI: 10.31489/2016m4/94-98.

[12] Oni Leke, Jagadish Singh. Out-of-plane equilibrium points of extra-solar planets in the central binaries PSR B1620-26 and Kepler-16 with clusters of material points and variable masses // New Astronomy. – 2023. – Vol. 99. – P. 101958. DOI: 10.1016/j.newast.2022.101958

[13] Oni Leke, Tajudeen Oluwafemi Amuda. Instability of In-plane Equilibrium Points of the Restricted Three-Body Problem with P-R Drag and Variable Masses // Proceedings of the International Astronomical Union. – 2024. – Vol. 20. – P. 105 – 113. DOI: 10.1017/S174392132500047X.

[14] Minglibayev M. Zh., Saparova M. R. Uravneniya dvizheniya ekzoplanetnoi sistemy TRAPPIST-1 s peremennymi massami v sisteme koordinat Yakobi [Equations of motion of the TRAPPIST-1 exoplanetary system with variable masses in the Jacobi coordinate system] // Vestnik Nacional'noi inzhenernoi akademii Respubliki Kazahstan. – 2026. – Vol. 99, № 1. – P. 219-228. DOI: https://doi.org/10.47533/2026.1606-146X.1-18 – (In Russian).

[15] Minglibayev M. Zh. K kanonicheskoi teorii vozmushchenii v nebesnoi mehanike tel peremennoi massy [On the canonical perturbation theory in celestial mechanics of bodies with variable mass] // Trudy AFI AN KazSSR. – Alma-Ata: Gylym, 1992. – Vol. 50. – P. 71-78. https://elibrary.kaznu.kz/wp-content/uploads/2025/11/minglibaev-m.d.-teoriya-vozmushheniya-na-baze-aperiodicheskogo-dvizheniya-po-kvazikonicheskomu-secheniyu.pdf – (In Russian).

ҒТАХР 27.31.44

https://doi.org/10.51889/2959-5894.2026.95.3.008

Рысқан А.Р. [1,2], Абдрахман Б.Қ.[*3], Айнакеева Н. Ж[1,4]

[1]Нархоз Университеті, Алматы қ., Қазақстан
[2] Калифорния университеті, Дэвис, АҚШ
[3]Абай атындағы Қазақ ұлттық педагогикалық университеті, Алматы қ., Қазақстан
[4]У.А.Джолдасбеков атындағы механика және машинатану институты, Алматы қ., Қазақстан
[*]*e-mail: balnurabdrakhman584@gmail.com*

ГИПЕРГЕОМЕТРИЯЛЫҚ ФУНКЦИЯЛАРЫ ҮШІН ЕКІНШІ РЕТТІ ДЕРБЕС ТУЫНДЫЛЫ ДИФФЕРЕНЦИАЛДЫҚ ТЕҢДЕУЛЕР ЖҮЙЕСІН ШЕШУ

Аңдатпа

Бұл мақалада төрт айнымалыдан тәуелді гипергеометриялық функцияларына қатысты екінші ретті дербес туындылы дифференциалдық теңдеулер жүйесі қарастырылады. Мұндай функциялар қазіргі математиканың әртүрлі маңызды салаларында, мысалы, математикалық физика, ішектер теориясы және механикада кеңінен қолданылады. Атап айтқанда, көп айнымалы гипергеометриялық функциялардың дифференциалдық теңдеулер жүйесі арқылы сипатталуы олардың құрылымдық қасиеттерін, шекаралық шарттардағы шешімдерінің бар-жоғын түсінуге мүмкіндік береді. Бұл жұмыста Экстон енгізген көп айнымалы гипергеометриялық функциялар класына жататын функцияларының шешімдері сипатталады. Зерттеу барысында аталған функциялар класы үшін дифференциалдық операторлар құрылып, олардың фундаментальды шешімдерінің аналитикалық сипаттамалары алынды. Алынған нәтижелер көп айнымалы функциялардың қолданбалы есептердегі рөлін тереңірек түсінуге мүмкіндік береді. Бұл мақаланың қорытынды нәтижелері көп айнымалы функциялардың заманауи қолданбалы есептердегі маңызды рөлін айқындап, математикалық модельдеу саласындағы жаңа ізденістерге жол ашады. Есептеулерде алынған шешімдердің нақтылығы мен негізділігі классикалық дифференциалдық теңдеулер теориясының әдістерімен дәлелденген.

Түйін сөздер: гипергеометриялық қатарлар, төрт айнымалы гипергеометриялық функциясы, Экстон гипергеометриялық функциялары, жүйенің сызықтық тәуелсіз шешімдері.

Рыскан А.Р.[1,2], Абдрахман Б.К.[*3], Айнакеева Н. Ж[1,4]
[1]Университет Нархоз, г.Алматы, Казахстан
[2]Калифорнийский университет, Дэвис, США
[3]Казахский национальный педагогический университет имени Абая, г.Алматы, Казахстан
[4]Институт механики и машиноведения им.академика У.А.Джолдасбекова, г.Алматы, Казахстан

РЕШЕНИЕ СИСТЕМЫ ДИФФЕРЕНЦИАЛЬНЫХ УРАВНЕНИЙ ВТОРОГО ПОРЯДКА В ЧАСТНЫХ ПРОИЗВОДНЫХ ДЛЯ ГИПЕРГЕОМЕТРИЧЕСКИХ ФУНКЦИЙ

Аннотация

В данной работе исследуется система дифференциальных уравнений в частных производных второго порядка, связанных с гипергеометрическими функциями четырех переменных $F_{24}^{(4)}$ $F_{25}^{(4)}$. Такие функции находят широкое применение в различных важных областях современной математики, таких как математическая физика, теория струн и механика. Особое внимание уделяется описанию многомерных гипергеометрических функций с помощью систем дифференциальных уравнений, что позволяет исследовать их внутреннюю структуру и анализировать существование решений при различных граничных условиях. В данной работе изучаются функции $F_{24}^{(4)}$ $F_{25}^{(4)}$, принадлежащие классу многомерных гипергеометрических функций, введенному Экстоном. В ходе исследования были построены соответствующие дифференциальные операторы для этого класса функций и получены аналитические формулы для их фундаментальных решений. Представленные результаты демонстрируют важность многомерных гипергеометрических функций для решения современных прикладных задач и указывают на перспективные направления дальнейших исследований в области

математического моделирования. Надежность и точность полученных вычислительных результатов подтверждаются методами классической теории дифференциальных уравнений.

Ключевые слова: гипергеометрический ряд, гипергеометрическая функция четырех переменных, гипергеометрические функции Экстона, линейно независимые решения системы.

Ryskan A.R.[1,2], Abdrakhman B.K.[*3], Ainakeyeva N.Zh[1,4]
[1]Narxoz University, Almaty, Kazakhstan
[2]UC Davis, Davis, USA
[3]Abai Kazakh National Pedagogical University, Almaty, Kazakhstan
[4]Joldasbekov Institute of Mechanics and Engineering, Almaty, Kazakhstan

SOLVING A SYSTEM OF SECOND-ORDER PARTIAL DIFFERENTIAL EQUATIONS FOR HYPERGEOMETRIC FUNCTIONS

Abstract

This research focuses on the analysis of a second-order system of partial differential equations linked with the four-variable hypergeometric functions $F_{24}^{(4)}$ $F_{25}^{(4)}$. These specific functions play a critical role in numerous branches of contemporary mathematics and theoretical physics, notably in quantum mechanics and string theory. By treating multidimensional hypergeometric functions within the framework of differential systems, this work explores their structural properties and evaluates solution behavior under specific boundary constraints. The subject of this inquiry involves the function classes $F_{24}^{(4)}$ and $F_{25}^{(4)}$ originally proposed by Exton. Throughout this investigation, distinct differential operators were established for the aforementioned functions, enabling the derivation of exact analytical expressions for their fundamental solutions. The outcomes highlighted in this paper underscore the significance of multivariate hypergeometric functions in addressing contemporary applied problems and open up new avenues for mathematical modeling. Ultimately, the validity of the computed data is rigorously verified through classical techniques of differential equation theory.

Keywords: hypergeometric series, four-variable hypergeometric function, Exton hypergeometric functions, linearly independent solutions of the system.

Кіріспе

Гипергеометриялық функциялар математикалық талдаудың маңызды салаларының бірі болып табылатын арнайы функциялар теориясының негізін құрайды [1, 2]. Бұл функциялар алғаш рет XVIII ғасырда Леонард Эйлер мен Карл Фридрих Гаусстың еңбектерінде кездескен [1,2]. 1812 жылы Гаусс бір айнымалының $F_1\left(a,b,c,z\right)$ классикалық гипергеометриялық функциясын қарастырып, оның қасиеттерін, сондай-ақ дифференциалдық теңдеулермен байланысын зерттеді [1, 2]. Кейіннен бұл теория XX ғасырда Куммер мен Риманның еңбектерінде тереңірек зерттелді [2]. Гипергеометриялық функцияларды көп айнымалыға кеңейту XX ғасырдың басында басталып, кейіннен жаңа бағыттардың дамуымен өрістей түсті [3]. Бұл салада Аппель (екі айнымалы функциялар), Лауричелла ($F_A^{(n)}, F_B^{(n)}, F_C^{(n)}, F_D^{(n)}$, көп айнымалы функциялар) Және Хорн [4] айтарлықтай үлес қосты. XX ғасырдың екінші жартысында Шривастава мен Экстонның еңбектерінде олардың қасиеттері мен қолдану аясын едәуір кеңейте отырып, көп айнымалы гипергеометриялық функциялардың жаңа кластары енгізілді [5, 6]. Атап айтқанда, Экстон екі, үш және төрт айнымалыға байланысты гипергеометриялық функциялардың толық жүйелерін ұсынды [7]. Ол көп айнымалы функцияларды $K_1, K_2, K_3, ..., K_{21}$ белгіленген функциялар арқылы жіктеді және әр класс үшін сәйкес дифференциалдық теңдеулер жүйесін құрды [8], [9]. Алайда, көп айнымалы гипергеометриялық функциялардың жаңа түрлеріне сәйкес келетін дифференциалдық теңдеулер жүйесінің жалпы шешімдерін табу және математикалық физиканың шеттік есептеріне қолдану мәселелері әлі де толық шешімін таппаған. Қазіргі математикалық талдау мен математикалық физика есептерін шешуде көп айнымалы гипергеометриялық

функцияларды қолдану қажеттілігі туындап отыр.Бұл зерттеудің өзектілігі көп айнымалы гипергеометриялық функциялардың жалпы теориясын одан әрі дамыту, олардың жүйелік қасиеттерін анықтау және қолдану аясын кеңейту қажеттілгімен айқындалады.

Аппель мен Лауричелланың классикалық еңбектерінде [3], Шриваставаның жалпыланған модельдерінде [6] және соңғы уақыттағы шеттік есептерге арналған зерттеулерде қарастырылған дәстүрлі тәсілдермен салыстырғанда, бұл зерттеуде қарастырылатын теңдеулер жүйесінің құрылымы кеңейтілген, олардың шешімдері тек белгеле бір нүктеде емес, жалпыланған сингулярлық жүйе аясында толық сипатталған. Бұл зерттеуде $F_{24}^{(4)}$ $F_{25}^{(4)}$ функцияларына сәйкес дифференциалдық теңдеулер жүйесі құрылады, олардың шешімдерінің қасиеттері зерттеледі [10]. Бұл жұмыстың мақсаты-гипергеометриялық функциялар теориясы аясында көп айнымалы функциялармен байланысты теңдеулер жүйесін шешу әдістерін ұсыну және олардың қолданылу маңызын көрсету. Алынған нәтижелер болашақта математикалық модельдерді және көп айнымалы гипергеометриялық функцияларды қамтитын шеттік есептерді шешу үшін пайдалы болуы мүмкін [11–15]. Соңғы жылдары көп айнымалы гипергеометриялық функциялардың шешімдер кеңістігін құру, аналитикалық жалғастыру формулаларын алу және дифференциалдық қатынастарын зерттеу бағытында бірқатар жұмыстар жарияланды [16–18]. Бұл зерттеулер Лауричелла типті функциялардың аналитикалық қасиеттерін тереңірек түсінуге және олардың дифференциалдық теңдеулер жүйелерін шешу әдістерін дамытуға мүмкіндік береді.

Зерттеудің ғылыми жаңалығы. Бұл зерттеуде төрт айнымалы гипергеометриялық $F_{24}^{(4)}$ және $F_{25}^{(4)}$ функцияларына сәйкес екінші ретті дербес туындылы дифференциалдық теңдеулер жүйесі қарастырылып, оның төрт сызықтық тәуелсіз шешімі **алғаш рет** аналитикалық түрде алынды. Алынған шешімдер Лауричелла типті гипергеометриялық функциялар арқылы өрнектеліп, олардың аналитикалық қасиеттері зерттелді.

Зерттеу әдістемесі

Бұл зерттеуде төрт айнымалыға байланысты гипергеометриялық $F_{24}^{(4)}$ $F_{25}^{(4)}$ функция үшін екінші ретті дербес туындылы дифференциалдық теңдеулер жүйесін шешудің аналитикалық әдістері қолданылады. Бұл функциялар арнайы дифференциалдық жүйелердің шешімдері бола отырып, физика мен қолданбалы математикадағы көптеген есептерді шешудің тиімді құралы бола алады. Жүргізілген талдау олардың симметрияларын, аналитикалық қасиеттерін, құрылымдық ерекшеліктерін одан әрі зерттеуге негіз болады. Экстон енгізген Лауричелла типіндегі гипергеометриялық функциялар үшін сәйкес дифференциалдық теңдеулер жүйесі құрылады, содан кейін алгебралық түрлендірулерді қолдана отырып стандартты екінші ретті түрге айналады. Шешімдер айнымалыларды дәрежелі алмастыру әдісі арқылы ізделеді. Лауричелла функциясының жіктеу формулаларын қолдану арқылы бұл шешімдер бір айнымалы Гаусс гипергеометриялық функцияларының көбейтіндісі түрінде өрнектеледі. Нәтижесінде теңдеулер жүйесінің сызықтық тәуелсіз шешімдері алынады. Ұсынылып отырған әдіс көптеген айнымалылардың арнайы функциялары теориясының ішкі құрылымын тереңірек түсінуге және жаңа аналитикалық формаларды іздеуге перспективалы жолдар ашады.

Зерттеу нәтижелері

$F_{24}^{(4)}$ $F_{25}^{(4)}$ функцияларының дербес туындылы дифференциалдық теңдеулер жүйесінің шешімі Гипергеометриялық $F_{24}^{(4)}$ функциясын қарастырайық:

$$F_{24}^{(4)}\left(a_1,a_2,b_1,b_2,b_3;c_1,c_2,c_3;x,y,z,t\right)=$$

$$=\sum_{m,n,p,q=0}^{\infty}\frac{\left(a_1\right)_{m+n+p}\left(a_2\right)_q\left(b_1\right)_{m+n}\left(b_2\right)_p\left(b_3\right)_q}{\left(c_1\right)_m\left(c_2\right)_{n+q}\left(c_3\right)_p}\frac{x^m}{m!}\frac{y^n}{n!}\frac{z^p}{p!}\frac{t^q}{q!},\qquad(1)$$

Мұндағы: $c\neq 0,-1,-2,-3,.....;$ және $i=1,.....,3.$

Алдымен, Кампэ-де-Ферье қатарын қолданамыз, [4],

$$A_{m,n,p,q}=\frac{\left(a_1\right)_{m+n+p}\left(a_2\right)_q\left(b_1\right)_{m+n}\left(b_2\right)_p\left(b_3\right)_q}{\left(c_1\right)_m\left(c_2\right)_{n+q}\left(c_3\right)_p m!n!p!q!}$$

$$A_{m+1,n,p,q}=\frac{\left(a_1\right)_{m+n+p+1}\left(a_2\right)_q\left(b_1\right)_{m+n+1}\left(b_2\right)_p\left(b_3\right)_q}{\left(c_1\right)_{m+1}\left(c_2\right)_{n+q}\left(c_3\right)_p (m+1)!n!p!q!}$$

$$A_{m,n+1,p,q}=\frac{\left(a_1\right)_{m+n+p+1}\left(a_2\right)_q\left(b_1\right)_{m+n+1}\left(b_2\right)_p\left(b_3\right)_q}{\left(c_1\right)_m\left(c_2\right)_{n+q+1}\left(c_3\right)_p m!(n+1)!p!q!}\qquad(2)$$

$$A_{m,n,p+1,q}=\frac{\left(a_1\right)_{m+n+p+1}\left(a_2\right)_q\left(b_1\right)_{m+n}\left(b_2\right)_{p+1}\left(b_3\right)_q}{\left(c_1\right)_m\left(c_2\right)_{n+q}\left(c_3\right)_{p+1} m!n!(p+1)!q!}$$

$$A_{m,n,p,q+1}=\frac{\left(a_1\right)_{m+n+p}\left(a_2\right)_{q+1}\left(b_1\right)_{m+n}\left(b_2\right)_p\left(b_3\right)_{q+1}}{\left(c_1\right)_m\left(c_2\right)_{n+q+1}\left(c_3\right)_p m!n!p!(q+1)!}$$

Мұндағы $A_{m,n,p,q}$ Кампэ-де-Ферье қатарының коэффициенттері, ал, $a_1,a_2,b_1,b_2,b_3;c_1,c_2,c_3;x,y,z,t$ параметрлер[12].

$$\frac{A_{m+1,n,p,q}}{A_{m,n,p,q}}=\frac{\left(a_1\right)_{m+n+p+1}\left(a_2\right)_q\left(b_1\right)_{m+n+1}\left(b_2\right)_p\left(b_3\right)_q}{\left(c_1\right)_{m+1}\left(c_2\right)_{n+q}\left(c_3\right)_p (m+1)!n!p!q!}\frac{\left(c_1\right)_m\left(c_2\right)_{n+q}\left(c_3\right)_p m!n!p!q!}{\left(a_1\right)_{m+n+p}\left(a_2\right)_q\left(b_1\right)_{m+n}\left(b_2\right)_p\left(b_3\right)_q}$$

$$\frac{A_{m,n+1,p,q}}{A_{m,n,p,q}}=\frac{\left(a_1\right)_{m+n+p+1}\left(a_2\right)_q\left(b_1\right)_{m+n+1}\left(b_2\right)_p\left(b_3\right)_q}{\left(c_1\right)_m\left(c_2\right)_{n+q+1}\left(c_3\right)_p m!(n+1)!p!q!}\frac{\left(c_1\right)_m\left(c_2\right)_{n+q}\left(c_3\right)_p m!n!p!q!}{\left(a_1\right)_{m+n+p}\left(a_2\right)_q\left(b_1\right)_{m+n}\left(b_2\right)_p\left(b_3\right)_q}\qquad(3)$$

$$\frac{A_{m,n,p+1,q}}{A_{m,n,p,q}}=\frac{\left(a_1\right)_{m+n+p+1}\left(a_2\right)_q\left(b_1\right)_{m+n}\left(b_2\right)_{p+1}\left(b_3\right)_q}{\left(c_1\right)_m\left(c_2\right)_{n+q}\left(c_3\right)_{p+1} m!n!(p+1)!q!}\frac{\left(c_1\right)_m\left(c_2\right)_{n+q}\left(c_3\right)_p m!n!p!q!}{\left(a_1\right)_{m+n+p}\left(a_2\right)_q\left(b_1\right)_{m+n}\left(b_2\right)_p\left(b_3\right)_q}$$

$$\frac{A_{m,n,p,q+1}}{A_{m,n,p,q}}=\frac{\left(a_1\right)_{m+n+p}\left(a_2\right)_{q+1}\left(b_1\right)_{m+n}\left(b_2\right)_p\left(b_3\right)_{q+1}}{\left(c_1\right)_m\left(c_2\right)_{n+q+1}\left(c_3\right)_p m!n!p!(q+1)!}\frac{\left(c_1\right)_m\left(c_2\right)_{n+q}\left(c_3\right)_p m!n!p!q!}{\left(a_1\right)_{m+n+p}\left(a_2\right)_q\left(b_1\right)_{m+n}\left(b_2\right)_p\left(b_3\right)_q}$$

(2) қатардан m-ге байланысты коэффициенттердің бір бағыттағы рекурренттік қатынасын (3) аламыз, m,n,p,q:

$$\frac{A_{m+1,n,p,q}}{A_{m,n,p,q}}=\frac{\left(a_1+m+n+p\right)\left(b_1+m+n\right)}{\left(c_1+m\right)\left(m+1\right)}$$

$$\frac{A_{m,n+1,p,q}}{A_{m,n,p,q}}=\frac{\left(a_1+m+n+p\right)\left(b_1+m+n\right)}{\left(c_2+n+q\right)\left(n+1\right)}$$

$$\frac{A_{m,n,p+1,q}}{A_{m,n,p,q}}=\frac{\left(a_1+m+n+p\right)\left(b_2+p\right)}{\left(c_3+p\right)\left(p+1\right)}\qquad(4)$$

$$\frac{A_{m,n,p,q+1}}{A_{m,n,p,q}}=\frac{\left(a_2+q\right)\left(b_3+q\right)}{\left(c_2+n+q\right)\left(q+1\right)}$$

Мұндағы $m = x\dfrac{\partial}{\partial x}$, $n = y\dfrac{\partial}{\partial y}$, $p = z\dfrac{\partial}{\partial z}$, $q = t\dfrac{\partial}{\partial t}$ дифференциалдық операторлар ретінде анықталады.

Көп айнымалы гипергеометриялық функциялар теориясына сәйкес, гипергеометриялық функциясы $F_{24}^{(4)}$ үшін дербес туындылы теңдеулер жүйесі келесідей беріледі:

$$\begin{aligned}
&\left(c_1 + x\frac{\partial}{\partial x}\right)\left(x\frac{\partial}{\partial x}+1\right)x^{-1}u - \left(a_1 + x\frac{\partial}{\partial x} + y\frac{\partial}{\partial y} + z\frac{\partial}{\partial z}\right)\left(b_1 + x\frac{\partial}{\partial x} + y\frac{\partial}{\partial y}\right)u = 0\\
&\left(c_2 + y\frac{\partial}{\partial y} + t\frac{\partial}{\partial t}\right)\left(y\frac{\partial}{\partial y}+1\right)y^{-1}u - \left(a_1 + x\frac{\partial}{\partial x} + y\frac{\partial}{\partial y} + z\frac{\partial}{\partial z}\right)\left(b_1 + x\frac{\partial}{\partial x} + y\frac{\partial}{\partial y}\right)u = 0\\
&\left(c_3 + z\frac{\partial}{\partial z}\right)\left(z\frac{\partial}{\partial z}+1\right)z^{-1}u - \left(a_1 + x\frac{\partial}{\partial x} + y\frac{\partial}{\partial y} + z\frac{\partial}{\partial z}\right)\left(b_2 + z\frac{\partial}{\partial z}\right)u = 0\\
&\left(c_2 + y\frac{\partial}{\partial y} + t\frac{\partial}{\partial t}\right)\left(t\frac{\partial}{\partial t}+1\right)t^{-1}u - \left(a_2 + t\frac{\partial}{\partial t}\right)\left(b_3 + t\frac{\partial}{\partial t}\right)u = 0
\end{aligned} \tag{5}$$

(5)- теңдеулер жүйесіне қарапайым есептеулерді қолдана отырып, біз келесі екінші ретті дербес туындылы теңдеулер жүйесін аламыз:

$$\begin{cases}
x(1-x)u_{xx} - 2xyu_{xy} - zxu_{xz} - yyu_{yy} - yzu_{yz} + [c_1 - (a_1 + b_1 + 1)x]u_x - (a_1 + b_1 + 1)yu_y - b_1 zu_z - a_1 b_1 u = 0\\
y(1-y)u_{yy} + tu_{yt} - yzu_{yz} - xxu_{xx} - 2xyu_{xy} - zxu_{xz} + [c_2 - (a_1 + b_1 + 1)y]u_y - (a_1 + b_1 + 1)xu_x - b_1 zu_z - a_1 b_1 u = 0\\
z(1-z)u_{zz} - zxu_{xz} - yzu_{yz} + [c_3 - (a_1 + b_2 + 1)z]u_z - (xu_x + yu_y + a_1 u)b_2 = 0\\
t(1-t)u_{tt} + yu_{ty} + [c_2 - (a_2 + b_3 + 1)t]u_t - a_2 b_3 u = 0
\end{cases} \tag{6}$$

Мұндағы: $u(x,y,z,t) = F_{24}^{(4)}(a_1, a_2, b_1, b_2, b_3; c_1, c_2, c_3; x, y, z, t)$

Біз сызықтық тәуелсіз шешімдерді келесі түрде іздейміз:

$$u = x^{\alpha} y^{\beta} z^{\gamma} t^{\delta} w \tag{7}$$

Мұндағы, w белгісіз функция, $\alpha, \beta, \gamma, \delta$ анықталатын тұрақтылар. Жаңа айнымалыларды $x^{\alpha}, y^{\beta}, z^{\gamma}, t^{\delta}$ (5) жүйедегі u функциясының орнына қою арқылы дифференциалдық теңдеулердің келесі жүйесін аламыз:

$$\begin{cases}
x(1-x)w_{xx} - yyw_{yy} - zxw_{xz} - 2xyw_{xy} - yzw_{yz} + [C_1 - (A_1 + B_1 + 1)x]w_x -\\
(A_1 + B_1 + 1)yw_y - B_1 zw_z + [\alpha(c_1 + \alpha - 1)x^{-1} - A_1 \cdot B_1]w = 0\\
y(1-y)w_{yy} - x^2 w_{xx} - 2xyw_{xy} - xzw_{xz} - yzw_{yz} + tw_{yt} + \left[C_2 - (A_1 + B_1 + 1)y\right]w_y\\
-(A_1 + B_1 + 1)xw_x - B_1 zw_z + t\beta y^{-1} w_t + \left[\beta(c_2 + \beta + \delta - 1)y^{-1} - A_1 \cdot B_1\right]w = 0\\
z(1-z)w_{zz} - xzw_{xz} - yzw_{yz} + [C_3 - (A_1 + B_2 + 1)z]w_z - B_2(xw_x + yw_y)\\
+[(c_3 + \gamma - 1)\gamma z^{-1} - A_1 B_2]w = 0\\
(1-t)tw_{tt} + xw_{xt} + \delta t^{-1} xw_x + [C_4 - (A_2 + B_3 + 1)t]w_t + [\delta(c_2 + \beta + \delta - 1)t^{-1} - A_2 B_3]w = 0
\end{cases} \tag{8}$$

Мұндағы

$$\begin{aligned}
&A_1 = \alpha+\beta+\gamma+a_1\\
&A_2 = \delta+a_2\\
&B_1 = b_1+\alpha+\beta,\ B_2 = b_2+\gamma,\ B_3 = b_3+\delta,\\
&C_1 = c_1+2\alpha\\
&C_2 = c_2+2\beta+\delta\\
&C_4 = c_2+\beta+2\delta\\
&C_3 = c_3+2\gamma
\end{aligned} \tag{9}$$

Жоғарыдағы (6) жүйесі (8) жүйесіне ұқсас. Сондықтан, келесі шарттар орындалуы керек:

$$\begin{cases}
\beta = 0\\
\delta = 0\\
\alpha\left(c_1+\alpha-1\right)=0\\
\beta\left(c_2+\beta+\delta-1\right)=0\\
\gamma(c_3+\gamma-1)=0\\
\delta(c_2+\beta+\delta-1)=0
\end{cases} \tag{10}$$

(9) жүйесінің шешімдерін төрт комбинация түрінде жазуға болады:

	1	2	3	4
α:	0	$1-c_1$	0	$1-c_1$
β:	0	0	0	0
γ:	0	0	$1-c_3$	$1-c_3$
δ:	0	0	0	0

(11)

Осылайша, жүйенің (11) шешімдерін (7) – ке енгізіп, келесі гипергеометриялық функциялар түрінде төрт айнымалының (8) екінші ретті дифференциалдық теңдеулер жүйесінің сызықтық тәуелсіз шешімдерін аламыз:

$$u\left(x,y,z,t\right) = x^{\alpha}y^{\beta}z^{\gamma}t^{\delta}F_{24}^{(4)}\left(\alpha+\gamma+a_1,a_2,b_1+\alpha,b_2+\gamma,b_3;c_1+2\alpha,c_2,c_3+2\gamma;x,y,z,t\right) =$$

$$= \sum_{m,n,p,q=0}^{\infty}\frac{\left(\alpha+\gamma+a_1\right)_{m+n+p}\left(a_2\right)_q\left(b_1+\alpha\right)_{m+n}\left(b_2+\gamma\right)_p\left(b_3\right)_q}{\left(c_1+2\alpha\right)_m\left(c_2\right)_{n+q}\left(c_3+2\gamma\right)_p}\frac{x^m}{m!}\frac{y^n}{n!}\frac{z^p}{p!}\frac{t^q}{q!},$$

$$u_1\left(x,y,z,t\right)F_{24}^{(4)}\left(a_1,a_2,b_1,b_2,b_3;c_1,c_2,c_3;x,y,z,t\right) =$$

$$= \sum_{m,n,p,q=0}^{\infty}\frac{\left(a_1\right)_{m+n+p}\left(a_2\right)_q\left(b_1\right)_{m+n}\left(b_2\right)_p\left(b_3\right)_q}{\left(c_1\right)_m\left(c_2\right)_{n+q}\left(c_3\right)_p}\frac{x^m}{m!}\frac{y^n}{n!}\frac{z^p}{p!}\frac{t^q}{q!},$$

$$u_2\left(x,y,z,t\right) = x^{1-c_1}F_{24}^{(4)}\left(1-c_1+\gamma+a_1,a_2,b_1+1-c_1,b_2+\gamma,b_3;2-c_1,c_2,c_3+2\gamma;x,y,z,t\right) =$$

$$= \sum_{m,n,p,q=0}^{\infty}\frac{\left(1-c_1+\gamma+a_1\right)_{m+n+p}\left(a_2\right)_q\left(b_1+1-c_1\right)_{m+n}\left(b_2+\gamma\right)_p\left(b_3\right)_q}{\left(2-c_1\right)_m\left(c_2\right)_{n+q}\left(c_3+2\gamma\right)_p}\frac{x^m}{m!}\frac{y^n}{n!}\frac{z^p}{p!}\frac{t^q}{q!},$$

$$u_3(x,y,z,t)=z^{1-c_3}tF_{24}^{(4)}(\alpha+\gamma+a_1,a_2,b_1+\alpha,b_2+1-c_3,b_3;c_1+2\alpha,c_2,2-c_3;x,y,z,t)=$$
$$=\sum_{m,n,p,q=0}^{\infty}\frac{(\alpha+\gamma+a_1)_{m+n+p}(a_2)_q(b_1+\alpha)_{m+n}(b_2+1-c_3)_p(b_3)_q}{(c_1+2\alpha)_m(c_2)_{n+q}(2-c_3)_p}\frac{x^m}{m!}\frac{y^n}{n!}\frac{z^p}{p!}\frac{t^q}{q!},$$

$$u_4(x,y,z,t)=x^{1-c_1}z^{1-c_3}F_{24}^{(4)}(1-c_1+1-c_3+a_1,a_2,b_1+1-c_1,b_2+1-c_3,b_3;2-c_1,c_2,2-c_3;x,y,z,t)=$$
$$=\sum_{m,n,p,q=0}^{\infty}\frac{(1-c_1+\gamma+a_1)_{m+n+p}(a_2)_q(b_1+1-c_1)_{m+n}(b_2+1-c_3)_p(b_3)_q}{(2-c_1)_m(c_2)_{n+q}(2-c_3)_p}\frac{x^m}{m!}\frac{y^n}{n!}\frac{z^p}{p!}\frac{t^q}{q!},$$

Төрт айнымалы гипергеометриялық функциясы $F_{25}^{(4)}$ келесі теңдеулер жүйесін қанағаттандырады:

$$\begin{cases}x(1-x)u_{xx}+(1-x)tu_{xt}-xyu_{xy}-xzu_{xz}-ytu_{yt}-ztu_{zt}+\left[c_1-(a_1+b_1+1)x\right]u_x-b_1yu_y\\-b_1zu_z-a_1tu_t-a_1b_1u=0\\y(1-y)u_{yy}-xyu_{xy}-yzu_{yz}-b_2xu_x+\left[c_2-(a_1+b_2+1)y\right]u_y-b_2zu_z-a_1b_2u=0\\z(1-z)u_{zz}-xzu_{xz}-yzu_{yz}-xb_3u_x-yb_3u_y+\left[c_3-(a_1+b_3+1)z\right]u_z-a_1b_3u=0\\t(1-t)u_{tt}+(1-t)xu_{xt}-a_2xu_x+\left[c_1-(a_2+b_1+1)t\right]u_t-a_2b_1u=0\end{cases}\quad(12)$$

Мұндағы $u(x,y,z,t)=F_{25}^{(4)}(a_1,a_2,b_1,b_2,b_3;c_1,c_2,c_3;x,y,z,t).$ Дифференциалдық теңдеулер жүйесінің (12) 4 сызықты тәуелсіз шешімі бар.

$$u_1(x,y,z,t)$$
$$=F_{25}^{(4)}(a_1,a_2,b_1,b_2,b_3;c_1,c_2,c_3;x,y,z,t)=\sum_{m,n,p,q=0}^{\infty}\frac{(a_1)_{m+n+p}(a_2)_q(b_1)_{m+q}(b_2)_n(b_3)_p}{(c_1)_{m+q}(c_2)_n(c_3)_p}\frac{x^m}{m!}\frac{y^n}{n!}\frac{z^p}{p!}\frac{t^q}{q!}$$

$$u_2(x,y,z,t)$$
$$=y^{1-c_2}F_{25}^{(4)}(a_1+1-c_2,a_2,b_1,b_2+1-c_2,b_3;c_1,2-c_2,c_3,;x,y,z,t)$$
$$=y^{1-c_2}\sum_{m,n,p,q=0}^{\infty}\frac{(a_1+1-c_2)_{m+n+p}(a_2)_q(b_1)_{m+q}(b_2+1-c_2)_n(b_3)_p}{(c_1)_{m+q}(2-c_2)_n(c_3)_p}\frac{x^m}{m!}\frac{y^n}{n!}\frac{z^p}{p!}\frac{t^q}{q!}$$

$$u_3(x,y,z,t)$$
$$=z^{1-c_3}F_{25}^{(4)}(a_1+1-c_3,a_2,b_1,b_2,b_3+1-c_3;c_1,c_2,2-c_3,;x,y,z,t)$$
$$=z^{1-c_3}\sum_{m,n,p,q=0}^{\infty}\frac{(a_1+1-c_3)_{m+n+p}(a_2)_q(b_1)_{m+q}(b_2)_n(b_3+1-c_3)_p}{(c_1)_{m+q}(c_2)_n(2-c_3)_p}\frac{x^m}{m!}\frac{y^n}{n!}\frac{z^p}{p!}\frac{t^q}{q!}$$

$$u_4(x,y,z,t)$$
$$=y^{\beta}z^{1-c_3}F_{25}^{(4)}(a_1+2-c_2-c_3,a_2,b_1,b_2+1-c_2,b_3+1-c_3;c_1,2-c_2,2-c_3,;x,y,z,t)$$
$$=x^{\alpha}y^{\beta}z^{\gamma}t^{\delta}\sum_{m,n,p,q=0}^{\infty}\frac{(a_1+2-c_2-c_3)_{m+n+p}(a_2)_q(b_1)_{m+q}(b_2+1-c_2)_n(b_3+1-c_3)_p}{(c_1)_{m+q}(2-c_2)_n(2-c_3)_p}\frac{x^m}{m!}\frac{y^n}{n!}\frac{z^p}{p!}\frac{t^q}{q!}$$

Талқылау

Осы зерттеуде ұсынылған төрт айнымалы гипергеометриялық $F_{24}^{(4)}$ $F_{25}^{(4)}$ функцияларының аналитикалық шешімдері қазіргі математикалық физикада күрделі есептерді шешудің жаңа мүмкіндіктерін ұсынады. Біз ұсынған тәсіл көпөлшемді кеңістіктердегі толқындардың таралуы мен жылу тасымалдануын жоғары дәлдікпен модельдеуге мүмкіндік береді. Анықталған шешімдердің иргелі жүйесі кез-келген жалпы шешімді аналитикалық түрде тұжырымдауға мүмкіндік береді.

Анықталған сызықтық тәуелсіз шешімдер мен жіктеу формулалары есептеу тиімділігін айтарлықтай арттырады. Дегенмен, параметрлердің нақты мәндері үшін осы шешімдердің жинақталу облысын нақтылай түсу қосымша зерттеулерді талап етеді. Болашақта, бұл әдісті жоғары ретті операторлық теңдеулеріне дейін қолдану математикалық физикадағы шеттік есептерді шешудің қуатты құралына айналуы мүмкін.

Қорытынды

Бұл зерттеу жұмысында төрт айнымалының гипергеометриялық $F_{24}^{(4)}$ $F_{25}^{(4)}$ функциялары үшін екінші ретті дербес туындылы дифференциалдық теңдеулер жүйесінің шешімдері қарастырылды. Осы жүйелердің төрт сызықтық тәуелсіз шешімдері аналитикалық жолмен алынды. Зерттеудің негізгі ғылыми нәтижесі қарастырылған дифференциалдық теңдеулер жүйесінің төрт сызықтық тәуелсіз шешімін **алғаш рет** аналитикалық түрде алу және оларды көп айнымалы гипергеометриялық функциялар теориясы аясында сипаттау болып табылады.

Осы функцияларды дифференциалдық жүйелердің шешімдері ретінде ұсына отырып, олардың шешу жолдарын тереңірек түсінуге қол жеткізуге болады. Сонымен қатар, қазіргі нәтижелер гипергеометриялық функциялардың көпөлшемді түрлерін зерттеудің маңыздылығын көрсетеді. Төрт айнымалы жағдай үшін алынған шешімдер теорияны көбірек айнымалылардың функцияларына кеңейтуге негіз болады және арнайы функциялардың жаңа кластарын анықтауды жеңілдетеді.Болашақ зерттеулер әртүрлі шекаралық жағдайларда осы шешімдердің тұрақтылығын, олардың асимптотикалық қасиетін және ықтимал интегралды көріністерін ескеруі керек. Мұндай жұмыс гипергеометриялық функциялар мен олардың туындыларының кеңістіктік модельдеуде, спектрлік анализде және математикалық зерттеудің кең салаларында қолданылуын күшейтеді.Бұл зерттеудің ғылыми маңыздылығы алынған аналитикалық шешімдердің гипергеометриялық функциялар теориясын дамытуға қосқан үлесімен айқындалады. Ұсынылған шешімдер математикалық физикаданың көпөлшемді модельдеріне, шеттік есептерге және сингулярлық коэффициенттері бар дифференциалдық теңдеулерді шешуге қолданылады. Қорытындылай келе, бұл зерттеу көп айнымалы функциялар теориясына айтарлықтай үлес қосады және олардың теориялық және практикалық өзектілігі туралы түсінік береді.

Алғыс білдіру

Бұл зерттеуді Қазақстан Республикасы Ғылым және жоғары білім министрлігінің Ғылым комитеті қолдады (Грант BR31715767).

Пайдаланылған дереккөздердің тізімі

[1] Appell P., Kampe de Feriet J. Fonctions hypergeometriques et hyperspheriques. Polynomes d'Hermite / Paris: Gauthier – Villars, 1926. - 434 p. URL: https://books.google.com/books?id=znr7fqnNPyoC

[2] Hasanov A., Ergashev T. Solutions of systems of the partial differential equations of Kampé de Fériet type functions // arXiv preprint arXiv:1803.03116. – 2018. – DOI: 10.48550/arXiv.1803.03116

[3] Exton H. Multiple Hypergeometric Functions and Applications. Halsted Press, London, 1976. URL: https://books.google.com/books?id=QwqoAAAAIAAJ

[4] Hasanov A., Srivastava H.M. Some decomposition formulas associated with the Lauricella function $F_A^{(r)}$ and other multiple hypergeometric functions // Appl. Math. Lett. – 2006. – V. 19. №2. – P. 113–121. DOI: https://doi.org/10.1016/j.aml.2005.03.009

[5] Moiseev, E. I. On the solvability of a nonlocal boundary value problem / E. I. Moiseev // Differential Equations. – 2001. – Vol. 37, No. 11. – P. 1643–1646 DOI: https://doi.org/10.1023/A:1017937403853.

[6] Srivastava H.M., Karlsson P.W. Multiple Gaussian Hypergeometric Series. Halsted Press, Ellis Horwood, Chichester, UK; John Wiley & Sons, New York, 1985. URL: https://archive.org/details/multiplegaussian0000sriv

[7] Berdyshev A.S., Ryskan A.R. Boundary value problem for the four-dimensional Gellerstedt equation. Bulletin of the Karaganda University. – 2021. – V. 4, №104. P. 35-48 DOI: https://doi.org/10.31489/2021M4/35-48.

[8] Hasanov A. Fundamental solutions for degenerated elliptic equation with two perpendicular lines of degeneration // International Journal of Applied Mathematics and Statistics, Vol. 13, No. 8, 2008, pp. 41–49. URL: http://www.ceser.in/ceserp/index.php/ijamas/article/view/448

[9] Hasanov A. Fundamental solutions of generalized bi-axially symmetric Helmholtz equation // Complex Variables and Elliptic Equations, Vol. 52, No. 8, 2007, pp. 673–683. DOI: https://doi.org/10.1080/17476930701300375.

[10] Hasanov A., Karimov E.T. Fundamental solutions for a class of three-dimensional elliptic equations with singular coefficients // Applied Mathematics Letters, – 2009. – №22. – P. 1828–1832. DOI: https://doi.org/10.1016/j.aml.2009.07.006

[11] Hasanov A., Berdyshev A. S., Ryskan A. R. Fundamental solutions for a class of four-dimensional degenerate elliptic equation. ComplexVar. EllipticEqu. – 2020. – V. 65, - Issue 4. P. 632-647. DOI: https://doi.org/10.1080/17476933.2019.1606803

[12] *Reichelt T., Schulze M., Sevenheck C., Walther U. Algebraic Aspects of Hypergeometric Differential Equations. Beiträge zur Algebra und Geometrie / Contributions to Algebra and Geometry. 2021. Vol. 62, pp. 137–203. DOI:* https://doi.org/10.1007/s13366-020-00560-1

[13] Matsubara-Heo S.J., Oshima T. Generalized hypergeometric functions with several variables // Indagationes Mathematicae. – 2025. – Vol. 36, No. 2. – P. 507–566. – DOI: https://doi.org/10.1016/j.indag.2024.07.007

[14] Hasanov A., Younis J., Aydi H. On Decomposition Formulas Related to the Gaussian Hypergeometric Functions in Three Variables // Journal of Function Spaces. – 2021. – Article ID 5538583. DOI: https://doi.org/10.1155/2021/5538583

[15] Arzikulov, Z. O., & Ergashev, T. G. (2025). Particular solutions of the multidimensional singular ultrahyperbolic equation generalizing the telegraph and Helmholtz equations. Bulletin of the Karaganda University. Mathematics Series, 2(118), 16–27. DOI: https://doi.org/10.31489/2025M2/16-27

[16] Bezrodnykh S. I. Constructing bases in solution space of the system of equations for the Lauricella Function $F_D^{(N)}$ // Integral Transforms and Special Functions. – 2023. – Vol. 34, No. 11. – P. 813–834. DOI: 10.1080/10652469.2023.2212396

[17] Bezrodnykh S. I. Formulas for computing the Lauricella function in the case of crowding of variables // Computational Mathematics and Mathematical Physics. – 2022. – Vol. 62, No. 12. – P. 2069–2090. DOI: 10.1134/S0965542522120041

[18] Bezrodnykh S. I. Generalizations of the Jacobi identity to the case of the Lauricella function $F_D^{(N)}$ // Integral Transforms and Special Functions. – 2024. – Vol. 35, No. 10. – P. 521–539. DOI: 10.1080/10652469.2024.2351396

IRSTI 83.03.51

https://doi.org/10.51889/2959-5894.2026.95.3.009

K.M. Shapen [1], A.T. Rakhymova [1*], A. Akhmedov [2], Zh.D. Baishemirov [3, 4]

[1]L.N. Gumilyov Eurasian National University, Astana, Kazakhstan
[2]Dallas College, Dallas, USA
[3]Kazakh-British Technical University, Almaty, Kazakhstan
[4]Narxoz University, Almaty, Kazakhstan
[*]*e-mail: aigerim_rakhimova@mail.ru*

BAYESIAN CAUSAL DEMAND MODELING WITH LATENT RANKING FACTORS IN E-COMMERCE

Abstract

E-commerce demand forecasting is complicated by the effects of latent ranking factors and deliberate price changes that distort observed sales relative to true demand. This paper presents a systematic review of modern approaches to demand modeling from the perspective of integrating probabilistic forecasting, Bayesian uncertainty assessment, and causal analysis. Three key areas are considered: probabilistic deep time series models (DeepAR, Temporal Fusion Transformer, N-BEATS, PatchTST, foundation models Chronos and TimesFM); Bayesian neural networks and ensemble methods for uncertainty assessment; and methods for causal modeling of price effects, including double machine learning and structural causal models. A comparative analysis of methods is conducted based on the criteria for accounting for uncertainty, ranking, and causality. A fundamental gap is identified: the absence of architectures that jointly account for Bayesian uncertainty, position bias, and causal price effects under conditions of unobservable product visibility. A direction for further research is proposed related to the development of integrated Bayesian-causal models with latent ranking variables for reconstructing true demand in e-commerce conditions.

Keywords: demand forecasting, e-commerce, Bayesian neural networks, causal modeling, position bias.

Қ.М. Шапен [1], А.Т. Рахымова [1]А, А. Ахмедов [2], Ж.Д. Байшемиров[3,4]
[1]Л.Н. Гумилев атындағы Еуразия ұлттық университеті, Астана қ., Қазақстан
[2] Даллас колледжі, Даллас, АҚШ
[3]Қазақстан-Британ техникалық университеті, Алматы қ., Қазақстан
[4]Нархоз университеті, Алматы қ., Қазақстан

ЭЛЕКТРОНДЫҚ КОММЕРЦИЯДА ЛАТЕНТТІ ФАКТОРЛАРЫ БАР РАНЖИРЛЕУДІ ЕСКЕРЕТІН СҰРАНЫСТЫ БАЙЕСТІК СЕБЕП-САЛДАРЛЫҚ МОДЕЛДЕУ

Аңдатпа

Электрондық коммерцияға сұранысты болжау жасырын рейтинг факторларының және бақыланатын сатуды шынайы сұранысқа қатысты бұрмалайтын бағаның өзгеруінің әсерінен күрделене түседі. Бұл мақалада ықтималдық болжауды, Байес белгісіздік бағалауын және себеп-салдарлық талдауды біріктіру тұрғысынан сұранысты моделдеуге заманауи тәсілдердің жүйелі шолуы ұсынылған. Үш негізгі бағыт қарастырылады: ықтималдық терең уақыт қатарлары моделдері (DeepAR, Temporal Fusion Transformer, N-BEATS, PatchTST, Chronos және TimesFM негіздік моделдері); Байес нейрондық желілері және белгісіздікті бағалауға арналған ансамбльдік әдістер; баға әсерлерін себептік моделдеу әдістері, соның ішінде Double Machine Learning және құрылымдық себептік моделдер. Әдістердің салыстырмалы талдауы белгісіздік, рейтинг және себеп-салдарлыққа байланысты есепке алу критерийлеріне негізделшп жүргізілген. Байес белгісіздікті, позициялық ауытқуды және тауардың бақыланбайтын көрінгіштігі жағдайында бағаның себеп-салдарлық әсерлерін бірге ескеретін архитектуралардың болмауы - негізгі іргелі кемшілік болып анықталды. Электрондық коммерция жағдайында шынайы сұранысты анықтау үшін жасырын рейтингтік айнымалылары бар интеграцияланған Байес себеп- салдарлық моделдерді әзірлеуге қатысты әрі қарай зерттеу бағыттары ұсынылады.

Түйін сөздер: сұранысты болжау, электрондық коммерция, байес нейрондық желілері, себеп-салдарлық моделдеу, позициялық бейімділік.

К.М. Шапен [1], А.Т. Рахымова [1] , А. Ахмедов [2], Ж.Д. Байшемиров[3]
[1] Евразийский национальный университет имени Л.Н. Гумилева, г.Астана, Казахстан
[2] Даллас колледж, Даллас, США
[3]Казахстанско-Британский технический университет, г.Алматы, Казахстан
[4]Университет Нархоз, г. Алматы, Казахстан

БАЙЕСОВСКО-КАУЗАЛЬНОЕ МОДЕЛИРОВАНИЕ СПРОСА С УЧЁТОМ ЛАТЕНТНЫХ ФАКТОРОВ РАНЖИРОВАНИЯ В ЭЛЕКТРОННОЙ КОММЕРЦИИ

Аннотация

Прогнозирование спроса в электронной коммерции осложняется действием латентных факторов ранжирования и целенаправленных изменений цен, искажающих наблюдаемые продажи относительно истинного спроса. Настоящая работа представляет систематический обзор современных подходов к моделированию спроса с позиций интеграции вероятностного прогнозирования, байесовской оценки неопределённости и каузального анализа. Рассмотрены три ключевых направления: вероятностные глубокие модели временных рядов (DeepAR, Temporal Fusion Transformer, N-BEATS, PatchTST, foundation models Chronos и TimesFM); байесовские нейронные сети и ансамблевые методы оценки неопределённости; методы причинного моделирования ценовых эффектов, включая Double Machine Learning и структурные причинные модели. Проведён сравнительный анализ методов по критериям учёта неопределённости, ранжирования и каузальности. Выявлен фундаментальный пробел - отсутствие архитектур, совместно учитывающих байесовскую неопределённость, position bias и причинные эффекты цены в условиях ненаблюдаемой видимости товара. Предложено направление дальнейших исследований, связанное с разработкой интегрированных байесовско-каузальных моделей с латентными переменными ранжирования для восстановления истинного спроса в условиях электронной торговли.

Ключевые слова: прогнозирование спроса, электронная коммерция, байесовские нейронные сети, каузальное моделирование, position bias.

Introduction

E-commerce is characterized by highly dynamic demand and limited observability of key sales drivers. Unlike traditional retail, where the seller independently determines product assortment, placement, and pricing, in e-commerce, the key driver of sales volume is product placement in search results, determined by the platform's algorithm and not directly observable. In these conditions, the central management task becomes inventory planning and pricing based on historical sales, which represent a distorted realization of true customer preferences. A fundamental feature of e-commerce data is the systematic discrepancy between observed demand and true demand. Observed sales y_t are formed as the product of latent demand d_t and latent visibility v_t, where v_t depends on the product position, competition, and the parameters of the ranking algorithm. In scientific literature, this phenomenon is described using the terms position bias and exposure bias. Position bias reflects the empirically established decrease in the likelihood of a click and purchase as a product moves away from the top search results, regardless of its relevance. Exposure bias means that products with low visibility do not receive enough impressions to realize their potential, resulting in their historical sales being systematically underestimated relative to actual demand [1-4].

For forecasting and decision-making purposes, latent visibility acts as an unobservable hidden factor. Traditional time series models interpret a decline in sales as a drop in demand, while in reality it may be due to a change in the ranking algorithm, increased competition, or a redistribution of traffic between categorie. Ignoring this factor leads to systematic errors in assessing trends, seasonality, and reactions to promotions, as well as to a decrease in the generalizing ability of models during periods of structural changes in the platform. Existing approaches to demand forecasting only partially address this problem. Classical statistical methods, including ARIMA, ETS, and Prophet, provide interpretable results on stationary series and are widely used due to their computational simplicity. However, they are primarily focused on obtaining point forecasts and do not provide a reliable estimate of the uncertainty needed to calculate safety stocks in highly volatile environments [5].

The development of deep learning has led to the emergence of probabilistic architectures capable of modeling the full predictive distribution. The DeepAR model implements autoregressive probabilistic forecasting based on recurrent networks, the Temporal Fusion Transformer integrates an attention mechanism that accounts for static and dynamic covariates, the N-BEATS and N-HiTS architectures enable series decomposition without manual feature engineering, and PatchTST demonstrates effectiveness on long sequences thanks to its patch representatio. The advent of foundation models, including Chronos and TimesFM, has enabled the use of pretraining on large time series corpora to improve the robustness of forecasts with limited observation history [6-11].

Despite the progress made in the accuracy of probabilistic forecasts, these models retain a common methodological premise: observed sales are considered a direct manifestation of demand. Price, promotional activity, and search ranking are introduced as standard correlation factors without distinguishing their causal nature. As a result, the models reproduce the observed patterns, but do not allow for a correct assessment of the consequences of targeted price changes, since they do not take into account the indirect influence of price on visibility through the ranking mechanism [12].

Of particular interest in this regard are Bayesian machine learning methods, which allow for quantitative assessment of prediction uncertainty and account for incomplete observed information. The use of Bayesian inference with variational approximations and ensemble methods provides a basis for joint modeling of latent demand, uncertainty parameters, and hidden visibility factors. Dropout-based approaches, as a Bayesian approximation, and deep ensembles enable scalable estimation of uncertainty associated with insufficient model knowledge in neural network architectures [13, 14].

Causal analysis methods based on the framework of structural causal models, directed acyclic graphs, and do-calculus provide a formal toolkit for estimating the effects of interventions. The Double Machine Learning approach provides consistent estimates for high-dimensional latent factors, provided they are fully observable. However, the application of these methods in e-commerce is limited because the key latent variable -product visibility -is latent and depends on unobservable parameters of the ranking algorithm [15-17].

To account for unobservable hidden factors, causal inference methods using proxy variables have been developed in recent years. These approaches utilize auxiliary proxy variables and bridge functions to identify causal effects in the presence of hidden factors. In e-commerce, proxy-based methods allow for reconstructing the impact of visibility based on external position measurements, impression data, and other indirect signals, creating the conditions for accurately estimating price elasticity in latent rankings [18, 19].

Thus, in the current literature, there is a significant gap between three approaches: probabilistic time series forecasting, Bayesian uncertainty assessment, and causal modeling of control actions. None of the existing approaches simultaneously accounts for probabilistic uncertainty, latent ranking factors, and the causal effects of pricing decisions. From a methodological perspective, the demand forecasting problem is transformed into the problem of reconstructing latent demand from indirect observations under conditions of uncertainty and structural causal dependencies [20].

This paper aims to systematize these areas and develop a unified theoretical model for demand analysis under conditions of hidden ranking and targeted price changes. The paper sequentially examines probabilistic deep models, Bayesian uncertainty assessment methods, approaches to causal modeling, and methods for accounting for positional bias. It then formulates a research agenda aimed at developing integrated Bayesian-causal architectures with latent visibility variables.

Research methodology

Classical demand forecasting methods

Historically, retail demand forecasting has relied on statistical time series models developed for conditions of relatively stable demand and full observability. ARIMA models and their seasonal extensions, SARIMA, formalize the dependence of the current value on past observations and forecast errors. Their strength lies in their interpretability and theoretical validity for stationary processes. In

e-commerce, these assumptions are often violated due to promotions, changes in ranking algorithms, the emergence of new competitors, and other factors that lead to structural breaks in the time series.

Exponential smoothing in ETS variants provides more flexible adaptation to trends and seasonality and is widely used due to its computational simplicity. Prophet is an additive time series model designed for automated accounting of trends, seasonality, and calendar effects. These models remain popular for practical applications due to their ease of implementation and low computational costs, but they are primarily focused on producing point forecasts and do not provide a reliable estimate of uncertainty.

A separate direction is associated with the use of gradient boosting, in particular XGBoost and LightGBM. With proper feature preparation, including previous sales values, moving averages, and calendar effects, these methods often outperform classical statistics in accuracy. Their limitation is the lack of a built-in probabilistic inference mechanism. Confidence intervals are constructed retroactively using quantile regression, which does not solve the problem of separating random and model uncertainty.

From a methodological perspective, the key limitation of the approaches considered is the assumption that observed sales represent an unbiased realization of demand $y_t = d_t + \varepsilon_t$. In e-commerce, this assumption is not met, since sales are distorted by visibility effects $y_t = d_t \cdot v_t$. The likelihood of purchase depends not only on consumer preferences but also on algorithmic selection and ranking mechanisms, which are described in the literature as position bias and exposure bias. As a result, models are trained on ranking noise, which reduces their robustness to changes in the distribution algorithms.

A comparative table illustrating the capabilities of classical methods is given below (Table 1.).

Table 1. Comparative analysis of classical methods

Method	*Considers seasonal effects*	*Probabilistic forecast*	*Incorporates ranking*	*Causal inter-pretability*	*Inter-pretability*
ARIMA	*Yes*	*No*	*No*	*No*	*High*
ETS	*Yes*	*No*	*No*	*No*	*High*
Prophet	*Yes*	*Partially*	*No*	*No*	*Medium*
XGBoost	*Partially*	*No*	*No*	*No*	*Low*

The table illustrates that none of the classical approaches simultaneously provides probabilistic estimation, accounting for ranking effects, and the possibility of causal interpretation. This circumstance necessitates a transition to probabilistic deep models and, further, to Bayesian and latent architectures capable of modeling v_t as a latent variable, as discussed in the following sections.

Deep Learning Models for Time Series

The development of deep learning has led to the emergence of a new generation of models for probabilistic forecasting and working with heterogeneous e-commerce data. Unlike classical approaches, these architectures are capable of training on thousands of product series simultaneously and accounting for complex nonlinear relationships.

The first widely adopted solution was DeepAR, which implements probabilistic forecasting based on recurrent networks. The model is trained to predict the parameters of the distribution of future sales, which allows for obtaining quantile estimates necessary for calculating safety stock. The most commonly used distributions are the negative binomial or Gaussian distributions, depending on the nature of the forecast series. Thanks to its probabilistic forecasting and ability to scale to a large number of product series, DeepAR has become one of the most widely used architectures in demand forecasting problems [6].

Further development involves the Temporal Fusion Transformer, which combines an attention mechanism with explicit consideration of static product characteristics and time-varying factors such

as price and promotional activity. A key advantage of TFT is its relative interpretability compared to other deep architectures, as the attention mechanism allows for the analysis of both feature importance and the time intervals that have the greatest impact on prediction [7].

The N-BEATS and N-HiTS architectures offer an alternative approach based on deep time series decomposition through stacks of fully connected blocks. N-BEATS provides automatic trend and seasonality separation without manual feature engineering, while N-HiTS complements this principle with hierarchical interpolation to improve accuracy over long horizons. Both models demonstrate high performance with limited external data [8].

A significant shift occurred with the advent of transformer solutions for long sequences. PatchTST implements a series partitioning into patches followed by the application of a self-attention mechanism, which allows for the effective capture of long-term dependencies and outperforms recurrent architectures on data with pronounced seasonality [9].

At the same time, the Time Series Foundation Models (TSFM) movement is being developed, represented by the Chronos, TimesFM, and Lag-Llama models. Unlike the traditional approach, where a model is trained on data from a single company, foundation models utilize knowledge extracted from a large number of heterogeneous time series, which allows for increased forecast robustness in the face of limited observation history [10, 11].

A particularly promising approach is combining time series with textual and visual representations of a product. Information from reviews, descriptions, and images allows for modeling hidden product quality characteristics that are not directly reflected in sales history. Thus, multimodal models partially address the problem of reconstructing latent demand.

It should be noted that the development of deep models has been accompanied by a significant increase in forecasting accuracy and the emergence of probabilistic estimates. However, the central object of forecasting has remained unchanged - most architectures are still focused on modeling observed sales y_t. The issue of separating true demand d_t from visibility v_t and ranking effects remains largely unresolved.

Despite progress, modern deep learning models treat price, stock, and search ranking as mere correlated features, ignoring their causal nature and the influence of hidden ranking factors. As a result, models are trained on data systematically skewed by position and exposure bias [1-3].

Thus, the development of deep learning has provided significant progress in forecasting observed sales, but has not resolved the fundamental problem of identifying true demand in the context of algorithmically driven product visibility and latent ranking factors. This limitation necessitates a transition from deterministic deep architectures to probabilistic models with explicit representations of uncertainty and latent variables, discussed in the next section.

Bayesian Approaches to Uncertainty Estimation

The transition from point forecasts to probabilistic estimates is a key development in e-commerce demand models. For inventory management tasks, knowing expected sales is insufficient; a quantitative assessment of the risk of stockouts and excess inventory is required. Bayesian machine learning methods allow for the formal consideration and quantification of various sources of forecast uncertainty.

It is common to distinguish between random and model components of uncertainty. Let y_t be sales at time t, and x_t be covariates (price, promotion). The predictive distribution is decomposed as:

$$p(y_t|x_t, \mathcal{D}) = \int p(y_t|x_t, \omega)\, p(\omega|\mathcal{D}) d\omega \quad (1)$$

where $p(y_t|x_t, \omega)$ reflects random noise, and $p(\omega|\mathcal{D})$ is the uncertainty of the parameters ω given D.

The first reflects the unavoidable variability of consumer behavior, the second is associated with a lack of data, which is especially important for new products and periods of change in ranking algorithms.

The basic tool is Bayesian neural networks, in which weights are represented by distributions. Training is achieved through variational inference with optimization of the lower variational bound:

$$\mathcal{L}(\theta) = \mathbb{E}_{q_\theta(\omega)}[\log p(\mathcal{D}|\omega)] - \mathrm{KL}\big(q_\theta(\omega)||p(\omega)\big) \quad (2)$$

where $\mathcal{L}(\theta)$ is the variational lower bound of the marginal likelihood logarithm to be maximized with respect to the parameters θ;

θ is the parameter of the variational approximation;

ω is the vector of model parameters;

$q_\theta(\omega)$ is the variational distribution approximating the true posterior distribution $p(\omega|\mathcal{D})$;

$\mathbb{E}_{q_\theta(\omega)}[\cdot]$ is the mathematical expectation for the distribution $q_\theta(\omega)$;

$p(\mathcal{D}|\omega)$ is the likelihood function of the training data D with fixed parameters ω;

$\log p(\mathcal{D}|\omega)$ is the logarithm of the likelihood, characterizing the quality of the data approximation;

$p(\omega)$ is the prior distribution of the model parameters;

$\mathrm{KL}\big(q_\theta(\omega)||p(\omega)\big)$ is the Kullback-Leibler divergence between the variational distribution and the prior, acting as a regularizer.

The first term corresponds to the expected log-likelihood and ensures consistency with the data, the second term penalizes deviation from a priori assumptions and controls for model complexity.

In practice, the MC Dropout method, interpreted as approximate Bayesian inference, has become widely used. Instead of a normal approximation, empirical quantiles of the predictive distribution are used, which are more robust to asymmetric sales distributions.

With *T* stochastic passes, is constructed a quantile interval:

$$\hat{Q}_{\alpha/2} = \mathrm{Quantile}_{\alpha/2}\left(\left\{\hat{y}_t^{(i)}\right\}\right), \hat{Q}_{1-\alpha/2} = \mathrm{Quantile}_{1-\alpha/2}\left(\left\{\hat{y}_t^{(i)}\right\}\right) \quad (3)$$

$$C_t^{1-\alpha} = \left[\hat{Q}_{\alpha/2}, \hat{Q}_{1-\alpha/2}\right]$$

where $\hat{Q}_{\alpha/2}$ is the estimate of the lower quantile of level $\alpha/2$ of the predictive distribution;

$\hat{Q}_{1-\alpha/2}$ is the estimate of the upper quantile of level $1-\alpha/2$ of the predictive distribution;

$\mathrm{Quantile}_p(\cdot)$ is the operator for calculating the empirical quantile of level p;

$\left\{\hat{y}_t^{(i)}\right\}$ is the set of forecast samples for time t obtained by the Monte Carlo method, $i = \overline{1,N}$;

$\hat{y}_t^{(i)}$ is the *i*-th forecast realization corresponding to the sample of parameters $\omega^{\{(i)\}} \sim q_\theta(\omega)$;

α is the significance level, $\alpha \in (0,1)$;

$1-\alpha$ is the specified confidence level of the predictive interval;

$C_t^{1-\alpha}$c is the two-sided predictive interval of confidence level $1-\alpha$ for the target variable y_t at time t;

N is the number of Monte Carlo samples

A common alternative is to use deep ensembles of *M* independently trained networks:

$$p(y_t|x_t) \approx \frac{1}{M}\sum_{m=1}^{M} p(y_t|x_t, \omega_m) \quad (4)$$

where $p(y_t|x_t)$ is the predictive distribution of the target variable y_t given the features x_t;

M is the number of Monte Carlo samples;

$\omega_m \sim q_\theta(\omega)$ is the *m*-th sample of the model parameters from the variational distribution;

$p(y_t|x_t, \omega_m)$ is the conditional distribution of the output given the fixed parameters ω_m;

x_t is the vector of input features at time t;

y_t is the predicted value.

For time series forecasting, Bayesian DeepAR and Bayesian TFT have been proposed. In DeepAR, the network models the parameters of the sales distribution:

$$y_t \sim \text{NegBin}\left(\mu_\theta(h_t), \sigma_\theta^2(h_t)\right), \qquad h_t = f_\theta(y_{1:t-1}, x_{1:t}) \qquad \theta \sim q_\theta \tag{5}$$

where y_t is the target variable at time t;
$\mu_\theta(h_t), \sigma_\theta^2(h_t)$ are the mean and variance calculated by the network from the hidden state h_t;
h_t is the hidden state;
f_θ is the recurrence function (LSTM/GRU);
$y_{1:t-1}$ are the past values of y
$x_{1:t}$ are the past and current features;
$\theta \sim q_\theta$ are the parameters sampled from the variance distribution.

Or in general terms

$$y_t \sim p_\theta(y_t|x_t) \tag{6}$$

where $y_t \sim p_\theta(y_t|x_t)$ is the target variable y_t that follows the parametric distribution p_θ given x_t;
p_θ is a family of distributions;
θ are the distribution parameters, $\theta \sim q_\theta$;
x_t is generally the entire context: $x_t = \{y_{1:t-1}, x_{1:t}\}$;
$h_t = f_\theta(y_{1:t-1}, x_{1:t})$ is the deterministic latent representation of the context.

For inventory control, the quantile of the predictive distribution on the supply horizon L is used. The safety stock is defined as:

$$s_t = Q_\tau(y_{t:t+L}|x_t) - \mathbb{E}[y_{t:t+L}|x_t] \tag{7}$$

where s_t is the measure of forecast asymmetry/bias at time t;
$Q_\tau(y_{t:t+L}|x_t)$ is the τ -quantile of the predictive distribution of the sum (or trajectory) $y_{t:t+L}$ given x_t;
$\tau \in (0,1)$ is the quantile level (τ =0.9 or 0.95 for the right tail);
$y_{t:t+L} = \sum_{k=0}^{L} y_{t+k}$ or the vector $(y_t, \dots, y_{t+L})$ is the forecast horizon of length L;
$\mathbb{E}[y_{t:t+L}|x_t]$ is the mathematical expectation of the same predictive distribution;
x_t is the context at time t.
To calibrate confidence intervals, prediction methods with guaranteed coverage are used, providing finite-sample coverage guarantees at the $1-\alpha$ level, provided that the order of observations does not affect their distribution:

$$P\left(y_{t+1} \in C_t(x_{t+1})\right) \geq 1 - \alpha \tag{8}$$

In the context of marketplaces, Bayesian inference allows one to treat true demand d_t as a latent variable. Observed sales are modeled taking into account latent visibility:

$$y_t \sim p(y_t|d_t, v_t), \qquad d_t \sim p(d_t|x_t, w_d), \qquad v_t \sim p(v_t|p_t, r_t, w_v) \tag{9}$$

where v_t characterizes the probability of displaying a product to a user and depends on platform ranking factors, including price – p_t and rating – r_t. The joint posterior is approximated through variational inference:

$$p(d_{1:T}, v_{1:T}, w_d, w_v|y_{1:T}) \propto \prod_{t=1}^{T} p(y_t|d_t, v_t)p(d_t|x_t, w_d)p(v_t|p_t, r_t, w_v)p(w_d)p(w_v) \tag{10}$$

where, $d_{1:T} = (d_1, \dots, d_T)$ is the latent demand trajectory for period $1{:}\,T$;
$v_{1:T} = (v_1, \dots, v_T)$ is the latent supply/visibility trajectory for period $1{:}\,T$;
w_d, w_v are the parameter vectors of the supply and demand models;

$y_{1:T}$ are the observed data (sales) for period $1:T$;

$p(y_t|d_t, v_t)$ is the likelihood of the observation at time t given demand d_t and supply v_t;

$p(d_t|x_t, w_d)$ is the demand model dependent on features x_t and parameters w_d;

$p(v_t|p_t, r_t, w_v)$ is the supply model, depending on the price p_t, rating/relevance r_t, and parameters w_v;

$p(w_d)p(w_v)$ are the prior distributions of the parameters;

x_t are the exogenous demand features at time t;

p_t is the price at time t;

r_t is the rating or visibility measure at time t.

Joint estimation allows us to separate demand variation due to customer preferences from sales variation caused by ranking algorithms.

Thus, Bayesian approaches provide a quantitative assessment of forecast uncertainty and enable modeling of hidden factors influencing observed sales. This provides the foundation for moving from sales forecasting to true demand assessment and forms the methodological basis for the further application of latent and causal models.

Causal Modeling of Price Effects with Latent Ranking

Standard forecasting models estimate the conditional distribution $p(y_t|p_t, x_t)$. Price control requires an intervention distribution $p(y_t|do(p_t), x_t)$. The difference stems from the causal structure of the marketplace, where price influences sales across two channels.

The structural causal model is formalized with an explicit separation of the intervening variable and the common factor. Let u_t be the latent platform factor (e.g., a seasonal category trend, a marketplace promotional campaign), and v_t be the product visibility:

$$\begin{gathered} u_t \to p_t, u_t \to v_t, u_t \to y_t, \\ p_t \to d_t, p_t \to v_t, d_t \to y_t, v_t \to y_t, \end{gathered} \tag{11}$$

Here v_t is a mediator of the price effect through ranking, and u_t is an unobserved latent factor that creates a spurious correlation between price and sales.

To estimate the causal effect in a partially linear setting, the model is used:

$$y_t = \theta_0 p_t + g(x_t, u_t) + \varepsilon_t \tag{12}$$

where θ_0 represents the average price elasticity. When u_t is unobservable, the standard estimate is biased. Chernozhukov's Double Machine Learning method allows one to obtain an unbiased estimate under the condition that hidden factors are observable [16]. Two-step procedure:

$$\begin{gathered} \hat{l}(x_t) = \mathbb{E}[p_t|x_t], \hat{m}(x_t) = \mathbb{E}[y_t|x_t] \\ \hat{\theta} = \left(\frac{1}{n}\sum_t (p_t - \hat{l}_t)^2\right)^{-1} \frac{1}{n}\sum_t (p_t - \hat{l}_t)(y_t - \hat{m}_t) \end{gathered} \tag{13}$$

In the presence of latent u_t this condition is violated.

To identify unobservable hidden factors, methods of proximal causal inference are used. Let z_t be a proxy variable for v_t (e.g., an external position estimate), and w_t be a proxy for u_t. Within the Miao-Tchetgen Tchetgen approach, a bridge function h is introduced that satisfies the integral equation [18]:

$$\mathbb{E}[y_t|p_t, z_t] = \int h(p_t, w)\, h(w|z_t) dw \tag{14}$$

This equation is a Fredholm equation of the first kind. If a solution exists, the causal effect is identified as:

$$\mathbb{E}[y_t|do(p^*)] = \int h(p^*, w)\, p(w) dw \tag{15}$$

Under the Bayesian causal framework with Pearl's do-operator, truncated factorization is applied. The joint model [15]:

$$p(y_t, d_t, v_t, u_t|p_t) = p(y_t|d_t, v_t, u_t)p(d_t|p_t, u_t)p(v_t|p_t, u_t)p(u_t) \quad (16)$$

Following the $do(p^*)$ intervention, the distribution is:

$$p\big(y_t\big|do(p^*)\big) = \iiint p(y_t|d_t, v_t, u_t)p(d_t|do(p^*), u_t)p(v_t|do(p^*), u_t)p(u_t)\, dd_t dv_t du_t \quad (17)$$

where $p(d_t|do(p^*), u_t) = p(d_t|p^*, u_t)$ and $p(v_t|do(p^*), u_t) = p(v_t|p^*, u_t)$ according to the truncated factorization rule. The expected effect:

$$\mathbb{E}\big(y_t\big|do(p^*)\big) = \iiint \mathbb{E}[y_t|d_t, v_t, u_t]p(d_t|p^*, u_t)p(v_t|p^*, u_t)p(u_t)\, dd_t dv_t du_t \quad (18)$$

This formula clearly separates the direct effect $p^* \to d_t$, the mediation effect $p^* \to v_t \to y_t$ and the influence of the latent hidden factor u_t.

Thus, correct causal modeling requires distinguishing between the ranking mediator and the latent underlying platform factor. Proximal inference methods enable identification in the presence of proxy variables, while the Bayesian formulation allows for the integration of parameter uncertainty assessment with causal identification.

Latent Models for Ranking and Visibility

Observed sales y_t are the result of the interaction of true demand d_t and algorithmic visibility v_t. The seller does not directly observe v_t, which leads to the problem of reconstructing hidden states from distorted observations.

The basic generative model is formulated as a hierarchical system:

$$d_t \sim p(d_t|x_t, w_d),\ v_t \sim p(v_t|p_t, r_t, u_t, w_v),\ y_t \sim p\big(y_t\big|d_t, v_t, w_y\big) \quad (19)$$

where u_t is the platform's latent factor. For counted sales, a visibility-controlled distribution is used:

$$y_t \sim \text{Poisson}(d_t \cdot v_t) \text{ or } y_t \sim \text{NegBin}(d_t \cdot v_t, \phi) \quad (20)$$

Visibility is modeled as a latent random variable on the interval *(0,1)*. A beta distribution with parameters dependent on the ranking factors is used:

$$\begin{gathered} \text{logit}\big(\mu_{v,t}\big) = \alpha_p \log p_t + \alpha_r r_t + \beta^T u_t \\ v_t \sim \text{Beta}(a_t, b_t), a_t = \mu_{v,t}\tau, b_t = (1 - \mu_{v,t})\tau \end{gathered} \quad (21)$$

where τ is the concentration parameter, the concentration parameter, $\alpha_p < 0$ reflects the penalty for a high price. Alternative parameterization via the logit-normal distribution:

$$\text{logit}(v_t) \sim \mathcal{N}\big(\alpha_p \log p_t + \alpha_r r_t + \beta^T u_t, \sigma_v^2\big) \quad (22)$$

Variational inference is used to estimate latent variables. Approximation of the posterior distribution:

$$q_\phi(d_{1:T}, v_{1:T}, u_{1:T}) = \prod_t q_{\phi_d}(d_t) q_{\phi_v}(v_t) q_{\phi_u}(u_t) \quad (23)$$

Variational lower bound on the log-likelihood for the latent ranking model:

$$\mathcal{L}(\phi, w) = \mathbb{E}_q\big|\sum_t \log p(y_t|d_t, v_t)\big| - \text{KL}(q_{\phi_d}||p_{w_d}) - \text{KL}(q_{\phi_v}||p_{w_v}) - \text{KL}(q_{\phi_u}||p_u) \quad (24)$$

Optimization is carried out using stochastic gradient descent with the reparameterization trick.

To take into account the temporal dynamics, a state space model is introduced:

$$d_t = p_d d_{t-1} + \eta_{d,t}, \text{logit}(v_t) = p_v \text{logit}(v_{t-1}) + \eta_{v,t}, \eta_{\cdot,t} \sim \mathcal{N}(0, \sigma^2)$$
$$y_t \sim p(y_t | d_t, v_t) \quad (25)$$

Hidden state filtering is performed using a particle filter or a variational Kalman filter. The posterior distribution:

$$p(d_t, v_t | y_{1:t}) \propto p(y_t | d_t, v_t) \int p(d_t, v_t | d_{t-1}, v_{t-1}) p(d_{t-1}, v_{t-1} | y_{1:t}) dd_{t-1} dv_{t-1} \quad (26)$$

To identify the model in the absence of direct observations v_t, proxy variables z_t and constraints on the structure are used:

$$z_t \sim p(z_t | v_t, \sigma_z^2), \mathbb{E}[v_t] = \bar{v} \quad (27)$$

where $\bar{v}$ is given by a priori knowledge of the average visibility of the category.

Integration with Bayesian forecasting is achieved through a joint predictive distribution:

$$p(y_{t+1} | y_{1:t}) = \int p(y_{t+1} | d_{t+1}, v_{t+1}) p(d_{t+1}, v_{t+1} | y_{1:t}) dd_{t+1} dv_{t+1} \quad (28)$$

This formula allows us to build forecasts that take into account both demand uncertainty and future visibility uncertainty.

Thus, latent ranking modeling enables the reconstruction of true demand from distorted sales observations. The use of hierarchical Bayesian models with time dependencies and proxy variables provides the basis for the correct estimation of price elasticity and inventory management under algorithmic visibility management.

Combining Bayesian and Causal Approaches to Price and Stock Management

Sections 4-6 comprise three independent components: forecast uncertainty assessment, identification of causal price effects, and reconstruction of latent visibility. Marketplace management tasks require their combined use within a unified probabilistic inference framework.

The complete generative model integrates a hierarchical structure:

$$p(y_{1:T}, d_{1:T}, v_{1:T}, w) = \prod_{t=1}^{T} p(y_t | d_t, v_t) p(d_t | p_t, x_t, u_t, w_d) p(v_t | p_t, r_t, u_t, w_v) p(u_t) p(w) \quad (29)$$

where $w = \{w_d, w_v\}$.

Posterior distribution for observed sales:

$$p(d_{1:T}, v_{1:T}, u_{1:T}, w | y_{1:T}) \propto p(y_{1:T}, d_{1:T}, v_{1:T}, u_{1:T}, w) \quad (30)$$

Variational approximation:

$$q_\phi(d, v, u, w) = q_{\phi_w}(w) \prod_t q_{\phi_d}(d_t) q_{\phi_v}(v_t) q_{\phi_u}(u_t) \quad (31)$$

An intervention predictive distribution is used for decision making. When setting the price p^*:

$$p(y_{t+1} | do(p^*), y_{1:t}) = \int p(y_{t+1} | d_{t+1}, v_{t+1}) p(d_{t+1} | p^*, u_{t+1}) p(v_{t+1} | p^*, u_{t+1}) p(u_{t+1} | y_{1:t}) dd_{t+1} dv_{t+1} du_{t+1} \quad (32)$$

This distribution takes into account both the epistemic uncertainty of the parameters w and the stochastic variability of demand and visibility.

The optimal price is determined by maximizing expected utility taking into account risk:

$$p_t^* = \arg\max_p \{\mathbb{E}[\pi(p,y)|do(p),\mathcal{D}] - \lambda \cdot \mathrm{Var}[\pi(p,y)|do(p),\mathcal{D}]\} \quad (33)$$

where $\pi(p,y) = (p - \mathrm{c}) \cdot \min(y,s) - h \cdot \max(0, s-y)$ is a profit function with cost *c*, inventory *s* and storage costs h. The parameter λ reflects the degree of risk aversion.

The intervention distribution quantile is used for inventory control:

$$s_t(p^*) = Q_\tau(y_{t:t+L}|do(p^*), y_{1:t}) \quad (34)$$

The joint optimization of price and stock is formulated as:

$$(p^*, s^*) = \arg\max_{p,s} \mathbb{E}_{p(y|do(p))}[\pi(p,y,s)] \quad (35)$$

with a service level constraint $P(y_{t:t+L} \le s|do(p)) \ge \tau$.

For computation, inference is carried out by sampling from the variational posterior. The algorithm:

1. Sample $w^{(i)} \sim q_{\phi_w}$, $u_{t+1}^{(i)} \sim q_{\phi_u}$
2. For a given p^* generate $d_{t+1}^{(i)} \sim p(d|p^*, u^{(i)}, w_d^{(i)})$, $v_{t+1}^{(i)} \sim p(v|p^*, u^{(i)}, w_v^{(i)})$
3. Generate $y_{t+1}^{(i)} \sim p(y|u^{(i)}, v^{(i)})$
4. Estimate $\mathbb{E}[\pi]$ and $\mathrm{Var}[\pi]$ from the sample $\{y^{(i)}\}_{i=1}^M$.

Conformal calibration is applied to the intervention distribution to provide finite-sample guarantees of coverage of forecast intervals when the pricing policy changes.

Thus, the integration of Bayesian inference, causal identification, and latent ranking modeling forms a methodological basis for decision-making in the face of algorithmic uncertainty in marketplaces. The proposed framework enables the transition from correlational sales forecasting to risk-adjusted optimization of true demand.

Modern Architectures and Research Directions

The integration of Bayesian, causal, and latent components has been implemented in several classes of models developed since 2020.

Probabilistic Transformers. The Temporal Fusion Transformer architecture and its Bayesian variants use an attention mechanism to model long-term dependencies [7]

$$h_t = \mathrm{Attetion}(Q_t, K_{1:t}, V_{1:t}),\ y_t \sim p_\theta(y_t|h_t) \quad (36)$$

The Bayesian version introduces a distribution on the attention parameters $W_Q, W_K, W_V \sim q_\phi$, which allows for estimating forecast uncertainty. For count data, the Negative Binomial or Zero-Inflated Poisson output is used.

Deep state-space models. Deep State Space Models use neural networks to parameterize transition and emission distributions [21]:

$$p(y_*|x_*, C) = \int p(y_*|x_*, z) q(z|C) dz \quad (37)$$

where context C is constructed from similar products. The Bayesian version provides uncertainty estimates for items without history.

Causal Transformers. DeepIV's Causal Transformer and EconML architectures integrate instrumental variables into the attention mechanism [22]:

$$\hat{p}_t = f_1(z_t, x_t),\ \hat{y}_t = f_2(\hat{p}_t, x_t) \quad (38)$$

where z_t is an instrument orthogonal to u_t. Training with orthogonal loss ensures unbiased learning for high-dimensional latent factors.

Principal Constraints of Existing Approaches:

1. Scalability of Bayesian inference. Full variational inference for millions of SKUs (articles) requires stochastic approximations with inducing points or amortized inference.
2. Latent visibility identifiability. Without external proxies z_t for v_t, the model suffers from non-identifiability. Weakly supervised methods based on position data are needed.
3. Non-stationarity of ranking algorithms. The w_t parameters change with platform updates, violating the exchangeability assumption for conformal prediction. Adaptive methods with distribution shift detection are needed.
4. Computational complexity of do-inference. Estimation $\mathbb{E}[y|do(p^*)]$ requires multiple sampling from the joint posterior distribution. For real-time, surrogate models and amortized causal inference are used.

Promising areas include the development of fundamental demand models pre-trained on data from multiple marketplaces with subsequent adaptation through Bayesian retraining, as well as the integration of differential privacy methods for learning on platform data without disclosing ranking algorithms.

Thus, the current state of research is characterized by a transition from isolated forecasting methods to integrated probabilistic-causal systems capable of functioning under conditions of algorithmic uncertainty.

Results of the study

The analysis demonstrates a systematic shift in demand forecasting methods from frequentist ARIMA and ETS models to deep probabilistic architectures. DeepAR, Temporal Fusion Transformer, N-BEATS, and PatchTST models provide scalability to thousands of series and form a baseline for probabilistic forecasts. The introduction of Bayesian methods has enabled a transition from point estimates to quantitative uncertainty assessment with separation of aleatory and epistemic components.

Discussion

A key limitation of existing approaches is the assumption of observability of the data generation process. In marketplaces, this assumption is violated due to algorithmic visibility management. Observed sales y_t are the product of true demand d_t and latent visibility v_t, where v_t depends on price through a ranking mechanism. Standard models interpret the decrease in y_t with an increase in price as a manifestation of price elasticity, ignoring the indirect effect $p_t \to v_t \to y_t$.

A comparative analysis shows the absence of architectures that simultaneously satisfy three requirements: Bayesian estimation of parameter uncertainty, accounting for latent ranking, and identification of the causal effects of price interventions. Double Machine Learning methods provide consistency with observable latent factors, but lose their properties in the presence of latent u_t. Proximal causal methods need valid proxies for v_t - a condition only partially met in real-world marketplaces

Computational limitations are related to the scale of the problem. Full Bayesian inference for the joint distribution $p(d_{1:T}, v_{1:T}, u_{1:T}, w|y_{1:T})$ has complexity $O(T \cdot M)$, where M is the number of samples. Catalogs with millions of SKUs require amortized inference schemes and distributed computation. Additional complexity arises from the non-stationarity of the ranking algorithms, which violates the exchangeability assumption required for conformal guarantees.

Promising areas include the development of fundamental demand models with pre-training on multi-platform data, the integration of differential privacy methods for training without disclosing ranking algorithms, and the creation of a digital twin of demand capable of simulating the consequences of pricing and assortment decisions under conditions of incomplete information.

Conclusion

An analysis of modern literature demonstrates the evolution of e-commerce demand forecasting methods from classical statistical models to deep probabilistic architectures and causal approaches. DeepAR, Temporal Fusion Transformer, N-BEATS, PatchTST, and foundation models have improved the accuracy of probabilistic forecasts and created a foundation for scaling. Bayesian methods have made it possible to formalize uncertainty assessment and separate the aleatory and epistemic components of risk.

It has been established that the key feature of e-commerce platforms - algorithmically controlled product visibility - remains outside the scope of most existing approaches. Observed sales represent a distorted representation of true demand, dependent on an unobservable ranking factor. Standard models interpret changes in visibility as fluctuations in preferences, leading to systematic errors in estimating price elasticity.

A comparative analysis revealed the lack of a universal method that simultaneously provides Bayesian uncertainty assessment, latent ranking considerations, and causal modeling of price interventions. This circumstance identifies a promising research direction related to the development of integrated Bayesian-causal architectures with latent visibility variables.

The proposed research program includes joint demand and ranking modeling, the development of proximal inference methods for unobservable latent factors, and the creation of computationally efficient schemes for intervention forecasting. Implementation of these areas will enable a transition from forecasting observed sales to reconstructing true demand and assessing the consequences of management decisions under conditions of incomplete information.

Thus, further development of the field is associated not with increasing the accuracy of point forecasts, but with the construction of models capable of taking into account the causal structure of marketplace data and ensuring stable decision-making in conditions of algorithmic uncertainty.

References

[1] Joachims T., Granka L., Pan B., et al. Accurately interpreting clickthrough data as implicit feedback. SIGIR Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval, 2005. 154-161. doi.org/10.1145/1076034.1076063

[2] Craswell N., Zoeter O., Taylor M., Ramsey B. An experimental comparison of click position-bias models. Web Search and Data Mining, 2008. 87-94. DOI:10.1145/1341531.1341545

[3] Wang X., Bendersky M., Metzler D., Najork M. Learning to rank with selection bias in personal search. SIGIR, 2016. 115-124. doi.org/10.1145/2911451.2911537

[4] Zohreh Ovaisi, Ragib Ahsan, Yifan Zhang, Kathryn Vasilaky, and Elena Zheleva. 2020. Correcting for Selection Bias in Learning-to-rank Systems. In Proceedings of The Web Conference 2020 (WWW '20), April 20–24, 2020, Taipei, Taiwan. ACM, New York, USA. 11 Pages. DOI: https://doi.org/10.1145/3366423.3380255

[5] Hyndman R. J., Athanasopoulos G. Forecasting: principles and practice, 3rd ed. Melbourne, Australia: OTexts, 2021. Accessed: Dec. 20, 2025. https://otexts.com/fpp3/

[6] Salinas D., Flunkert V., Gasthaus J., Januschowski T. DeepAR: Probabilistic forecasting with autoregressive recurrent networks. International Journal of Forecasting, vol. 36, issue 3, 1181-1191. DOI: 10.1016/j.ijforecast.2019.07.001

[7] Lim B., Arik S.O., Loeff N., Pfister T. Temporal Fusion Transformers for interpretable multi-horizon time series forecasting. International Journal of Forecasting, vol. 37(4), 2021. 1748-1764. DOI: 10.1016/j.ijforecast.2021.03.012

[8] Oreshkin B.N., Carpov D., Chapados N., Bengio Y. N-BEATS: Neural basis expansion analysis for interpretable time series forecasting. ICLR, 2020. 31 Pages. https://doi.org/10.48550/arXiv.1905.10437

[9] Nie Y., Nguyen N.H., Sinthong P., Kalagnanam J. A time series is worth 64 words: Long-term forecasting with transformers. ICLR, 2023. 1-22. doi.org/10.48550/arXiv.2211.14730

[10] Ansari A.F., Stella L., Turkmen C., et al. Chronos: Learning the language of time series. Transactions on Machine Learning Research, 2024. 1-43. DOI:10.48550/arXiv.2403.07815

[11] Abhimanyu D., Weihao K., Rajat S., Yichen Zh. A decoder-only foundation model for time-series forecasting. Proceedings of the 41st International Conference on Machine Learning, PMLR 235, 2024. 10148-10167. doi.org/10.48550/arXiv.2310.10688

[12] Januschowski T., Gasthaus J., Wang Y., et al. Criteria for classifying forecasting methods. International Journal of Forecasting, 36(1), 2020. 167-177. DOI:10.1016/j.ijforecast.2019.05.008

[13] Gal Y., Ghahramani Z. Dropout as a Bayesian approximation: Representing model uncertainty in deep learning. Proceedings of The 33rd International Conference on Machine Learning, PMLR, 2016. 1050-1059. doi.org/10.48550/arXiv.1506.02142

[14] Lakshminarayanan B., Pritzel A., Blundell C. Simple and scalable predictive uncertainty estimation using deep ensembles. 31st Conference on Neural Information Processing Systems (NIPS 2017), USA 2017. doi.org/10.48550/arXiv.1612.01474

[15] Pearl J. Causality: Models, Reasoning, and Inference. Cambridge University Press, 2000. Econometric Theory, 19, 2003, 675–685. DOI:10+10170S0266466603004109

[16] Chernozhukov V., Chetverikov D., Demirer M., Duflo E., Hansen Ch., Newey Wh., Robins J. Double/debiased machine learning for treatment and structural parameters. The Econometrics Journal, Volume 21, Issue 1, 1 February 2018, 1–68. https://doi.org/10.1111/ectj.12097

[17] Bottou L., Peters J., Quiñonero-Candela J., et al. Counterfactual reasoning and learning systems. Journal of Machine Learning Research 14 (2013). 3207-3260. doi.org/10.48550/arXiv.1209.2355

[18] Miao W., Geng Z., Tchetgen Tchetgen E.J. Identifying causal effects with proxy variables of an unmeasured confounder. Biometrika, 2016. DOI:10.1093/biomet/asy038

[19] Eric J Tchetgen, Andrew Ying, Yifan Cui, Xu Shi, Wang Miao. An introduction to proximal causal learning. Statist. Sci. 39(3), 2024. 375-390. DOI: 10.1214/23-STS911

[20] Schölkopf B., Locatello F., Bauer S., Ke N.R., Kalchbrenner N., Goyal A., Bengio Y. Toward Causal Representation Learning. Proceedings of the IEEE, Vol. 109, No. 5, 2021. 612–634. doi.org/10.48550/arXiv.2102.11107

[21] Rangapuram S.S., Seeger M.W., Gasthaus J., Stella L., Wang Y., Januschowski T. Deep State Space Models for Time Series Forecasting. Advances in Neural Information Processing Systems. Vol. 31, 2018. 7785-7794. DOI: https://proceedings.neurips.cc/paper_files/paper/2018/file/5cf68969fb67aa6082363a6d4e6468e2- Paper.pdf

[22] Hartford J., Lewis G., Leyton-Brown K., Taddy M. Deep IV: A Flexible Approach for Counterfactual Prediction. Proceedings of the 34th International Conference on Machine Learning. 2017. 1414-1423. DOI: https://proceedings.mlr.press/v70/hartford17a.html

МАТЕМАТИКАНЫ ОҚЫТУ ӘДІСТЕМЕСІ

МЕТОДИКА ПРЕПОДАВАНИЯ МАТЕМАТИКИ

METHODS OF TEACHING MATHEMATICS

ҒТАХР 27.39.00

https://doi.org/10.51889/2959-5894.2026.95.3.010

А.Б. Биргебаев[1], М.Е. Қыдырақын[1*]

[1]Абай атындағы Қазақ Ұлттық педагогикалық университеті, Алматы қ., Қазақстан
[*]*e-mail: kydyrakynmarina@gmail.com*

ФУНКЦИОНАЛДЫҚ АНАЛИЗДЕН ЕСЕПТЕР ШЫҒАРУДЫҢ ЕРЕКШЕЛІКТЕРІ

Аңдатпа

Функционалдық анализ және нақты айнымалы функциялар теориясы пәндері бойынша педагогикалық институттар мен университеттерде оқытылатын теориялық материалдарды игеруде есептер мен жаттығулар жинағының орны ерекше екені белгілі. Орыс тілінде бұл бағытта әртүрлі деңгейде жазылған есептер жинақтары жеткілікті. Олардың негізгілерінің қатарында Ю.С. Очан, А.А. Кирилов пен А.Д. Гвишиани, С.А. Теляковский, П.Л. Ульянов, А.Н. Бахвалов, М.И. Дьяченко, К.С. Казарян, П. Сифуэнтес, В.А. Петров, Н.Я. Виленкин, М.И. Граев еңбектерін атауға болады. Аталған есептер жинақтарында негізінен жиындар және өлшемдер теориясына көп көңіл бөлініп, көптеген есептердің шығару жолдары көрсетілген. Ал интегралдар теориясы, операторлар теориясы, сепарабельділік, кеңістіктердің толықтығы, компактілік және олардың есеп шығаруға қолданылуы жеткілікті деңгейде қамтылмаған. Қазақ тілінде университеттердің оқу бағдарламасына сай аталған пәндер бойынша жазылған есептер жинағы жоқтың қасы. Бірлі-жарым жарық көрген кітаптардың ішінде функционалдық анализ бойынша Ғ. Ақышевтің еңбегін атап өтуге болады. Бұл еңбекте функционалдық анализдің бірнеше тарауы қамтылып, есептердің шығару жолдары көрсетілген, алайда тиражы өте аз. Зерттеудің мақсаты – функционалдық талдаудың негізгі бағыттарындағы есептерді шешудің әдіснамалық аспектілерін анықтау және жүйелеу, сондай-ақ университеттік оқу курсының мазмұнын ескере отырып, оларды қазақ тілінде баяндау тәсілдерін әзірлеу. Бұл пән математиканың іргелі салаларының бірі болғандықтан, болашақ ғалым-зерттеушілердің (магистранттар, PhD докторанттар) оны терең меңгермей ғылыммен айналысуы мүмкін емес. Функционалдық анализ нақты анализдің жалғасы ретінде қарастырылып, оның есептері өзара байланыста зерттеледі. Осы тұрғыда көптеген есептер әдістемелік жағынан талданып, олардың шығару жолдары көрсетілді. Зерттеу жұмысында жиындар теориясы, өлшем теориясы, Лебег интегралы және функционалдық талдау саласындағы негізгі есептер түрлері жүйеленеді, оларды шешудің әдіснамалық тәсілдері айқындалады және әртүрлі күрделілік деңгейіндегі жаттығулар жүйесі әзірленеді. Алынған нәтижелерді функционалдық талдау пәнін қазақ тілінде оқытуда, сондай-ақ оқу-әдістемелік материалдарды әзірлеуде пайдалануға болады.

Түйін сөздер: жиын, өлшемдер теориясы, интегралдар теориясы, операторлар, компактілік, сепарабельділік, кеңістіктің толықтығы.

А.Б. Биргебаев[1], М.Е. Кыдыракын[1]

[1]Казахский национальный педагогический университет имени Абая, г. Алматы, Казахстан

ОСОБЕННОСТИ РЕШЕНИЯ ЗАДАЧ ПО ФУНКЦИОНАЛЬНОМУ АНАЛИЗУ

Аннотация

Известно, что при освоении теоретического материала по дисциплинам функционального анализа и теории функций действительного переменного в педагогических институтах и университетах особое

место занимают сборники задач и упражнений. На русском языке существует достаточное количество задачников различного уровня по данному направлению. Среди наиболее известных можно отметить труды Ю.С. Очана, А.А. Кирилова и А.Д. Гвишиани, С.А. Теляковского, П.Л. Ульянова, А.Н. Бахвалова, М.И. Дьяченко, К.С. Казаряна, П. Сифуэнтеса, В.А. Петрова, Н.Я. Виленкина, М.И. Граева. В указанных сборниках основное внимание уделяется теории множеств и теории меры, приводятся решения большого количества задач. В то же время теория интегралов, теория операторов, сепарабельность, полнота пространств, компактность и их применение при решении задач освещены недостаточно полно. На казахском языке сборников задач, соответствующих университетским учебным программам по данным дисциплинам, практически не существует. Среди немногочисленных изданий можно отметить труд Г. Акышева по функциональному анализу, в котором охвачены несколько разделов данной дисциплины и приведены решения задач, однако тираж книги крайне ограничен. Цель исследования заключается в выявлении и систематизации методических особенностей решения задач по основным разделам функционального анализа, а также в разработке подходов к их изложению на казахском языке с учетом содержания университетского курса. Поскольку данная дисциплина является одной из фундаментальных областей математики, будущие исследователи (магистранты и PhD-докторанты) не могут заниматься научной деятельностью, не освоив её в достаточной степени. Функциональный анализ рассматривается как продолжение реального анализа, а его задачи изучаются во взаимосвязи. В этом контексте многие задачи были исследованы с методической точки зрения, и приведены их подробные решения. В результате исследования систематизированы основные типы задач по теории множеств, теории меры, интегралу Лебега и функциональному анализу, определены методические подходы к их решению и разработана система упражнений различного уровня сложности. Полученные результаты могут быть использованы при преподовании функционального анализа на казахском языке и при разработке учебно-методических материалов.

Ключевые слова: множество, теория меры, теория интегралов, операторы, компактность, сепарабельность, полнота пространства.

A.B. Birgebaev[1], M.E. Kydyrakyn[1]
[1]Kazakh National Pedagogical University named after Abai, Almaty, Kazakhstan

FEATURES OF PROBLEM SOLVING IN FUNCTIONAL ANALYSIS

Abstract

It is well known that problem books and exercise collections play a special role in mastering theoretical material in functional analysis and the theory of functions of a real variable at pedagogical institutes and universities. In Russian, there exists a sufficient number of problem collections of various levels in this field. Among the most notable are the works of Yu.S. Ochan, A.A. Kirilov and A.D. Gvishiani, S.A. Telyakovsky, P.L. Ulyanov, A.N. Bakhvalov, M.I. Dyachenko, K.S. Kazaryan, P. Sifuentes, V.A. Petrov, N.Ya. Vilenkin, and M.I. Graev. These collections mainly focus on set theory and measure theory, and provide solutions to a large number of problems. However, the theory of integrals, operator theory, separability, completeness of spaces, compactness, and their applications to problem solving are insufficiently covered. In the Kazakh language, there are almost no problem collections corresponding to university curricula in these disciplines. Among the few published works, one may note G. Akyshev's book on functional analysis, which covers several sections of the subject and includes solutions to problems, although its circulation is very limited. The aim of the study is to identify and systematize the methodological aspects of solving problems in the core areas of functional analysis, as well as to develop approaches to presenting them in the Kazakh language, taking into account the content of the university course. Since this discipline is one of the fundamental areas of mathematics, future researchers (Master's and PhD students) cannot engage in scientific work without mastering it thoroughly. Functional analysis is considered a continuation of real analysis, and its problems are studied in close interconnection. From a methodological perspective, many problems were analyzed and their detailed solutions were provided. The study systematizes the main types of problems in set theory, measure theory, the Lebesgue integral, and functional analysis; defines methodological approaches to solving them; and develops a system of exercises of varying difficulty. The results obtained can be used in teaching functional analysis in the Kazakh language and in developing educational and methodological materials.

Keywords: set, measure theory, integral theory, operators, compactness, separability, completeness of spaces.

Кіріспе

Негізгі ережелер

Зерттеудің негізгі идеясы нақты айнымалы функциялар теориясы мен функционалдық анализдің теориялық мазмұнын есептер шешу үдерісімен өзара сабақтастықта қарастыруға негізделеді. Зерттеу барысында аталған пәндердің негізгі бөлімдері бойынша есептер жүйеленіп, оларды шешуде қолданылатын математикалық ұғымдар, теоремалар мен әдістердің ерекшеліктері айқындалды. Теориялық материалды типтік есептермен байланыстыра ұсыну арқылы математикалық тұжырымдардың есептерді шешудегі қолдану жолдары көрсетілді. Ұсынылған есептер жүйесі білім алушылардың теориялық білімдерін есептерді шешуде қолдануына және функционалдық анализ бойынша қазақ тіліндегі оқу-әдістемелік материалдардың мазмұнын толықтыруға бағытталған.

Функционалдық анализ кәзіргі математиканың іргелі бағыттарының бірі болып табылады және математикалық анализ, дифференциалдық және интегралдық теңдеулер, операторлар теориясы, математикалық физика сияқты бірқатар салалардың теориялық негіздерін құрайды. Функционалдық анализді меңгеру барысында білім алушылардан анықтамалар мен теоремаларды білумен қатар, оларды нақты математикалық есептердішешуде дұрыс қолдана білу талап етіледі.Осыған байланысты теориялық материал мен есептер жүйесінің өзара байланысын қамтамасыз ету функционалды анализі оқытудың маңызды мәселелерінің бірі болып табылады.

Нақты айнымалы функциялар теориясы мен функционалдық анализдің негізгі ұғымдарының абстрактілі сипатта болуы оларды мегеруде есептер мен жаттығулардың маңызын арттырады. Атап айтқанда, жиындар мен өлшенетін функциялар, өлшем теориясы, Лебег интегралы, функционалдық кеңістіктер, сызықтық операторлар, кеңістіктердің толықтығы, сеперабельдігі және компактілік сияқты ұғымдары терең түсіну оларды тиісті есептерді шешу барысында қолданумен тығыз байланысты.Сондықтан теориялық материалды меңгерумен қатар, әрбір тақырыпқа сәйкес есептерді мақсатты түрде іріктеу, оларды белгілі бір жүйеге келтіру және шешу тәсілдерін көрсету қажеттілігі туындайды. Функционалдық анализ бен нақты айнымалы функциялар теориясы бойынша теориялық еңбектер, оқулықтар және есептер жинағы жеткілікті болғанмен, қазақ тілінде білім алатын студенттерге арналған оқу-әдістемелік материалдардың мазмұнын жетілдіру мәселесі өзектілігін сақтап отыр.

Осы бағытта зерттеу жүргізіп, жинақты жазуға түрткі болған белгілі математик, «Нақты анализ» кітабының авторы Қ. Наурызбаевтың мынадай пікірі болды: «… қазақ тілінде жазылған жаттығулар мен есептер жинағы бұл пән бойынша әзірге жоқ. Сондықтан курстың мазмұнын баяндау барысында кейбір маңызды ұғымдарға, фактілерге көңіл аудару мақсатымен оқушыларға тапсырма түрінде және тараулар соңында келтірілген жаттығу есептер мен сұрақтар түрінде бірқатар материал берілді. Бірақ оның көлемі тапшы – бұлар тек үлгі есебінде келтіріліп отыр. Осындай жинақтардың ұлт тілінде жарық көруі – алда тұрған кезекті істердің бірі». Бұл ретте, теориялық материалдарды типтік есептермен байланыстырып, математикалық ұғымдар мен тұжырымдардың есептерді шешуде қалай қалай қолданылатынын жүйелі түрде көрсететін материалдарды қалыптастыру маңызды. Әлбетте, бұл функционалдық анализ бойынша есептерді тек жинақтау емес, оларды теориялық мазмұны, қолданылатын математикалық аппарат және шешу тәсілдері тұрғысынан жүйелеу қажеттігін көрсетеді. Осыған байланысты зерттеудің мақсаты - нақты айнымалы функциялар теориясы мен функционалдық анализдің негігі бөлімдері бойынша есептерді жүйелеу, олардың теориялық материалмен байланысын анықтау және есептерді шешуде қолданылатын негізгі математикалық тәсілдердің ерекшеліктерін көрсету. Мақсатқа жету үшін мынандай зерттеу міндеттері айқындалады: нақты айнымалы функциялар теориясымен функционалды анализдің негізгі бөлімдеріне сәйкес есептерді іріктеу және жүйелеу; есептердің теориялық ұғымдар мен тұжырымдармен байланысын анықтау; типтік есептерді шешуде қолданылатын

математикалық әдістер мен тәсілдерді талдау; теориялық материалды есептер арқылы бекітуге мүмкіндік беретін тапсырмалар жүйесін қалыптастыру; алынған нәтижелерді қазақ тіліндегі оқу-әдістемелік жинаққа қолдану мүмкіндіктерін көрсету.

Зерттеу нысаны – жоғары оқу орындарында нақты айнымалы функциялар теориясы мен функционалдық анализді оқыту үдерісі.

Зерттеу пәні – нақты айнымалы функциялар теориясы мен функционалдық анализ бойынша есептерді іріктеу, жүйелеу және оларды шешуде теориялық нәтижелерді қолдану. Зерттеудің негізгі идеясы теориялық материал мен есептерді өзара сабақтастықа қарастырудан тұрады. Бұл ретте теориялық ұғымдар, анықтамалар мен теоремалар тиісті типтік есептермен байланыстырылып, оларды есепті шешу барысында таңдау мен қолдану жолдары қарастырылады. Мұндай құрылым теориялық білімді септер шешуде қолдану арасындағы байланысты жүйелі түрде көрсетуге мүмкіндік береді.

Зерттеу әдіснамасы

Ұсынылып отырған мақалада зерттеудің әдіснамасын оқу құралының нақты айнымалы функциялар теориясы мен функционалдық анализдің негізгі бөлімдері бойынша негізгі теориялық қағидалары мен есептердің шығару тәсілдері және оларды құрастыру әдістемесі, сондай ақ, теориялық білімді математикалық есептерді шешу барысына қолдануға бағытталған жүйелілік пен сабақтастық ұстанымдары құрайды. Зерттеу барысында теориялық талдау, салыстыру, жүйелеу, жіктеу және жалпылау әдістері қолданылды. Зерттеу материалдарын іріктеу барысында жиындар теориясы, өлшем теориясы, өлшенетін функциялар, Лебег интегралы, функционалдық кеңістіктер, сызықтық операторлар, кеңістіктің толықтығы, сепарабельділігі мен компактылығы сияқты нақты айнымалы функциялар теориясы мен функционалдық анализдің негізгі тақырыптары қамтылды. Әрбір тақырып бойынша теориялық қағидаларды тікелей қажет ететін типтік есептер қарастырылып, олардың шешу жолдары талданды. Осы негізде жазылғанл оқу құралы аталған бағытта елімізде қазақ тілінде жарық көрген алғашқы еңбектердің бірі болып табылады.

Қазіргі уақытта жоғары оқу орындарында қазақ тілінде білім алатын студенттер санының артуы бұл бағыттағы оқу-әдістемелік әдебиеттердің маңыздылығын арттырып отыр. Осыған байланысты құралдың басты мақсаты – білім алушыларға функционалдық анализ бен нақты айнымалы функциялар теориясының іргелі ұғымдарын есептер мен жаттығулар арқылы терең меңгерту, олардың математикалық мәдениетін қалыптастыру және логикалық ойлау қабілеттерін дамыту болып табылады.

Бүгінгі таңда аталған пәндерді терең игермей, математика саласында толыққанды ғылыми зерттеу жүргізу мүмкін емес. Сондықтан ұсынылып отырған жинақтың ғылыми және әдістемелік құндылығы жоғары болатыны белгілі.

Оқу-әдістемелік құрал авторлардың математика, физика және информатика институтында көп жылдар бойы жүргізген дәрістері мен практикалық, сондай-ақ семинарлық сабақтарының негізінде дайындалған. Сонымен қатар бұл еңбек университет студенттеріне, оқытушыларға және математиканы терең меңгеруге қызығушылық танытатын кең аудиторияға арналған.

Оқу құралының құрылымы университеттің оқу бағдарламасына сәйкес жасалған және осы уақытқа дейін орыс және қазақ тілдерінде жарық көрген еңбектер мен интернет ресурстардағы материалдар негізінде құрастырылған.

Құралдың бірінші бөлімі екі тараудан тұрады және жиындар теориясы мен функциялар теориясына қатысты есептер мен жаттығуларды қамтиды. Жиындар теориясы, өлшемдер теориясы, өлшенетін функциялар, Лебег интегралы және функционалдық анализге қатысты есептерді талдау барысында [1]-[6] әдебиеттер пайдаланылды.Теориялық материалдар [7]-[15] әдебиеттер негізінде іріктеліп алынған. Математиканың негізгі ұғымдарының бірі – жиын ұғымы екені белгілі. Жиындар теориясы нақты айнымалы функциялар теориясының іргетасы болып табылады. Бұл ғылым саласы XIX ғасырдың соңында Г. Кантордың еңбектерінен бастау алады. Дегенмен, бүгінгі күнге дейін жиын ұғымына толық әрі қатаң анықтама берілген

жоқ. Оның орнына сипаттамалық анықтамалар қолданылады. Мәселен, Г. Кантор жиынды «көптің біртұтас ретінде ойша біріктірілуі» деп сипаттаған. Жалпы алғанда, жиын деп белгілі бір қасиеттері немесе белгілері бойынша біріктірілген объектілердің жиынтығын айтамыз. Мысал ретінде натурал сандар жиынын, түзудегі нүктелер жиынын, жазықтықтағы түзулер жиынын немесе планеталар жиынын келтіруге болады. Егер кез келген объект үшін оның берілген жиынға жататынын немесе жатпайтынын анықтауға мүмкіндік болса, онда мұндай жиын анықталған деп есептеледі. Жиындар әдетте латын алфавитінің бас әріптерімен белгіленеді:

$$A, B, C, X, Y, \ldots .$$

Жиынды құрайтын объектілер оның элементтері деп аталады. Егер a элементі A жиынына тиесілі болса, онда бұл $a \in A$ түрінде белгіленеді. Ал a элементі A жиынына тиесілі болмаса, онда $a \notin A$ немесе $a \overline{\in} A$ түрінде жазылады. Мысалы, $\mathbb{Q}$ – рационал сандар жиыны болса, онда $\frac{1}{2} \in \mathbb{Q}$, ал $\sqrt{3} \overline{\in} \mathbb{Q}$.

Табиғатта ақырсыз жиындар кеңінен қарастырылады. Мұндай жиындар ашық және тұйық жиындар, сондай-ақ саналымды және саналымсыз жиындар болып жіктеледі. Егер X жиынында ашық жиындардың τ жүйесі (топология) енгізілсе, онда ол топологиялық кеңістікке айналады. Сол сияқты кез келген бос емес X жиынында әрбір x және y элементтер жұбына теріс емес $\rho(x, y)$ нақты саны сәйкес қойылып, төмендегі шарттарды қанағаттандырса, онда бұл функция метрика деп аталады:

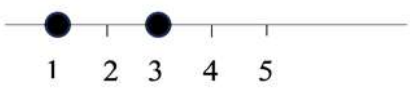
1. $\rho(x, y) > 0$ және $\rho(x, y) = 0 \Leftrightarrow x = y$ (тепе-теңдік аксиомасы);
2. $\rho(x, y) = \rho(y, x)$ (симметрия аксиомасы);
3. $\rho(x, z) \leq \rho(x, y) + \rho(y, z)$ (үшбұрыш аксиомасы).

Осылайша $\rho(x, y)$ саны x және y элементтері арасындағы қашықтықты (яғни X кеңістігінің метрикасын) сипаттайды. Ал көрсетілген үш шарт метриканың аксиомалары деп аталады. Мұндай (X, ρ) жұбы метрикалық кеңістік деп аталады. Метриканы әртүрлі тәсілдермен енгізу арқылы әр алуан қасиеттерге ие кеңістіктерді қарастыруға болады. Осы бағытта ұсынылып отырған жинақта көптеген есептер келтіріліп, олардың ерекшеліктері нақты мысалдар арқылы түсіндіріледі. Мысал ретінде келесі есепті қарастырайық.

Мысал 1. $B_1(x_1, r_1) \subset B_2(x_1, r_1)$ орындалатын метрикалық кеңістікке мысал келтіріңіз, мұнда $r_1 > r_2$ яғни радиусы үлкен шардың радиусы кіші шардың ішінде де жататын метрикалық кеңістіктің мысалын келтіру керек.

Шешуі: $X = \mathbb{N}$ жиынында $\rho(n, m) = |n - m|$ метрикасын қарастырайық. Ашық шарларды төмендегідей таңдайық:

$B_1 = B(1, 4) = \{1,2,3,4\}$, $B_2 = B(3,3) = \{1,2,3,4,5\}$



Сурет 1. $X = \mathbb{N}$ *метрикалық кеңістігіндегі ашық шарлардың өзара орналасуы*

Онда 1-суреттен және ашық шар мәндерінен $B_1 \subset B_2$, бірақ $r_1 > r_2$ екені байқалады.

Математикалық анализ курсынан белгілі болғандай, Риман интегралы негізінен үзіліссіз функциялар және үзіліс нүктелері шектеулі (ақырлы) болатын функциялар үшін анықталған. Алайда қарапайым Дирихле функциясы Риман бойынша интегралданбайды. Осыған байланысты интеграл ұғымын кеңейту қажеттілігі туындап, Лебег интегралы енгізілді. Дегенмен, Лебег интегралы үшін математикалық анализдің классикалық нәтижелерінің бірі – Ньютон-Лейбниц формуласы жалпы жағдайда орындала бермейді. Бұл мәселені шешу мақсатында абсолютті үзіліссіз функциялар ұғымы енгізілген.

Абсолютті үзіліссіз функциялар – кейбір нүктелерде туындысы болмауы мүмкін, бірақ кең мағынада анықталған функциялар класы. Бұл ұғым Ньютон-Лейбниц формуласының

кеңейтілген түрде орындалуын қамтамасыз етеді. Сонымен қатар, ол дифференциалдық және интегралдық есептеулер арасындағы байланысты сақтай отырып, функциялардың өзгерісі нөл өлшемді жиындарда «күрт» өзгермейтін процестерді сипаттайды. $f(x)$ функциясы, Егер Лебег бойынша интегралданатын $f'(x)$ функциясының алғашқы функциясы $[a, b]$ кесіндісінде $f(x)$ функциясы болса онда ол абсолютті үзіліссіз болады . Басқаша айтқанда, келесі теңдік орындалады:

$$f(x) = f(a) + \int_a^x f'(t)dt\,.$$

Керісінше, егер функция осы түрде берілсе және $f'(x)$ Лебег бойынша интегралданатын болса, онда $f(x)$ абсолютті үзіліссіз функция болады.

Абсолютті үзіліссіз функциялар Лебег өлшемі бойынша барлық дерлік нүктелерде туындысы бар функциялармен жұмыс істеуге мүмкіндік береді, тіпті кейбір нүктелерде туындысы анықталмаған жағдайларда да. Мысалы, $y = |x|$ функциясы $x = 0$ нүктесінде туындысы анықталмағанымен, абсолютті үзіліссіз функция болып табылады.

Аталған мәселелерді тереңірек талдау мақсатында оқу құралында бірқатар есептер қарастырылып, олардың кейбіреулерінің шешу жолдары егжей-тегжейлі көрсетілген. Риман және Лебег интегралдарының ерекшеліктері мен өзара байланыстарын айқындау үшін арнайы мысалдар келтірілген. Атап айтқанда, Дирихле функциясының интегралдануы мәселесі арқылы бұл екі интеграл ұғымының айырмашылықтары нақты көрсетіледі.

Оқу құралының екінші бөлімінің бірінші тарауында метрикалық кеңістіктің дербес жағдайлары болып табылатын нормаланған және евклид кеңістіктеріне қатысты есептер қарастырылған. Белгілі болғандай, нормаланған кеңістік толық болған жағдайда Банах кеңістігіне, ал евклид кеңістігі толық (ақырсыз) болған жағдайда Гильберт кеңістігіне айналады. Кеңістіктердің толықтығы мен сепарабелділігі функционалдық анализдің негізгі ұғымдарының бірі болғандықтан, оларды теориялық тұрғыда талдап, есептерде қолдану маңызды. Сонымен қатар, кеңістік ақырсыз өлшемді болған жағдайда шенелгендік ұғымының орнына компакттық және шала компакттық ұғымдары енгізіледі. Бұл ұғымдардың анықтамалары беріліп, оларды есептер шығаруда қолдану жолдары көрсетілген.

Төменде осы ұғымдарды түсіндіруге арналған мысалдар келтірілген.

Мысал 2. $\{x^n\}_{n=1}^{\infty}$ тізбегі $C[0;1]$ кеңістігінде шала компакт емес.

Мысал 3. Төмендегі функциялар жиынының компакттылығын зерттейік:

а) $sin(nx), n = 1,2 \ldots,\ C[0;1]$ кеңістігінде;

б) $sin(\alpha x), \alpha \in [1;2]$.

Шешуі: а) $x = 0$ деп алайық. Онда $y = \frac{\pi}{2n}$ болса,

$$sin(nx) = 0, \qquad sin(ny) = 1.$$

Сонда

$$|x - y| = \left|\frac{\pi}{2n}\right| \to 0 \quad (n \to \infty),$$

бірақ

$$|sin(nx) - sin(ny)| = |0 - 1| = 1.$$

Бұл жағдайда біршенді үзіліссіздік шарты орындалмайды. Демек, берілген функциялар жиыны шала компакт болмайды.

б) Келесі өрнекті қарастырайық.

$$|\sin(\alpha x) - \sin(\alpha y)| = 2\left|sin\frac{\alpha(x-y)}{2}\right|\left|cos\frac{\alpha(x+y)}{2}\right| \le 2\frac{\alpha(x-y)}{2} \le 2|x-y|.$$

Осыдан $\forall \varepsilon > 0\ \exists\ \delta = \frac{\varepsilon}{2} : |x - y| < \delta \Longrightarrow |\sin(\alpha x) - \sin(\alpha y)| \le \varepsilon$ яғни жиын біршенді үзіліссіз. Бұл жерде Арцел-Асколи теоремасы пайдаланылды.

Оқу құралының екінші бөлімінің екінші тарауы операторлар мен функционалдар теориясына арналған. Дифференциалдық теңдеулерді зерттеу барысында ең алдымен олардың шешімдерінің бар болуы, жалғыздығы және корректілігі мәселелерін анықтау қажет екені белгілі. Кез келген дифференциалдық теңдеуді дифференциалдық оператор түрінде жазуға болатындықтан, берілген функционалдық кеңістікте теңдеуді тікелей шешкенге қарағанда, сәйкес операторды қарастырып, оның кері операторының бар-жоғын және шенелгендігін зерттеу жеткілікті болып табылады. Егер шешімдердің тегістігін анықтау керек болса, ондай жағдайда коэрцитвті бағалаулар пайдаланылады [16, 17].

Айталық, X, Y – $K(\mathbb{R}, \mathbb{C})$ өрісіндегі сызықтық кеңістіктер болсын. Егер кез келген $\alpha, \beta \in K$ және $x_1, x_2 \in X$ үшін

$$A(\alpha x_1 + \beta x_2) = \alpha A x_1 + \beta A x_2$$

шарты орындалса, онда $A : X \to Y$ бейнелеуі сызықтық оператор деп аталады.

Мысал 4. $A(x) = kx + b$ операторын қарастырайық. Сызықтық оператор үшін міндетті түрде $A(0) = 0$ шарты орындалуы тиіс. Ал бұл жаңдайда $A(0) = k \cdot 0 + b = b,\ b \neq 0 \Rightarrow A(0) \neq 0$. Демек, $A(x) = kx + b$ – сызықты оператор емес.

Сонымен қатар,

$$A(x_1 + x_2) = k(x_1 + x_2) + b = kx_1 + kx_2 + b,$$

ал

$$A(x_1) + A(x_2) = (kx_1 + b) + (kx_2 + b) = kx_1 + kx_2 + 2b.$$

Осыдан $b \neq 0$ болғанда

$$A(x_1 + x_2) \neq A(x_1) + A(x_2),$$

яғни оператор сызықты емес. Егер $b = 0$ болса, онда $y = kx$ – сызықты оператор.

Енді X, Y – нормаланған кеңістіктер, A – сызықтық оператор болсын. Онда

$$\|A\| = \sup_{x \neq 0} \frac{\|Ax\|}{\|x\|}$$

шамасы операторың нормасы деп аталады. X жиынын Y жиынына бейнелейтін барлық сызықтық операторлар жиынын $L(X, Y)$ деп белгілейді. Егер $\|A\| < \infty$ болса, онда $A \in L(X, Y)$ операторы сызықты шенелген деп аталады. X кеңістігін Y кеңістігіне бейнелейтін барлық шенелген операторлар жиыны $B(X, Y)$ деп белгіленеді.

Тұжырым. Сызықтық операторың нормасы үшін келесі теңдіктер орындалады:

$$\|A\| = \sup_{x \neq 0} \frac{\|Ax\|}{\|x\|} = \sup_{\|x\|=1} \|Ax\| = \sup_{\|x\| \leq 1} \|Ax\|.$$

Ескерту. Кез келген x үшін

$$\|Ax\| \leq \|A\| \|x\|$$

теңсіздігі орындалады. Егер $\|Ax\| \leq ... \leq C\|x\|$ бағалауын жүргізсек, онда $\|A\| \leq C$ шығады. Ал операторың нормасы C шамасына тең болатынын дәлелдеу үшін:

1) Егер $\|x\| = 1$ болатын x саны табылып, $\|Ax\| = C$ орындалса, онда супремум нақты максимумға тең болады, яғни $\|A\| = C$;

2) Егер $\|x_n\| = 1$ болатын тізбек үшін $\|Ax_n\| \leq C$ (төменнен ұмтылса), онда одан $\|A\| = C$ болатыны шығады.

Егер X нормаланған кеңістігінде $|x_0| = 1$ болатын x_0 векторы табылып,

$$\|Ax_0\| = C,$$

және барлық $x \in X$ үшін

$$\|Ax\| \leq C\|x\|$$

орындалса, онда

$$\|A\| = C.$$

Айтылған тұжырымды талдап көрейік:

1. Анықтама бойынша оператор нормасы:

$$\|A\| = \sup_{\|x\|=1} \|Ax\|.$$

2. Айталық, $|x_0| = 1$ болатын x_0 элементі үшін $\|Ax_0\| = C$ орындалсын. Онда анықтама бойынша

$$\|A\| \geq \|Ax_0\| = C.$$

3. Екінші жағынан, егер барлық x үшін $|Ax| \leq C\ \|x\|$ орындалса, онда $\|x\| = 1$ болғанда

$$\|Ax\| \leq C,$$

яғни

$$\sup_{\|x\|=1} \|Ax\| \leq C.$$

4. (2) және (3) нәтижелерін біріктірсек:

$$C \leq \|A\| \leq C,$$

демек,

$$\|A\| = C.$$

Қорытындылай келе, келесі тұжырымдар алынады:

- C саны $|x| = 1$ болғандағы $\|Ax\|$ мәндерінің жоғарғы шекарасы болып табылады;
- Егер осы шекараға жететін x_0 векторы табылса, онда супремум нақты максимумға тең болады;
- Осы жағдайда оператор нормасы дәл осы C санына тең болады.

Ақырлы өлшемді кеңістіктерде (мысалы, A матрица болған жағдайда) $x \longrightarrow |Ax|$ функциясы үзіліссіз, ал $|x| = 1$ сферасы компакт болғандықтан, максимум әрқашан бар болады. Сондықтан бұл жағдайда супремум міндетті түрде максимумға тең. Ал ақырсыз өлшемді кеңістіктерде максимум болмауы мүмкін. Мұндай жағдайда оператор нормасы супремум арқылы анықталып, оған жақындайтын тізбектер көмегімен сипатталады.

Мысал 5 . $A : \mathbb{R}^2 \longrightarrow \mathbb{R}^2$ сызықтық операторы

$$A = \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}$$

матрицасымен берілсін. $\mathbb{R}^2$-дегі евклид кеңістігінің стандартты нормасын қарастырып, оператордың нормасын табайық.

Шешуі:

1-қадам. Операторлың нормасының анықтамасы бойынша:

$$\|A\| = \sup_{\|x\|=1} \|Ax\|.$$

$x = (x_1, x_2)$ болса, онда

$$Ax = (2x_1, x_2), \qquad \|Ax\| = \sqrt{(2x_1)^2 + (x_2)^2} = \sqrt{4x_1^2 + x_2^2}.$$

2-қадам. $|x| = 1$ шарты:

$$x_1^2 + x_2^2 = 1.$$

Осыдан $x_2^2 = 1 - x_1^2$, сондықтан

$$\|Ax\|^2 = 4x_1^2 + x_2^2 = 4x_1^2 + (1 - x_1^2) = 3x_1^2 + 1.$$

3-қадам. Максимумды табу.

$f(x_1) = 3x_1^2 + 1$ функциясы $-1 \le x_1 \le 1$ аралығында $|x_1| = 1$ нүктелерінде максимумға жетеді.

Сондықтан

$$\|Ax\|^2 = 3 + 1 = 4, \qquad \max_{\|x\|=1} \|Ax\| = \sqrt{4} = 2.$$

4-қадам. Максимумның жететінін тексерейік.

$x = (1, 0)$ десек, онда $\|x\| = 1$,

$$Ax = (2, 0), \qquad \|Ax\| = 2.$$

Осыдан норманың максимумы жететіндей $x_0 = (1, 0)$ векторы табылды.

Нәтижесінде

$$\|A\| = 2.$$

Бұл мысал ақырлы өлшемді кеңістікте оператор нормасының максимум арқылы анықталатынын көрсетеді, яғни $x_0 = ((1, 0)), |x_0| = 1$ болатын вектор табылып, векторға оператордың әсерінің нормасы $\|Ax_0\| = 2$ болады және ол мән оператордың нормасымен тең.

Жалпы алғанда, екінші тарауда операторлардың нормасын табу мәселелерімен қатар рефлексивті кеңістіктер, Банах кеңістігіндегі түйіндес операторлар, компакт операторлар, кері операторлар, операторлардың спектрі және спектрлік радиусы сияқты маңызды тақырыптар қарастырылған. Теориялық материалдар нақты мысалдармен және есептердің шешу жолдарымен толықтырылып, олардың функционалдық анализдің дамуына және қолданбалы есептерді шешудегі маңызы жан-жақты сипатталған.

Зерттеу нәтижелері

Зерттеу нәтижесінде нақты айнымалы функциялар теориясы мен функционалдық анализдің негізгі бөлімдері боынша есептер мазмұндық және әдістемелік тұрғыдан жүйеленді.Есептерді шешуде қолданылатын негізгі теориялық ұғымдар мен тұжырымдардың өзара байланысы айқындалып, оларды типтік есептерді шешуге қолдану тәсілдері көрсетілді. Жоғарыда аталған әдебиеттердегі есептер мен теориялық материалдарға зерттеулер жүргізіліп оларды сабақтастыра ұсыну негізінде жиындар теориясы, өлшем теориясы, өлшенетін функциялар, Лебег интегралы, функционалдық кеңістіктер, сызықтық операторлар, кеңістіктің толықтығы, сепарабельділігі мен компактылығы тақырыптары бойынша тапсырмалар кешені қалыптастырылды. Алынған нәтижелер университетіміздің оқу бағдарламасына сай елімізде қазақ тіліндегі алғашқы оқу-әдістемелік құралының бірі ретінде жарық көрген «Функционалдық талдау (жаттығулар мен есептер жинағы) » жинағынының материалдарын жүйелеуге негіз болды.

Талқылау

Зерттеу барысында алынған нәтижелер функционалдық анализді оқытуда теориялық материал мен есептерді біртұтас жүйеде қарастырудың қажеттігін көрсетеді. Функционалдық анализдің көптеген ұғымдарының абстрактілі сипаты білім алушыдан анықтамалар мен теоремаларды білу мен қатар, оларды қандай жағдайда және қандай есептерді шығаруға қолданылатынын түсінуді талап етеді. Зерттеудің практикалық маңыздылығы нақты айнымалы функциялар теориясы мен функционалдық анализ бойынша қазақ тіліндегі оқу-әдістемелік материалдарды толықтыру мүмкіндігімен анықталады. Ұсынылып отырған оқу құралында осы екі пәнді де қамтитын есептер мен жаттығулар жүйелі түрде жинақталған. Айта кету керек, бұл пәндер бойынша қазақ тіліндегі терминология әлі толық қалыптасып үлгермеген. Осыған байланысты терминдердің бірізділігін сақтау мақсатында қазақ тілінде

жарық көрген оқулықтар мен математика саласындағы қазақша-орысша терминологиялық сөздіктер пайдаланылды [16].

Қорытынды

Зерттеу барысында негізгі тақырыптар бойынша есептер жүйеленіп, оларды шешуде қолданылатын теориялық ұғымдар, тұжырымдар және математикалық тәсілдер арасындағы байланыстар айқындалып, өзара сабақтастығын қамтамасыз етудің маңыздығы көрсетілді. Сонымен қатар, оқу-әдістемелікматериалдар аталған пәндердің барлық бөлімдерін толық қамтиды деп айтуға болмайды. Дегенмен, бұл еңбек қазақ тіліндегі ғылыми әдебиетті дамытуға, сондай-ақ болашақта нақты анализ, функционалдық анализ және математиканың басқа да күрделі салалары бойынша ұлттық тілде жарық көретін еңбектердің пайда болуына өз үлесін қосады деп сенім білдіреміз. Жоғарыда айтылғандарды қорытындылай келе, қазақ тілінде «Нақты айнымалы функциялар теориясы» мен функционалдық анализ пәндері бойынша есептер мен жаттығуларды қамтитын «Функционалдық анализ (жаттығулар мен есептер жинағы)» атты оқу құралы жарық көріп, оқырман қауымға ұсынылды (440 бет, 720 есеп пен жаттығулар). Аталған оқу құралының ғылыми-әдістемелік құндылығы жоғарыда келтірілген тұжырымдармен негізделеді. Жұмыс Л.Н. Гумилев атындағы Еуразия ұлттық университеті математика факультетінде (профессор Оспанов Қ.), Дулати атындағы Тараз университеті математика факультетінде (профессор Мұратбеков М.Б.), сондай-ақ Абай атындағы Қазақ ұлттық педагогикалық университеті математика факультетінің ғылыми семинарларында (профессор Бердышев А.С.) талқылаудан өтті.

Пайдаланылған дереккөздердің тізімі

[1] Очан Ю. С. Сборник задач по математическому анализу. Учеб.пособие для студентов физ-мат. фак. пед.ин-тов – М.: Просвещение, 1981, 271 стр.

[2] Глазырина ,П.Ю. Дейкалова,М.В. Коркина, Л. Ф. Функциональный анализ: учебное пособие Уральский федеральный университет. – Екатеринбург : Издательство Уральского университета, 2016. – 214 с. – ISBN 978-5-7996-1771-4.

[3] ТреногинВ.А., Писаревский Б.М.., Соболева Т.С. Задачи и упражнения по функциональному анализу 2-еизд. Физматлит,2005,240 стр.

[4] Ульянов П. Л., Бахвалов А. Н., Дьяченко М. И., Казарян К. С., Сифуэнтес П. Действительный анализ в задачах. – М.: Физматлит, 2016.

[5] Петров В. А., Виленкин Н. Я., Граев М. И. Элементы функционального анализа в задачах. – М.: Просвещение, 1978.

[6] Ақышев Ғ. Функционалдық анализ: есептер жинағы. – Қарағанды: ҚарМУ баспасы, 2018.

[7] Колмогоров А. Н., Фомин С. В. Элементы теории функций и функционального анализа. – М.: Наука, 2010.

[8] Натансон И. П. Теория функций вещественной переменной. – М.:физматлит, 2017.

[9] Треногин В. А. Функциональный анализ. – М.: физматлит, 2005.

[10] Люстерник Л. А., Соболев В. И. Элементы функционального анализа. – М.: Л., 2001.

[11] Наурызбаев Қ. Ж. Нақты анализ (нақты айнымалы функциялар теориясы). – Алматы: Қазақ университеті, 2004.

[12] Блиев Н. Қ. Функционалдық анализ: қысқаша курс. – Алматы: Қазақ университеті, 2014.

[13] Абдикаликова З., Ибатов А. Функционалдық анализ. 1-2 бөлім. – Алматы, 2020.

[14] Шейпак И. А. Функциональный анализ. Ч. 1-2: семинары МГУ. –М., 2016-2021.

[15] Треногин, В. А. Функциональный анализ : учебник / В. А. Треногин. – Москва : ФИЗМАТЛИТ, 2002. – 488 с.

[16] Қазақша-орысша, орысша-қазақша терминалогиялық сөздік (математика). –Алматы: Рауан, 1999.

[17] Birgebayev, A.; Muratbekov, M.. Smoothness of solutions (separability) of the nonlinear stationary schr¨odinger equation. Journal of Mathematics, Mechanics and Computer Science, [S.l.], v. 115, n. 3, p. 25-35,

sep. 2022. ISSN 2617-4871. Date accessed: 30 nov. 2022. doi: https://doi.org/10.26577/JMMCS.2022.v115.i3.03.

[18] Berdyshev A.S., Birgebaev A.B., Cabada A. On the smoothness of solutions of the third order nonlinear differential equation. .//Boundary value problems. May 2017. –P. 1-11. DOI 10.1186/s13661-017-0799-4

[19] Бородин П.А., Савчук А.М. Шейпак И.А. Задачи по функциональному анализу. - Новое изд. - М.: МЦНМО, 2017.- 336 с.

References

[1] Ochan Yu. S. Sbornik zadach po matematicheskomu analizu: ucheb. posobie dlya studentov fiz.-mat. fak. ped. in-tov [Collection of Problems in Mathematical Analysis: Textbook for Students of Physics and Mathematics Faculties of Pedagogical Institutes]. Moscow: Prosveshchenie, 1981. 271p.

[2] Glazyrina P. Yu., Deykalova M. V., Korkina L. F. Funktsional'nyi analiz: uchebnoe posobie [Functional Analysis: Textbook]. Yekaterinburg: Izdatel'stvo Ural'skogo universiteta, 2016. 214 p. ISBN 978-5-7996-1771-4.

[3] Trenogin V. A., Pisarevskii B. M., Soboleva T. S. Zadachi i uprazhneniya po funktsional'nomu analizu [Problems and Exercises in Functional Analysis]. 2nd ed. Moscow: Fizmatlit, 2005. 240 p.

[4] Ul'yanov P. L., Bakhvalov A. N., D'yachenko M. I., Kazaryan K. S., Sifuentes P. Deistvitel'nyi analiz v zadachakh [Real Analysis in Problems]. Moscow: Fizmatlit, 2016.

[5] Petrov V. A., Vilenkin N. Ya., Graev M. I. Elementy funktsional'nogo analiza v zadachakh [Elements of Functional Analysis in Problems]. Moscow: Prosveshchenie, 1978.

[6] Akyshev G. Funktsionaldyq analiz: esepter zhinaǵy [Functional Analysis: Collection of Problems]. Karaganda: QarMU baspasy, 2018.

[7] Kolmogorov A. N., Fomin S. V. Elementy teorii funktsii i funktsional'nogo analiza [Elements of the Theory of Functions and Functional Analysis]. Moscow: Nauka, 2010.

[8] Natanson I. P. Teoriya funktsii veshchestvennoi peremennoi [Theory of Functions of a Real Variable]. Moscow: Fizmatlit, 2017.

[9] Trenogin V. A. Funktsional'nyi analiz [Functional Analysis]. Moscow: Fizmatlit, 2005.

[10] Lyusternik L. A., Sobolev V. I. Elementy funktsional'nogo analiza [Elements of Functional Analysis]. Moscow: Leningrad, 2001.

[11] Nauryzbaev K. Zh. Naqty analiz (naqty aýnymaly funktsiyalar teoriyasy) [Real Analysis (Theory of Functions of a Real Variable)]. Almaty: Qazaq Universiteti, 2004.

[12] Bliyev N. Q. Funktsionaldyq analiz: qysqasha kurs [Functional Analysis: A Short Course]. Almaty: Qazaq Universiteti, 2014.

[13] Abdikalikova Z., Ibatov A. Funktsionaldyq analiz. 1–2 bölim [Functional Analysis. Parts 1–2]. Almaty, 2020.

[14] Sheipak I. A. Funktsional'nyi analiz. Ch. 1–2: seminary MGU [Functional Analysis. Parts 1–2: MSU Seminars]. Moscow, 2016–2021.

[15] Trenogin V. A. Funktsional'nyi analiz: uchebnik [Functional Analysis: Textbook]. Moscow: Fizmatlit, 2002. 488 p.

[16] Qazaqsha-oryssha, oryssha-qazaqsha terminologiyalyq sözdik (matematika) [Kazakh-Russian, Russian-Kazakh Terminological Dictionary (Mathematics)]. Almaty: Rauan, 1999.

[17] Birgebayev, A.; Muratbekov, M.. Smoothness of solutions (separability) of the nonlinear stationary schr¨odinger equation. Journal of Mathematics, Mechanics and Computer Science, [S.l.], v. 115, n. 3, p. 25-35, sep. 2022. ISSN 2617-4871. Date accessed: 30 nov. 2022. doi: https://doi.org/10.26577/JMMCS.2022.v115.i3.03.

[18] Berdyshev A.S., Birgebaev A.B., Cabada A. On the smoothness of solutions of the third order nonlinear differential equation. .//Boundary value problems. May 2017. –P. 1-11. DOI 10.1186/s13661-017-0799-4

[19] Borodin P. A., Savchuk A. M., Sheipak I. A. Zadachi po funktsional'nomu analizu [Problems in Functional Analysis]. New ed. Moscow: MTsNMO, 2017. 336 p.

IRSTI 14.35.09

https://doi.org/10.51889/2959-5894.2026.95.3.011

T. S. Mamadaliev

O. Zhanibekov South Kazakhstan Pedagogical University, Shymkent, Kazakhstan
*e-mail: Sanzhar_sherlock@mail.ru

DIGITAL EDUCATIONAL RESOURCES IN INCLUSIVE MATHEMATICS TEACHING

Abstract

The study examines the potential of digital educational resources for mathematics teaching in inclusive education and the readiness of pre-service mathematics teachers to integrate these resources into their professional practice. The research was conducted at O. Zhanibekov South Kazakhstan Pedagogical University. The methodological design combines a questionnaire with practice-oriented mathematical tasks. GeoGebra and Desmos were incorporated into the practical component to assess participants' ability to visualize mathematical content, use digital tools purposefully, adapt mathematical tasks to inclusive learning requirements, and differentiate activities according to learners' educational needs. The practical tasks were evaluated using five criteria: mathematical accuracy, effective use of digital tools, visualization, inclusive adaptation, and task differentiation. The methodological framework therefore approaches digital competence not merely as technical proficiency but as the ability to integrate technology with mathematical content and inclusive pedagogical practices. The practical significance of the study lies in its potential contribution to mathematics teacher education by supporting the development of learning activities that integrate digital competence with the capacity to design accessible and differentiated mathematics instruction.

Keywords: inclusive education, mathematics teaching, digital educational resources, GeoGebra, Desmos, pre-service mathematics teachers, digital competence.

Т.С. Мамадалиев

Ө. Жәнібеков атындағы Оңтүстік Қазақстан педагогикалық университеті, Шымкент қ., Қазақстан

ИНКЛЮЗИВТІ МАТЕМАТИКАНЫ ОҚЫТУДАҒЫ ЦИФРЛЫҚ БІЛІМ БЕРУ РЕСУРСТАРЫ

Аңдатпа

Зерттеу инклюзивті білім беру жағдайында математиканы оқытуда цифрлық білім беру ресурстарын қолдану мүмкіндіктерін және болашақ математика мұғалімдерінің оларды кәсіби қызметте пайдалануға даярлығын зерттеуге бағытталған. Зерттеу Өзбекәлі Жәнібеков атындағы Оңтүстік Қазақстан педагогикалық университетінің базасында ұйымдастырылды. Зерттеу әдістемесі сауалнама мен практикалық математикалық тапсырмаларды кешенді қолдануға негізделді. Практикалық бөлімде GeoGebra және Desmos цифрлық орталары қолданылып, қатысушылардың математикалық мазмұнды визуализациялау, цифрлық құралдарды пайдалану, тапсырмаларды инклюзивті білім беру талаптарына бейімдеу және даралау қабілеттері бағаланды. Практикалық жұмыстар математикалық дұрыстық, цифрлық құралды пайдалану, визуализация, инклюзивті бейімдеу және тапсырмаларды даралау критерийлері бойынша талданды. Зерттеудің әдістемелік негізі цифрлық технологияларды техникалық меңгеруді ғана емес, оларды математикалық мазмұнмен және инклюзивті педагогикалық міндеттермен кіріктіре қолдануды қарастырады. Зерттеу нәтижелерінің практикалық маңызы болашақ математика мұғалімдерін даярлау бағдарламаларында цифрлық және инклюзивті құзыреттіліктерді өзара байланысты қалыптастыруға бағытталған практикалық тапсырмаларды жетілдіру мүмкіндігімен анықталады.

Түйін сөздер: инклюзивті білім беру, математиканы оқыту, цифрлық білім беру ресурстары, GeoGebra, Desmos, болашақ математика мұғалімдері, цифрлық құзыреттілік.

Т.С. Мамадалиев
Южно-Казахстанский педагогический университет имени О. Жанибекова, г.Шымкент, Казахстан

ЦИФРОВЫЕ ОБРАЗОВАТЕЛЬНЫЕ РЕСУРСЫ В ИНКЛЮЗИВНОМ ОБУЧЕНИИ МАТЕМАТИКЕ

Аннотация

Исследование направлено на изучение возможностей применения цифровых образовательных ресурсов при обучении математике в условиях инклюзивного образования и готовности будущих учителей математики к их использованию в профессиональной деятельности. Исследование организовано на базе Южно-Казахстанского педагогического университета имени О.Жанибекова. Методология исследования основана на комплексном применении анкетирования и практических математических заданий. В практической части использовались цифровые среды GeoGebra и Desmos, посредством которых оценивались способности участников визуализировать математическое содержание, использовать цифровые инструменты, адаптировать задания к требованиям инклюзивного образования и осуществлять их индивидуализацию. Практические работы анализировались по критериям математической корректности, использования цифрового инструмента, визуализации, инклюзивной адаптации и индивидуализации заданий. Методологический подход исследования предусматривает не только техническое освоение цифровых технологий, но и их интеграцию с математическим содержанием и задачами инклюзивной педагогики. Практическая значимость исследования определяется возможностью совершенствования подготовки будущих учителей математики посредством включения заданий, направленных на взаимосвязанное формирование цифровых и инклюзивных компетенций.

Ключевые слова: инклюзивное образование, обучение математике, цифровые образовательные ресурсы, GeoGebra, Desmos, будущие учителя математики, цифровая компетентность.

Introduction

The ongoing digital transformation of education requires not only the renewal of instructional content and teaching methods but also the development of accessible learning environments that accommodate learners' individual abilities and diverse educational needs. This requirement becomes particularly important in inclusive education, where meaningful inclusion extends beyond the physical presence of a learner in a mainstream classroom. It also involves ensuring access to learning materials, active participation in educational activities, and appropriate pedagogical support. Contemporary research on inclusive mathematics education therefore emphasizes accessibility and meaningful participation as essential conditions for equitable learning opportunities [1].

The implementation of inclusive approaches in mathematics education presents specific pedagogical challenges. Mathematics requires learners to understand abstract concepts, process symbolic information, identify logical relationships, and follow sequential procedures when solving problems. These characteristics may create additional barriers for students who experience difficulties in processing information, maintaining attention, understanding abstract representations, or completing multistep tasks. Consequently, the effectiveness of inclusive mathematics instruction depends not only on the mathematical content itself but also on how this content is represented, structured, differentiated, and adapted to learners' individual needs. Previous research indicates that effective inclusive mathematics education is closely associated with flexible instructional practices, appropriate teacher support, and the creation of accessible learning environments [2].

Digital educational resources provide considerable opportunities to address these challenges. They enable mathematical information to be represented through textual, graphical, visual, and interactive formats and allow teachers to modify the complexity of tasks, organize repeated practice, and provide timely feedback. Digital technologies can also facilitate differentiated instruction by allowing mathematical content to be presented in multiple ways and adjusted to different levels of learner readiness. From an inclusive perspective, these characteristics are particularly relevant because accessibility cannot be reduced to simplifying mathematical content; rather, it requires creating alternative pathways through which learners can engage with and understand mathematical ideas [3].

Dynamic mathematics environments have particular potential in this context. Tools such as GeoGebra and Desmos allow mathematical relationships to be represented dynamically and enable learners to observe changes in graphs, parameters, geometric objects, and algebraic expressions. Such environments can support the transition between symbolic and visual representations and make abstract mathematical relationships more observable. Research on technology-enhanced mathematics education has demonstrated the potential of dynamic mathematical environments to support conceptual understanding, visualization, and active engagement with mathematical content [4, 5]. At the same time, the educational value of these tools depends substantially on how teachers integrate them into instructional activities rather than on the availability of technology itself.

This issue places new demands on the professional preparation of future mathematics teachers. Digital competence in mathematics education involves more than the ability to operate a particular application or online platform. A future teacher should be able to select a digital resource according to a specific mathematical objective, design meaningful learning activities, visualize mathematical relationships, differentiate tasks, and adapt digital materials to learners with different educational needs. This understanding is consistent with approaches that integrate technological, pedagogical, and content knowledge in teacher education [6]. Recent studies involving pre-service mathematics teachers have similarly shown that effective adoption of GeoGebra is associated with pedagogical and technological knowledge, perceived usefulness, and the capacity to incorporate digital tools meaningfully into future teaching practice [7].

The inclusive dimension adds another layer to this professional competence. A mathematically correct digital task is not necessarily an accessible or inclusive task. For example, an accurately constructed function graph may still be difficult for a learner to interpret if instructions are overly complex, the activity contains too many simultaneous operations, or no differentiated support is provided. Future mathematics teachers therefore need experience in transforming standard mathematical tasks into accessible learning activities through step-by-step instructions, visual support, differentiated levels of complexity, and alternative forms of interaction with mathematical content. Studies of teacher preparation for inclusive mathematics education emphasize that beliefs, self-efficacy, and professional competence related to inclusion can be developed through purposeful teacher education rather than being regarded as automatically acquired general teaching skills [8].

Despite the growing body of research on digital technologies in mathematics education and the increasing attention paid to inclusive mathematics teaching, these two areas are still frequently examined separately. Studies of digital mathematics environments often focus on mathematical achievement, visualization, technology acceptance, or technological pedagogical content knowledge, whereas research on inclusive mathematics education primarily addresses accessibility, participation, differentiation, teacher beliefs, and support for diverse learners [2, 7–9]. Consequently, there remains a need to examine future mathematics teachers' digital competence together with their ability to adapt technology-supported mathematical tasks to inclusive educational requirements.

This study addresses this gap by considering digital and inclusive competence within a common practice-oriented framework. In addition to examining pre-service mathematics teachers' experiences and perceptions regarding digital educational resources, the study evaluates their performance in GeoGebra- and Desmos-based mathematical tasks. Particular attention is given to mathematical accuracy, the purposeful use of digital tools, visualization, inclusive adaptation, and task differentiation. This approach makes it possible to consider not only how future teachers evaluate their own readiness but also how this readiness is manifested when they are required to design and adapt mathematical learning activities.

Accordingly, the aim of the study is to examine pre-service mathematics teachers' readiness to use digital educational resources in inclusive mathematics teaching and to evaluate their ability to integrate mathematical content, digital tools, and inclusive instructional strategies in practice.

The study is guided by the following research question: *To what extent are pre-service mathematics teachers prepared to use digital educational resources for the visualization, adaptation, and differentiation of mathematical tasks in inclusive education?*

The research hypothesis assumes that pre-service mathematics teachers demonstrate stronger performance in the mathematical and visualization aspects of digital tasks than in their inclusive adaptation and differentiation, indicating a need for more systematic integration of digital and inclusive competencies within mathematics teacher education.

Research methodology

The study employed a quantitative descriptive research design to examine pre-service mathematics teachers' readiness to use digital educational resources in inclusive education. Data were collected through a questionnaire and a set of practice-oriented tasks involving digital mathematics tools. This combined approach made it possible to relate participants' self-reported perceptions of the potential of digital resources to their actual performance in representing mathematical content in digital environments and adapting mathematical tasks to the requirements of inclusive teaching.

Research Setting and Participants. The study was conducted at O. Zhanibekov South Kazakhstan Pedagogical University. The participants were students enrolled in educational programs for the preparation of future mathematics teachers. Participant selection took into account their engagement with mathematics-related courses, familiarity with digital educational resources, and preparation for future professional teaching practice.

Participation in the study was voluntary. The participants were informed about the purpose of the research and that the collected data would be used exclusively for scientific purposes. The confidentiality of personal information was maintained throughout the study, and the data were processed and reported in aggregated form.

Research Instruments. Empirical data were collected using two interrelated instruments: an author-developed questionnaire and a set of practical mathematical tasks. The author-developed questionnaire consisted of 20 items organized into five thematic sections: participants' general information; experience in using digital resources for mathematics teaching; perceptions of the didactic potential of digital tools; opportunities for using digital resources in inclusive mathematics education; and readiness to use these resources, including perceived barriers to their implementation. A five-point Likert scale was used to assess the main statements: 1 – strongly disagree; 2 – disagree; 3 – neither agree nor disagree; 4 – agree; 5 – strongly agree. In addition, the questionnaire examined the frequency with which participants used digital educational resources, the platforms they employed, and their overall self-reported readiness to integrate digital tools into inclusive mathematics teaching. To complement the questionnaire data with performance-based evidence, participants completed two practical tasks using the *GeoGebra and Desmos* digital mathematics environments.

Practical Mathematical Tasks. The first practical task focused on the visualization of a quadratic function. Participants were given the function

$$y=x^2-4x+3$$

and were asked to construct its graph in GeoGebra and determine the vertex of the parabola, its axis of symmetry, and its points of intersection with the coordinate axes. In the subsequent stage, the pre-service teachers were required to adapt the original mathematical task for a learner with special educational needs. The adaptation involved dividing the task into a clear sequence of steps, incorporating visual supports, providing accessible instructions, and using the features of the digital environment to support the learner's engagement with the task.

The second practical task examined the effect of a parameter on the graph of a quadratic function. Using Desmos, participants constructed the function

$$y=ax^2$$

and compared changes in the graph for a=0.5, a=1, a=2, and a=−1. They were asked to explain how changes in the parameter affected the orientation and shape of the parabola and to design an interactive mathematical activity for learners based on the resulting visualization. Participants were also required to adapt the activity for an inclusive classroom by developing at least two levels of difficulty.

The practical tasks were designed to assess not only the mathematical accuracy of participants' work but also their ability to use digital tools for didactic purposes and to adapt mathematical activities to learners' individual educational needs.

Assessment of Practical Tasks. The participants' work was evaluated according to five criteria: mathematical accuracy, use of digital tools, visualization of mathematical content, inclusive adaptation, and task differentiation. The assessment criteria and scoring descriptors are presented in Table 1.

Table 1. Assessment Criteria for the Practical Tasks

Criterion	*0 points*	*1 point*	*2 points*
Mathematical accuracy	*The solution contains significant errors*	*Partially correct*	*Fully correct*
Use of digital tools	*Not used*	*Digital features are partially utilized*	*Digital features are used effectively and purposefu*
Visualization	*No visualization is provi*	*Basic visualization is provided*	*Purposeful visualization clearly represents the mathematical content*
Inclusive adaptation	*No adaptation is provide*	*Individual elements are adapted*	*Systematically adapted t meet the learner's educational needs*
Task differentiation	*Not provided*	*Partially addressed*	*Provided at two or more levels of difficulty*

The maximum score for each criterion was 2 points. Thus, the total score for the practical tasks ranged from 0 to 10 points. For descriptive purposes, the results were classified into three levels: 0–3 points as low, 4-7 points as moderate, and 8-10 points as high. The study was conducted in several stages. At the first stage, scientific literature on inclusive mathematics education, digital educational resources, and the digital competence of pre-service mathematics teachers was reviewed. Based on this analysis, the content and structure of the research instruments were determined.

At the second stage, the participants completed the author-developed questionnaire. This stage was designed to identify their experience in using digital resources, perceptions of the potential of these resources for inclusive mathematics teaching, and self-reported level of readiness to integrate digital tools into their future professional practice. At the third stage, the participants completed practical mathematical tasks using *GeoGebra and Desmos.* In addition to visualizing mathematical objects in digital environments, they were required to adapt the proposed tasks in accordance with the principles and requirements of inclusive education.

At the fourth stage, the data obtained from the questionnaire and practical tasks were systematized and subjected to quantitative analysis. Participants' self-reported readiness scores were compared with their performance on the practical tasks. Descriptive statistical methods were used to analyze the research data. For categorical variables, absolute frequencies (f) and percentage distributions (P) were calculated using the following formula:

$$P = \left(\frac{f}{N}\right) \times 100\%$$

where f represents the number of participants who selected a particular response and N represents the total number of participants. For the Likert-scale items and scores obtained from the practical tasks, the arithmetic mean was calculated as follows:

$$\bar{x} = \frac{\sum_{i-1}^{N} x_i}{N}$$

where x_i represents the individual score of each participant and N represents the total number of participants.

These descriptive measures were used to characterize pre-service mathematics teachers' readiness to use digital educational resources in inclusive teaching, their perceptions of the didactic potential of digital tools, and their performance on the practical mathematical tasks.

Correlation analysis was employed to examine the relationship between participants' self-reported readiness and their practical performance. Participants' self-assessment scores regarding their readiness to use digital educational resources in inclusive mathematics teaching were compared with their total scores on the GeoGebra- and Desmos-based practical tasks. The relationship between these two variables was assessed using Pearson's correlation coefficient (r):

$$r = \frac{\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{N}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{N}(y_i - \bar{y})^2}}$$

where x_i represents a participant's self-reported readiness to use digital resources in inclusive mathematics teaching, y_i represents the participant's total score on the practical tasks, $\bar{x}$ and $\bar{y}$ represent the corresponding mean values, and N represents the total number of participants.

A positive correlation coefficient indicated that higher levels of self-reported readiness tended to be associated with higher practical performance scores, whereas a negative coefficient indicated an inverse relationship between the two variables. Statistical significance was evaluated at the $p<0.05$ level.

In addition, performance on the practical tasks was analyzed separately according to five criteria: mathematical accuracy, use of digital tools, visualization, inclusive adaptation, and task differentiation. Mean scores were calculated for each criterion, and the results obtained from the GeoGebra- and Desmos-based tasks were compared. This analysis provided a more differentiated assessment of participants' ability not only to work accurately with mathematical content but also to visualize it in a digital environment and adapt it to the requirements of inclusive education.

The quantitative findings were presented in tables and graphical form. The figures illustrated the percentage distributions obtained from the questionnaire, mean scores for the practical tasks, and the distribution of pre-service mathematics teachers across different levels of readiness. This approach enabled the participants' self-reported readiness identified through the questionnaire to be considered alongside their performance-based readiness demonstrated through the practical mathematical tasks.

Results of the study

The findings provided a multidimensional perspective on pre-service mathematics teachers' readiness to use digital educational resources in inclusive education. The analysis focused on several interrelated indicators, including participants' experience with digital resources, their performance on practical tasks completed in *GeoGebra and Desmos*, their levels of practical readiness, and the relationship between self-reported readiness and actual task performance. First, participants' experience with different digital educational resources used in mathematics teaching was analyzed. Since respondents were allowed to select more than one response option, the cumulative percentage

across the categories could exceed 100%. Participants' reported use of digital educational resources in mathematics teaching is presented in Table 2.

Table 2. Pre-Service Mathematics Teachers' Use of Digital Educational Resources

Digital Educational Resource	*n*	*%*
GeoGebra	*62*	*77.5*
Desmos	*48*	*60.0*
LearningApps	*44*	*55.0*
Quizizz	*39*	*48.8*
Kahoot!	*35*	*43.8*
E-textbooks	*53*	*66.3*
Video materials	*57*	*71.3*

According to the results presented in Table 2**,** GeoGebra was the most widely used mathematics-specific digital resource among the participants, with 62 respondents (77.5%) reporting its use. Video materials were used by 71.3% of the participants, while 66.3% reported using e-textbooks. Desmos was used by 60.0% of the respondents. Learning Apps was reported by 55.0% of the participants, whereas Quizizz and Kahoot! were used by 48.8% and 43.8%, respectively. These findings indicate that the participants were familiar not only with general digital educational resources but also with specialized platforms designed to support the visualization of mathematical content. The distribution of these indicators is presented graphically in Figure 1**.**

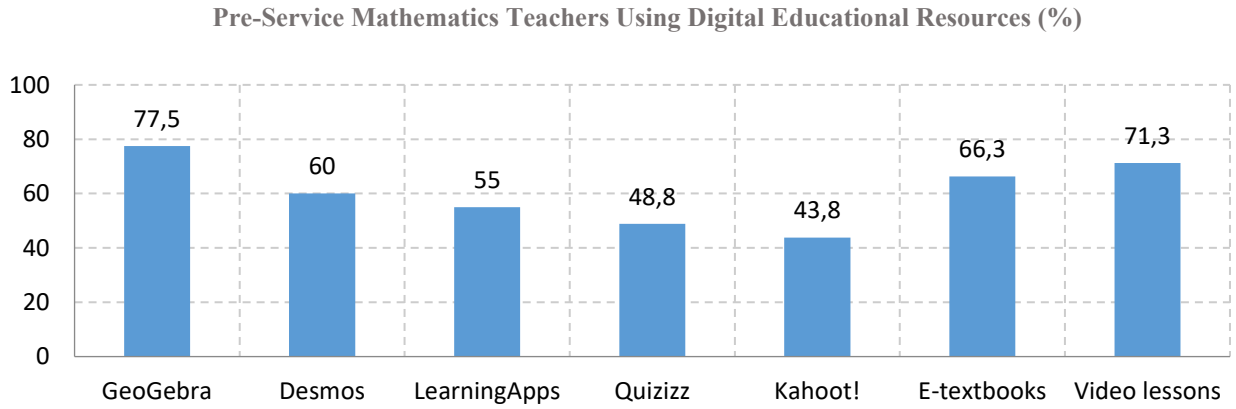


Figure 1. Percentage of Pre-Service Mathematics Teachers Using Digital Educational Resources

Figure 1 provides a comparative representation of the frequency with which different digital tools were used by the participants. The higher use of GeoGebra compared with the other platforms indicates that digital environments designed to visualize mathematical objects and functions occupy an important place in the participants' experience with technology-enhanced mathematics learning. During the practical stage of the study, participants' ability to use digital mathematics environments to complete specific mathematical tasks was assessed. The tasks performed in GeoGebra and Desmos were evaluated according to five criteria: mathematical accuracy, use of digital tools, visualization, inclusive adaptation, and task differentiation. The maximum score for each criterion was 2 points.

The mean scores for the GeoGebra and Desmos tasks across the five assessment criteria are presented in Table 3.

According to the data presented in Table 3**,** mathematical accuracy showed the highest mean score in both digital environments. The mean score for this criterion was M=1.72 for the GeoGebra task and M=1.66 for the Desmos task. Visualization also demonstrated relatively high scores, with mean values of M=1.68 and M=1.63, respectively. For the use of digital tools, the mean score was M=1.61 in GeoGebra and M=1.58 in Desmos.

Table 3. Mean Performance Scores for Practical Tasks Completed in GeoGebra and Desmos

Assessment Criterion	*GeoGebra, M*	*Desmos, M*
Mathematical accuracy	*1.72*	*1.66*
Use of digital tools	*1.61*	*1.58*
Visualization	*1.68*	*1.63*
Inclusive adaptation	*1.42*	*1.38*
Task differentiation	*1.35*	*1.31*

In contrast, inclusive adaptation and task differentiation yielded lower mean scores than the other assessment criteria. The mean score for inclusive adaptation was M=1.42 for the GeoGebra task and M=1.38 for the Desmos task. Similarly, the mean scores for task differentiation were M=1.35 and M=1.31, respectively. In contrast, inclusive adaptation and task differentiation yielded lower mean scores than the other assessment criteria. The mean score for inclusive adaptation was M = 1.42 for the GeoGebra task and M = 1.38 for the Desmos task. Similarly, the mean scores for task differentiation were M = 1.35 and M = 1.31, respectively. A comparison of the mean scores across the five assessment criteria for the GeoGebra and Desmos tasks is presented in Figure 2.

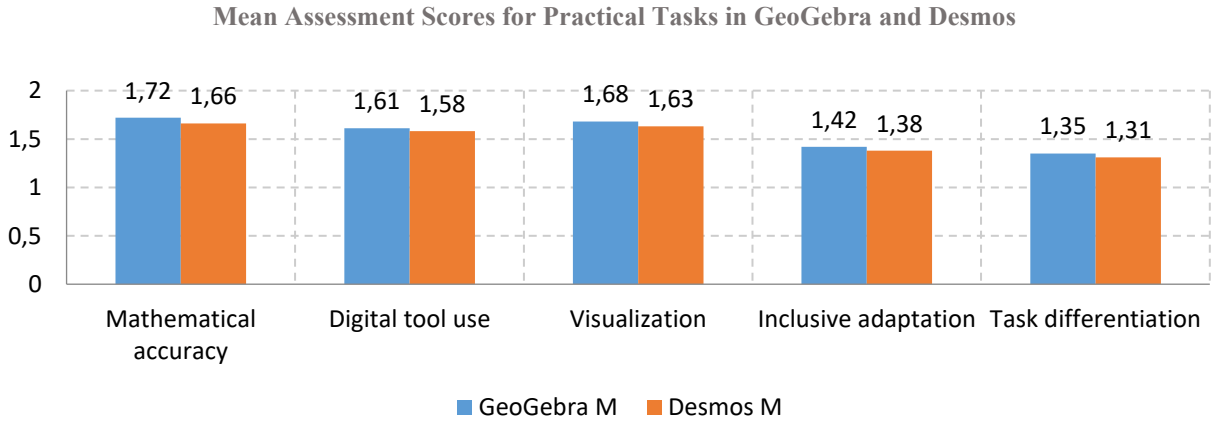


Figure 2. Mean Scores for the Assessment Criteria of Practical Tasks Completed in GeoGebra and Desmos

The results shown in Figure 2 indicate that mathematical accuracy and visualization received relatively higher mean scores in both digital environments, whereas inclusive adaptation and task differentiation received lower scores. Based on the total scores obtained from the practical tasks, participants' readiness to use digital resources in inclusive mathematics teaching was classified into three levels: low (0–3 points), moderate (4–7 points), and high (8–10 points). The distribution of participants across these readiness levels is presented in Table 4.

Table 4. Practical Readiness Levels of Pre-Service Mathematics Teachers

Readiness Level	*Score Range*	*n*	*%*
Low	*0–3*	*Excel result*	*Excel result*
Moderate	*4–7*	*Excel result*	*Excel result*
High	*8–10*	*Excel result*	*Excel result*
Total		*80*	*100.0*

The distribution of participants across the three readiness levels provides an overall indication of their practical competence in using digital resources for inclusive mathematics teaching. The results reflect differences in participants' ability to combine mathematical knowledge with the effective use

of digital tools, visualization, inclusive adaptation, and task differentiation. A graphical representation of the distribution across the three readiness levels is provided in Figure 3.

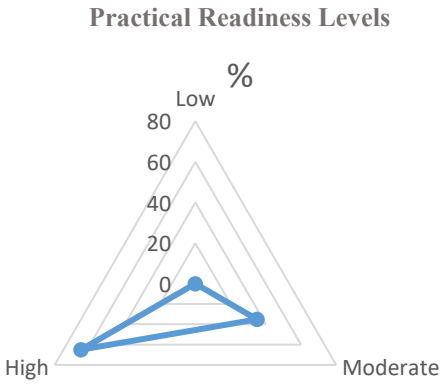


Figure 3. Practical Readiness Levels of Pre-Service Mathematics Teachers for Using Digital Resources in Inclusive Mathematics Teaching

At the next stage of the study, the relationship between participants' self-reported readiness to use digital resources in inclusive mathematics teaching and their total scores on the practical tasks was examined. Participants rated their perceived readiness on a 0–10-point scale, while practical readiness was determined based on their performance on the GeoGebra- and Desmos-based tasks.

Overall, the quantitative findings revealed variation in pre-service mathematics teachers' experience with digital educational resources. Among the resources examined, GeoGebra, video materials, and e-textbooks showed relatively high levels of use. In the assessment of practical tasks, comparatively higher scores were observed for mathematical accuracy and visualization, whereas inclusive adaptation and task differentiation received lower scores. The relationship between participants' self-reported readiness and their practical performance was further examined through correlation analysis.

A comparison of the mean scores across the five assessment criteria for the GeoGebra and Desmos tasks is presented in Figure 4.

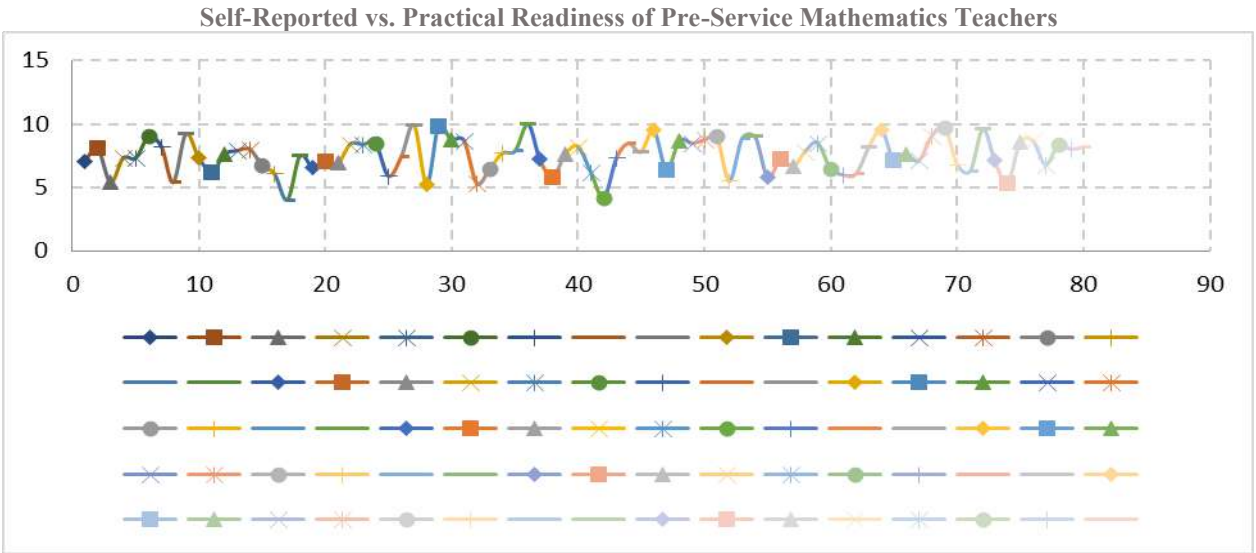


Figure 4. Relationship Between Self-Reported and Practical Readiness of Pre-Service Mathematics Teachers

Each point in the figure represents two indicators for an individual participant. The horizontal axis shows the participant's self-reported readiness to use digital resources in inclusive mathematics teaching, while the vertical axis represents the total score obtained from the practical tasks. The specific correlation coefficient was determined using an appropriate statistical test after examining the distributional characteristics of the data. Overall, the quantitative findings revealed variation in pre-service mathematics teachers' experience with digital educational resources. Among the resources examined, GeoGebra, video materials, and e-textbooks showed relatively high levels of use. In the assessment of practical tasks, comparatively higher scores were observed for mathematical accuracy and visualization, whereas inclusive adaptation and task differentiation received lower scores. The relationship between participants' self-reported readiness and their practical performance was further examined through correlation analysis.

Discussion

The findings indicate that although pre-service mathematics teachers had developed a relatively broad experience of using digital educational resources, their readiness to adapt these resources to the requirements of inclusive teaching was not equally developed. Participants' familiarity with resources such as GeoGebra, Desmos, e-textbooks, and video materials suggests that digital tools are becoming increasingly integrated into their professional preparation. However, a noticeable difference emerged between the technical use of digital tools and their pedagogically appropriate adaptation to learners with diverse educational needs. This distinction highlights the need to conceptualize digital competence not merely as technological proficiency, but as the ability to integrate mathematical content, pedagogical strategies, and individual learner needs.

GeoGebra was identified as one of the most frequently used mathematics-specific digital resources among the participants. The practical tasks also produced comparatively high scores for mathematical accuracy and visualization. This finding may be associated with GeoGebra's capacity to represent abstract mathematical objects dynamically, demonstrate changes in parameters, and facilitate the visual exploration of mathematical relationships. Previous research on pre-service mathematics teachers' acceptance of GeoGebra has similarly demonstrated that perceived usefulness and ease of use are associated with intentions to employ the platform in future professional practice [10].

This finding is also consistent with studies examining the relationship between technological, pedagogical, and content knowledge. Yildiz and Arpaci demonstrated that pre-service mathematics teachers' intentions to continue using GeoGebra are influenced by their technological pedagogical content knowledge (TPACK), performance expectancy, and social factors [11]. From this perspective, the ability to use a digital tool should not be considered separately from mathematical content. In addition to mastering the technical functions of a digital platform, future teachers need to understand the specific didactic purpose that the technology serves when explaining a particular mathematical concept.

The relatively high visualization scores obtained in the GeoGebra- and Desmos-based practical tasks are particularly noteworthy. Representing mathematical functions and their graphs in dynamic environments enables learners to observe relationships between parameters visually. Research conducted in the context of universities in Kazakhstan has also highlighted the potential of GeoGebra for accurately constructing mathematical objects, dynamically modifying them, and visually exploring mathematical properties [12]. Digital visualization in inclusive mathematics education can therefore be viewed not simply as a means of illustration but as a pedagogical strategy for improving access to abstract mathematical content.

At the same time, inclusive adaptation and task differentiation received lower scores than mathematical accuracy and visualization. This represents one of the key findings of the present study. Although a future teacher may be able to construct a mathematically correct graph in a digital environment, adapting the same material for a learner with special educational needs, differentiating it across several levels of complexity, or selecting appropriate forms of additional support may present greater challenges. Meaney et al., in examining the potential of digital tools to support diverse

learners, similarly identified considerable variation in how pre-service and practicing teachers evaluated the suitability of different digital tools for particular groups of learners [13].

These findings emphasize the importance of purposeful professional preparation for inclusive mathematics education. Scherer and Bertram found that pre-service teachers' beliefs and self-efficacy concerning inclusive mathematics teaching can change over the course of teacher education programs [14]. Thus, competence in inclusive mathematics teaching should not be regarded as a general professional skill that develops automatically. Instead, it should be systematically fostered through targeted learning experiences and practice-oriented tasks during university-based teacher education.

In this respect, the design of the practical tasks used in the present study is particularly relevant. Participants were required not merely to use ready-made digital materials but to construct mathematical tasks in digital environments, visualize mathematical relationships, and redesign the tasks in accordance with inclusive education requirements. Previous studies in which pre-service mathematics teachers designed mathematical modelling activities by integrating different digital tools have also emphasized the importance of connecting technology use with the design of subject-specific learning tasks in professional teacher preparation [15].

The findings therefore suggest that digital and inclusive competencies should not be developed as separate components of mathematics teacher education but rather within an integrated subject-specific and methodological framework. For example, instead of limiting instruction to the technical functions of GeoGebra or Desmos, teacher education programs could require students to develop standard, visualized, step-by-step, and differentiated versions of the same mathematical task. Research on the integration of interactive mathematical resources into teacher education similarly emphasizes the importance of preparing teachers not only as users of digital resources but also as designers capable of evaluating, adapting, and transforming these resources into meaningful learning materials [16].

The practical significance of the study lies in its implications for incorporating learning activities that connect digital mathematics environments with inclusive pedagogy into pre-service mathematics teacher education. Such activities may include visualizing mathematical content, managing the cognitive demands of tasks, developing step-by-step instructions, providing additional visual support, and designing tasks at different levels of complexity while maintaining mathematical accuracy. Studies involving targeted GeoGebra training for pre-service mathematics teachers have reported improvements in technological and pedagogical knowledge, further supporting the potential value of structured practical preparation at the university level [17].

Future studies could extend the sample to include pre-service mathematics teachers from several teacher education institutions and compare participants across different years of study. An experimental design comparing participants' performance before and after a targeted training program on the inclusive use of digital mathematics resources could provide stronger evidence of changes in professional competence. Future research could also develop and evaluate inclusive digital tasks across different areas of mathematics, including geometry, algebra, probability, and statistics.

Conclusion

The study demonstrates that the use of digital educational resources in inclusive mathematics teaching requires the professional preparation of pre-service mathematics teachers to integrate subject-specific, digital, and inclusive competencies. The research examined participants' experience with digital resources, the potential of these tools for visualizing mathematical content, and their ability to adapt practical tasks to inclusive education requirements. Performance on the GeoGebra- and Desmos-based practical tasks showed comparatively strong results in mathematical accuracy and visualization. These findings indicate that participants were able to use digital mathematics environments to construct function graphs, represent relationships between mathematical objects, and make abstract concepts more visually accessible.

At the same time, the comparatively lower scores for inclusive adaptation and task differentiation demonstrate that technical proficiency with digital tools alone is insufficient for their effective use in inclusive education. Future mathematics teachers need not only to understand the functional capabilities of digital resources but also to develop the ability to modify mathematical tasks according to learners' individual educational needs, structure activities into manageable stages, provide additional visual support, and differentiate levels of complexity.

The findings support the systematic integration of digital and inclusive competencies within pre-service mathematics teacher education. Practice-oriented activities based on GeoGebra, Desmos, and other digital mathematics environments could therefore be incorporated into teacher education programs to provide future teachers with opportunities to design, evaluate, and adapt mathematical tasks for diverse learners.

Overall, the initial research hypothesis was supported: digital educational resources provide important opportunities for mathematical visualization and task differentiation; however, their effective application in inclusive education depends on the future teacher's ability to adapt digital tools pedagogically to the specific educational needs of learners.

References

[1] Y. Abtahi and N. Planas, "Mathematics teaching and teacher education against marginalisation, or towards equity, diversity and inclusion," ZDM – Mathematics Education, vol. 56, pp. 307–318, 2024, doi: 10.1007/s11858-024-01602-x.

[2] L. Healy, E. Nardi, and I. Biza, "Interdependency, alternative forms of mathematical agency and joy as challenges to ableist narratives about the learning and teaching of mathematics," ZDM – Mathematics Education, vol. 56, no. 3, pp. 379–391, 2024, doi: 10.1007/s11858-024-01565-z.

[3] T. Meaney, H. L. Huru, and M. Kvivesen, "Preservice and inservice teachers' views on digital tools for diverse learners in mathematics education," NOMAD: Nordic Studies in Mathematics Education, vol. 28, no. 3–4, 2023, doi: 10.7146/nomad.v28i3-4.149192.

[4] E. K. Lotey, F. O. Boateng, Y. D. Arthur, B. Adu-Obeng, and J. F. Gordon, "Exploring GeoGebra acceptance: an implication for pre-service mathematics teachers' professional development," Cogent Education, vol. 12, no. 1, Art. no. 2514980, 2025, doi: 10.1080/2331186X.2025.2514980.

[5] E. Yildiz and I. Arpaci, "Understanding pre-service mathematics teachers' intentions to use GeoGebra: The role of technological pedagogical content knowledge," Education and Information Technologies, vol. 29, pp. 18817–18838, 2024, doi: 10.1007/s10639-024-12614-1.

[6] H. Seifert and A. Lindmeier, "Developing a performance-based assessment to measure pre-service secondary teachers' digital competence to use digital mathematics tools," Journal für Mathematik-Didaktik, vol. 45, Art. no. 25, 2024, doi: 10.1007/s13138-024-00251-7.

[7] M. Y. Koyunkaya and A. T. Dede, "Using different digital tools in designing and solving mathematical modelling problems," Education and Information Technologies, vol. 29, pp. 19035–19065, 2024, doi: 10.1007/s10639-024-12577-3.

[8] P. Scherer and J. Bertram, "Professionalisation for inclusive mathematics–teacher education programs and changes in pre-service teachers' beliefs and self-efficacy," ZDM – Mathematics Education, vol. 56, pp. 447–459, 2024, doi: 10.1007/s11858-024-01580-0.

[9] R. Vithal, K. Brodie, and R. Subbaye, "Equity in mathematics education," ZDM – Mathematics Education, vol. 56, pp. 153–164, 2024, doi: 10.1007/s11858-023-01504-4.

[10] E. K. Lotey, F. O. Boateng, Y. D. Arthur, B. Adu-Obeng, and J. F. Gordon, "Exploring GeoGebra acceptance: an implication for pre-service mathematics teachers' professional development," Cogent Education, vol. 12, no. 1, Art. no. 2514980, 2025, doi: 10.1080/2331186X.2025.2514980.

[11] E. Yildiz and I. Arpaci, "Understanding pre-service mathematics teachers' intentions to use GeoGebra: The role of technological pedagogical content knowledge," Education and Information Technologies, vol. 29, pp. 18817–18838, 2024, doi: 10.1007/s10639-024-12614-1.

[12] E. Saparbayeva, M. Abdualiyeva, Y. Torebek, N. Madiyarov, and A. Tursynbayev, "Leveraging digital tools to advance mathematics competencies among construction students," Cogent Education, vol. 11, no. 1, Art. no. 2319436, 2024, doi: 10.1080/2331186X.2024.2319436.

[13] T. Meaney, H. L. Huru, and M. Kvivesen, "Preservice and inservice teachers' views on digital tools for diverse learners in mathematics education," NOMAD: Nordic Studies in Mathematics Education, vol. 28, no. 3–4, 2023, doi: 10.7146/nomad.v28i3-4.149192.

[14] P. Scherer and J. Bertram, "Professionalisation for inclusive mathematics–teacher education programs and changes in pre-service teachers' beliefs and self-efficacy," ZDM – Mathematics Education, vol. 56, pp. 447–459, 2024, doi: 10.1007/s11858-024-01580-0.

[15] M. Y. Koyunkaya and A. T. Dede, "Using different digital tools in designing and solving mathematical modelling problems," Education and Information Technologies, vol. 29, pp. 19035–19065, 2024, doi: 10.1007/s10639-024-12577-3.

[16] F. Dilling, I. Witzke, K. Hörnberger, et al., "Co-designing teaching with digital technologies: a case study on mixed pre-service and in-service mathematics teacher design teams," ZDM – Mathematics Education, vol. 56, pp. 667–680, 2024, doi: 10.1007/s11858-024-01623-6.

[17] J. Quarder, G. Greefrath, S. Gerber, et al., "Pedagogical content knowledge for simulations and mathematical modelling with digital tools: a quasi-experimental study with pre-service mathematics teachers," ZDM – Mathematics Education, vol. 57, pp. 395–409, 2025, doi: 10.1007/s11858-025-01673-4.

ИНФОРМАТИКА

COMPUTER SCIENCE

IRSTI 28.23.37; 44.41.35

https://doi.org/10.51889/2959-5894.2026.95.3.012

A. Abdildayeva[1], G. Taganova[2*], A. Tazhibeyev[3], A. Kural[3], S. Askarov[3]

[1]International Engineering and Technology University, Almaty, Kazakhstan
[2]Astana International University, Astana, Kazakhstan
[3]UMAY R&D LLP, Astana, Kazakhstan
[*]*e-mail: guldana.kileuzhanova@gmail.com*

LEAKAGE-CONTROLLED RESIDUAL-CALIBRATED HYBRID MODEL FOR ONE-HOUR-AHEAD PHOTOVOLTAIC GENERATION FORECASTING

Abstract

This experimental research study evaluates a leakage-controlled residual-calibrated hybrid model for sequential one-hour-ahead photovoltaic generation forecasting during active station-hours. A source audit of operational records from three solar power plants identified 28,296 raw hourly rows but only 11,736 unique station-hour keys; 16,560 redundant source rows were collapsed only when known plan and actual values were identical, and no conflicting known plan/actual values were found. After restricting supervised modelling to unique station-hours with positive planned generation and known actual output, 4,848 observations remained and were chronologically divided into 3,038 training, 819 calibration, and 991 held-out test observations. The proposed Residual-Calibrated Neural Hybrid (RCNH) combines a two-hidden-layer Tabular MLP with a HistGradientBoosting residual calibrator trained exclusively on the calibration period. On the held-out test period, RCNH achieved MAE = 2,591.80 kWh, RMSE = 3,906.53 kWh, and R^2 = 0.908609; 95% three-day block-bootstrap intervals were 2,372.75-2,961.81 kWh, 3,537.16-4,505.29 kWh, and 0.848489-0.932518, respectively. Against the train-only Tabular MLP, RCNH reduced MAE by 11.36% (95% block-bootstrap interval 6.98-14.95%; Holm-adjusted two-sided p = 0.003998). However, a matched-label-budget diagnostic favored an MLP refitted on train plus calibration data (MAE = 2,464.63 kWh), and rolling-origin validation showed substantially stronger temporal stability for tree ensembles, with ExtraTrees attaining mean RMSE = 3,064.51 kWh versus 6,106.74 kWh for RCNH. These results show that residual calibration is competitive and beneficial in the final held-out regime but is not uniformly dominant; its effectiveness depends on calibration-window representativeness and temporal regime. The contribution is therefore an experimentally validated, leakage-controlled forecasting protocol and a quantified assessment of where residual calibration helps and where its robustness is limited.

Keywords: photovoltaic generation; one-hour-ahead forecasting; data leakage; residual calibration; chronological validation; block bootstrap; machine learning; experimental evaluation

А.А. Абдилдаева[1], Г. Таганова[2], А.А. Тажибаев[3], А.Ж. Құрал[3], С.С. Аскаров[3]
[1]Халықаралық инженерлік-технологиялық университеті, Алматы қ., Қазақстан
[2]Астана халықаралық университеті, Астана қ., Қазақстан
[3]UMAY R&D ЖШС, Астана қ., Қазақстан

ФОТОЭЛЕКТРЛІК ГЕНЕРАЦИЯНЫ БІР САҒАТ БҰРЫН БОЛЖАУҒА АРНАЛҒАН ДЕРЕКТЕРДІҢ АҒЫП КЕТУІН БАҚЫЛАЙТЫН ЖӘНЕ ҚАЛДЫҚТАРДЫ КАЛИБРЛЕЙТІН ГИБРИДТІ МОДЕЛЬ

Аңдатпа

Бұл ғылыми-эксперименттік зерттеуде белсенді сағаттардағы фотоэлектрлік генерацияны бір сағат алдын ала болжауға арналған деректердің ағып кетуін бақылайтын қалдықты-калибрленген гибридті модель бағаланды. Үш күн электр станциясының 28 296 бастапқы сағаттық жазбасы аудиттен өткізіліп, 11 736 бірегей «станция-сағат» кілті анықталды; бірдей мәндері бар 16 560 қайталама жазба біріктірілді. Жоспарлы генерациясы оң және нақты генерациясы белгілі 4 848 бірегей белсенді бақылау 3 038 оқыту, 819 калибрлеу және 991 тәуелсіз тест бақылауына хронологиялық түрде бөлінді. RCNH моделі екі жасырын қабатты Tabular MLP мен калибрлеу кезеңінің қалдықтарында оқытылған HistGradientBoosting калибраторын біріктіреді. Тест кезеңінде RCNH MAE = 2 591,80 кВт·сағ, RMSE = 3 906,53 кВт·сағ және R^2 = 0,908609 нәтижелерін көрсетті. Оқыту кезеңімен ғана үйретілген MLP-мен салыстырғанда MAE 11,36%-ға төмендеді; үш күндік block-bootstrap үшін 95% аралығы 6,98-14,95%, Holm түзетуінен кейінгі екіжақты p = 0,003998. Алайда train+calibration деректерімен қайта оқытылған MLP MAE = 2 464,63 кВт·сағ нәтижесін көрсетті, ал rolling-origin тексеруде ағаштық ансамбльдер анағұрлым тұрақты болды. Сондықтан RCNH бәсекеге қабілетті және белгілі бір тест режимінде тиімді, бірақ оның артықшылығы калибрлеу терезесінің репрезентативтілігі мен уақыттық режимге тәуелді.

Түйін сөздер: фотоэлектрлік генерация; бір сағат алдын ала болжау; деректердің ағып кетуі; қалдықтарды калибрлеу; хронологиялық валидация; block bootstrap; машиналық оқыту; эксперименттік бағалау

А.А. Абдилдаева[1], Г. Таганова[2], А.А. Тажибаев[3], А.Ж. Құрал[3], С.С. Аскаров[3]
[1]Международный инженерно-технологический университет, г.Алматы, Казахстан
[2]Международный университет Астана, г.Астана, Казахстан
[3]ТОО UMAY R&D, г.Астана, Казахстан

ГИБРИДНАЯ МОДЕЛЬ С КОНТРОЛЕМ УТЕЧКИ ДАННЫХ И КАЛИБРОВКОЙ ОСТАТКОВ ДЛЯ ПРОГНОЗИРОВАНИЯ ФОТОЭЛЕКТРИЧЕСКОЙ ГЕНЕРАЦИИ НА ОДИН ЧАС ВПЕРЁД

Аннотация

В данной научно-экспериментальной работе выполнена оценка гибридной модели с контролем утечки данных и калибровкой остатков для последовательного прогнозирования фотоэлектрической генерации на один час вперед в активные часы. Аудит эксплуатационных данных трех солнечных электростанций показал, что 28 296 исходных почасовых строк соответствуют 11 736 уникальным наблюдениям «станция-час»; 16 560 повторных строк с идентичными известными значениями были консолидированы. После ограничения области моделирования наблюдениями с положительной плановой и известной фактической генерацией получено 4 848 уникальных активных наблюдений, хронологически разделенных на 3 038 обучающих, 819 калибровочных и 991 независимое тестовое наблюдение. Модель RCNH объединяет двухслойный Tabular MLP и калибратор HistGradientBoosting, обучаемый только на остатках калибровочного периода. На тестовом периоде получены MAE = 2 591,80 кВт·ч, RMSE = 3 906,53 кВт·ч и R^2 = 0,908609. Относительно MLP, обученного только на train-периоде, MAE снижена на 11,36%; 95%-й интервал трехдневного блочного bootstrap составляет 6,98-14,95%, скорректированное по Holm двустороннее p = 0,003998. Вместе с тем диагностический MLP, переобученный на train+calibration, достиг MAE = 2 464,63 кВт·ч, а rolling-origin валидация показала более высокую временную устойчивость древесных ансамблей. Следовательно, RCNH является конкурентоспособной и эффективной в конкретном отложенном режиме, но не демонстрирует универсального превосходства; ее результат зависит от репрезентативности калибровочного окна и временного режима.

Ключевые слова: фотоэлектрическая генерация; прогнозирование на один час вперед; утечка данных; калибровка остатков; хронологическая валидация; блочный bootstrap; машинное обучение; экспериментальная оценка

Introduction

Main provisions. The experimental unit is the unique station-hour, with duplicate source exports consolidated before feature engineering. The study evaluates leakage-controlled sequential one-hour-ahead photovoltaic forecasting during active hours using only information available at the forecast origin. On the held-out period, the Residual-Calibrated Neural Hybrid (RCNH) improved the train-only MLP, but matched-label-budget and rolling-origin analyses showed that this advantage was not universal. Overall, model ranking depended on calibration-window representativeness, information budget, and temporal regime. Operational photovoltaic forecasting supports dispatch, deviation monitoring, reserve planning, and integration of variable renewable generation. The task is complicated by non-stationarity across hours, seasons, plant scales, and operating regimes [1-7]. Forecasting modules can also contribute to broader digital-twin and monitoring architectures for energy systems and photovoltaic installations [8,9]. In industrial data, repeated exports and variables whose availability at prediction time is uncertain can introduce data leakage and inflate apparent predictive accuracy [10]. This study addresses two risks that can inflate apparent forecasting accuracy: data leakage from current or future information and the treatment of repeated source rows as independent observations. Accordingly, modelling is restricted to unique active station-hours and historical predictors aligned to exact earlier timestamps.

The research question is whether residual calibration can improve a leakage-controlled neural forecast on a genuinely later period and whether any gain remains credible under time-aware uncertainty analysis and temporal robustness checks.

The objective is to develop and experimentally evaluate RCNH for one-hour-ahead photovoltaic generation forecasting. The study audits the operational data layer, constructs prediction-time admissible features, compares RCNH with operational, linear, tree-based, boosting, and neural baselines, and evaluates uncertainty, matched information budgets, rolling-origin robustness, and ablation effects. The working hypothesis is that residual calibration can improve a primary MLP when the calibration period is representative of the subsequent regime, while strong tabular models may be more stable under temporal shift. Prior photovoltaic forecasting studies cover physical, statistical, machine-learning, and hybrid approaches [3-7]. For temporally ordered forecasting problems, validation must respect chronology and dependence rather than rely uncritically on random resampling [11,12]. For medium-sized tabular datasets, linear models, Random Forest, ExtraTrees, gradient boosting, and XGBoost remain strong benchmarks [13-17]. The gap addressed here is methodological: duplicate source rows, undefined information availability, calibration/test contamination, and reliance on a single split can lead to optimistic conclusions. The proposed protocol therefore combines station-hour auditing, strict feature availability, a dedicated calibration period, an untouched test period, and temporal robustness diagnostics.

Research methodology

Data audit and experimental domain

The operational archive contains hourly plan-fact generation and several daily or event-based tables. Only variables with defensible availability at the forecast origin were retained. Same-day finalized KPIs, curtailment aggregates, maintenance and defect records were excluded when an hourly as-of join or availability timestamp could not be verified. A source audit reduced 28,296 raw hourly rows to 11,736 unique station-hour keys by consolidating only identical repeated values; no conflicting known plan/actual values were found. Among 9,792 unique station-hours with known actual generation, 4,848 observations with positive planned generation formed the active supervised domain. The resulting chronological partition is summarized in Table 1. The forecast is defined at the local interval start. The previous-hour final measurement is assumed available after that interval ends,

and the plan is treated as issued before the forecast interval; the latter remains an explicit operational assumption because plan issue timestamps are absent from the archive.

Table 1. Audited active-hour modelling domain and chronological split

Indicator	*Value*
Raw hourly rows	*28,296*
Raw rows with known actual	*22,320*
Unique station-hours	*11,736*
Redundant source rows collapsed	*16,560*
Unique station-hours with known actual	*9,792*
Unique active station-hours	*4,848*
Solar power plants	*3*
Unique active dates	*136*
Train	*3,038 rows / 81 dates (2022-04-01 to 2023-01-07)*
Calibration	*819 rows / 27 dates (2023-01-08 to 2023-02-16)*
Held-out test	*991 rows / 28 dates (2023-02-17 to 2023-03-23)*

Forecasting formulation and leakage-control strategy

Let s denote a solar power plant and t the forecasted hourly interval. The target is measured active generation and the same-interval plan is an admissible predictor. Supervised modelling is restricted to unique station-hours with positive planned generation and known actual output. The final feature set includes planned generation and transforms, calendar/cyclic descriptors, station identity, actual-generation and plan lags at 1, 2, 3, 24, 48, and 72 hours, the one-hour historical plan-fact residual, and rolling statistics over 3, 6, 12, and 24 hours. Historical values are aligned to exact timestamps before active-hour filtering; preprocessing is fitted on training data only. Leakage controls prohibit the current actual value, current plan-fact deviation or target ratio, future targets, same-day finalized aggregates, and event variables without verifiable availability timestamps. Pipeline tests additionally confirm that changing the current target or future actual values does not alter earlier feature vectors. The full leakage-controlled workflow is shown in Figure 1.

Residual-Calibrated Neural Hybrid and comparator models

RCNH is a two-stage model. The primary branch is a Tabular MLP with 128 and 64 hidden units, ReLU activation and Adam optimization; its epoch count is selected only on a chronological inner-training tail. The residual calibrator is a HistGradientBoostingRegressor trained exclusively on calibration-period residuals and admissible prediction context.

The final prediction adds the estimated residual correction to the fixed MLP forecast and constrains generation to non-negative values. The held-out test targets are never used for preprocessing, base-model fitting, calibration, hyperparameter selection, or model selection. The two-stage RCNH architecture is illustrated in Figure 2.

Comparators are the plan schedule, one-hour persistence, Ridge, RandomForest, ExtraTrees, HistGradientBoosting, XGBoost, and the standalone Tabular MLP. Hyperparameters were fixed before held-out evaluation; the main experiment uses seed 42, with seeds 43 and 44 used only as sensitivity checks.

Experimental and statistical evaluation protocol

The primary experiment uses a chronological split: 3,038 training observations (through 7 January 2023), 819 calibration observations (8 January-16 February 2023), and 991 held-out test observations (17 February-23 March 2023). Performance is assessed with MAE, RMSE, R^2, WAPE, and sMAPE, following established forecast-accuracy principles [18]. Although rank-based paired testing such as the Wilcoxon procedure is a classical nonparametric option for paired comparisons [19], the temporal dependence in this study motivated block-based inference.

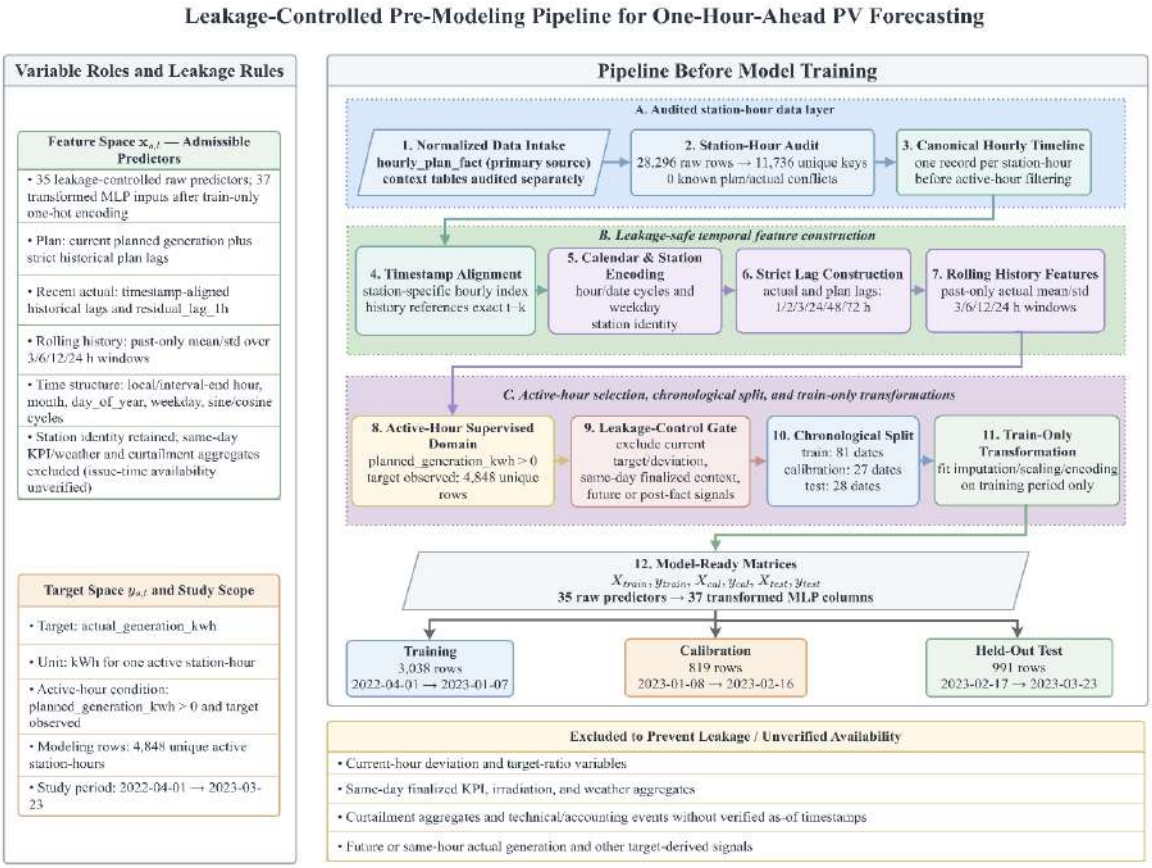


Figure 1. Leakage-controlled pre-modelling pipeline for one-hour-ahead PV generation forecasting

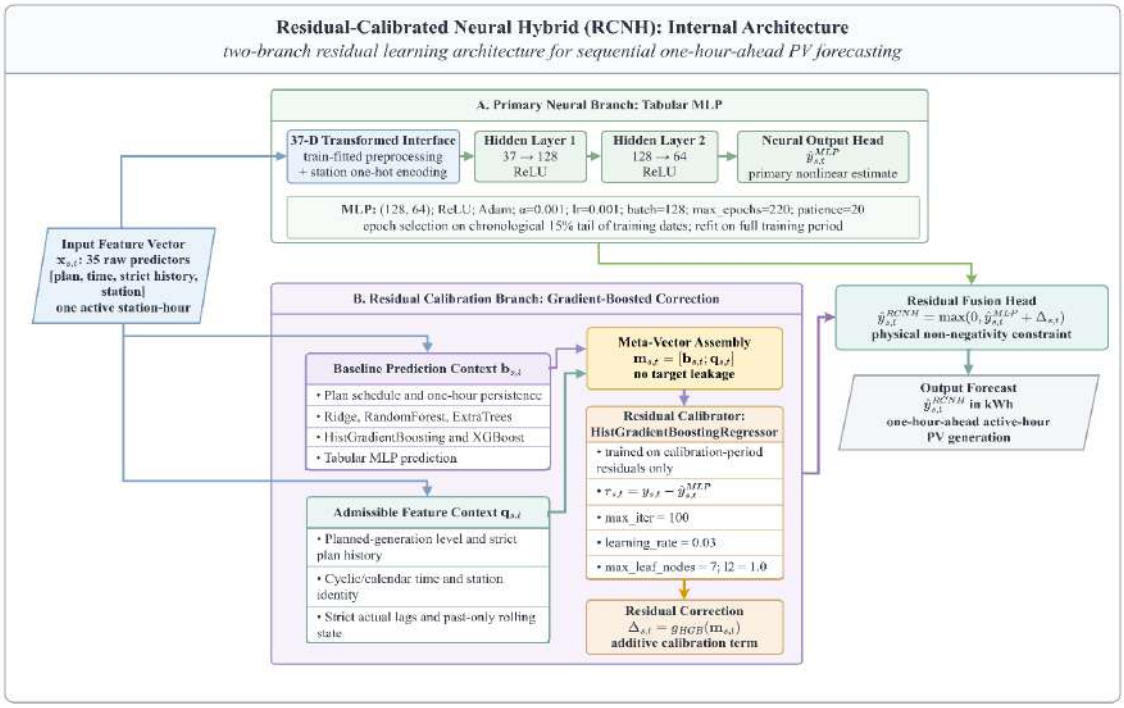


Figure 2. Architecture of the Residual-Calibrated Neural Hybrid forecasting model

Grouped permutation importance is used as model-specific interpretability evidence rather than as a causal ranking [20]. Uncertainty is quantified with a paired moving-day-block bootstrap using 2,000 replicates and three-day blocks; paired absolute-error differences are tested with a centered block-

bootstrap null distribution [21], and Holm correction is applied for multiple comparisons [22]. Robustness is examined through four chronological rolling-origin folds, calibration-window sensitivity, and a matched-label-budget diagnostic in which baselines are refitted on train plus calibration labels. Ablation experiments remove residual calibration, baseline-prediction context, lagged generation history, or plan-schedule features. Grouped permutation importance complements the ablation analysis by measuring sensitivity of the fitted model to perturbations of correlated feature groups [20].

Results of the study

Held-out performance and uncertainty

The comparative performance of all models on the held-out test period is presented in Table 2.

Table 2. Held-out test performance under the primary training/calibration protocol

Model	*MAE, kWh*	*RMSE, kWh*	*R^2*	*WAPE, %*	*sMAPE, %*
RCNH	*2,591.80*	*3,906.53*	*0.908609*	*20.24*	*51.66*
Ridge	*2,802.96*	*4,054.10*	*0.901574*	*21.89*	*50.57*
RandomForest	*2,828.21*	*4,264.43*	*0.891096*	*22.08*	*43.43*
Tabular MLP	*2,923.98*	*4,273.60*	*0.890627*	*22.83*	*51.02*
ExtraTrees	*2,950.09*	*4,318.46*	*0.888319*	*23.03*	*45.39*
XGBoost	*2,885.25*	*4,345.78*	*0.886901*	*22.53*	*45.30*
HistGradientBoosting	*2,951.80*	*4,459.40*	*0.880910*	*23.05*	*44.21*
One-hour persistence	*3,954.09*	*6,362.27*	*0.757593*	*30.87*	*68.57*
Plan schedule	*5,460.12*	*7,900.31*	*0.626226*	*42.63*	*67.20*

As shown in Table 2, on the held-out test period, RCNH achieved the lowest MAE and RMSE and the highest R^2 under the primary protocol: MAE = 2,591.80 kWh, RMSE = 3,906.53 kWh, and R^2 = 0.908609. Relative to the train-only Tabular MLP, MAE decreased by 11.36% and RMSE by 8.59%.

The three-day block-bootstrap 95% interval for RCNH MAE was 2,373-2,962 kWh and for RMSE 3,537-4,505 kWh. Against the train-only MLP, the relative MAE reduction was 11.36% (95% interval 6.98-14.95%), with Holm-adjusted two-sided p = 0.003998. The difference was also statistically supported against Ridge and the two operational baselines, whereas several close tree-based comparisons were not significant after multiplicity correction.

Matched-label-budget diagnostic

The results of the matched-label-budget diagnostic are presented in Table 3.

Table 3. Test performance when baseline models are refitted on train plus calibration labels

Model	*MAE, kWh*	*RMSE, kWh*	*R^2*
Tabular MLP	*2,464.63*	*3,735.04*	*0.916457*
ExtraTrees	*2,659.36*	*3,877.07*	*0.909982*
RCNH	*2,591.80*	*3,906.53*	*0.908609*
RandomForest	*2,610.07*	*3,950.60*	*0.906535*
XGBoost	*2,646.84*	*3,970.03*	*0.905614*
Ridge	*2,729.29*	*3,982.07*	*0.905041*
HistGradientBoosting	*2,735.35*	*4,089.26*	*0.899859*

Table 3 shows that the matched-label-budget diagnostic changes the interpretation of the primary result. When the standalone MLP is allowed to use the same 3,857 labeled observations available to the two-stage RCNH system, it achieves MAE = 2,464.63 kWh, RMSE = 3,735.04 kWh, and R^2 = 0.916457. Thus, part of RCNH's primary advantage reflects allocation of calibration labels rather than residual architecture alone.

Temporal robustness and sensitivity.

The numerical summary of the four rolling-origin validation folds is presented in Table 4.

Table 4. Numerical rolling-origin validation summary (four folds)

Model	*Mean MAE*	*SD MAE*	*Mean RMSE*	*SD RMSE*	*Mean R²*
ExtraTrees	*2,072*	*371*	*3,065*	*537*	*0.868*
RandomForest	*2,016*	*414*	*3,099*	*645*	*0.864*
XGBoost	*2,076*	*366*	*3,122*	*537*	*0.861*
HistGradientBoosting	*2,227*	*396*	*3,311*	*553*	*0.848*
Plan schedule	*2,820*	*873*	*4,516*	*1,348*	*0.693*
One-hour persistence	*2,968*	*626*	*4,863*	*847*	*0.679*
RCNH	*4,843*	*3,076*	*6,107*	*2,823*	*0.424*
Tabular MLP	*5,409*	*3,619*	*6,498*	*3,515*	*0.355*
Ridge	*5,111*	*3,604*	*7,196*	*4,812*	*0.129*

As summarized in Table 4, across four rolling-origin folds, ExtraTrees had the lowest mean RMSE (3,064.51 kWh), followed by RandomForest (3,098.58 kWh) and XGBoost (3,122.24 kWh), whereas RCNH reached 6,106.74 kWh with substantially larger fold-to-fold variation. Using only the most recent 50% or 75% of the calibration pool slightly improved average RCNH error relative to the full pool, reinforcing the dependence on calibration-window representativeness.

Ablation analysis

The effects of removing individual components of RCNH are presented in Table 5.

Table 5. RCNH ablation results on the final held-out test period

Experiment	*MAE, kWh*	*RMSE, kWh*	*R²*	*MAE change vs full*
Full RCNH	*2,591.80*	*3,906.53*	*0.908609*	*+0.00*
Without residual calibration	*2,923.98*	*4,273.60*	*0.890627*	*+332.18*
Without baseline ensemble context	*2,585.35*	*3,901.76*	*0.908832*	*-6.45*
Without lagged generation history	*7,111.10*	*10,013.91*	*0.399479*	*+4,519.31*
Without plan schedule features	*2,599.85*	*3,826.49*	*0.912316*	*+8.05*

Table 5 shows that removing residual calibration reproduced the standalone MLP and increased MAE by 332.18 kWh, while removing lagged generation history increased MAE to 7,111.10 kWh. By contrast, removing baseline-ensemble context or plan-schedule features produced little change after refitting, indicating strong redundancy among correlated predictors. Taken together, the results partially support the working hypothesis. Residual calibration improved the train-only MLP on the held-out test period, but the matched-label-budget and rolling-origin experiments showed that the improvement depends on calibration-window representativeness and temporal regime. Thus, the hypothesis was tested under both a fixed held-out split and additional temporal robustness checks, and the evidence supports a conditional rather than universal advantage of RCNH.

Discussion

The results support a bounded conclusion: residual calibration can materially improve a train-only neural predictor in a favorable held-out regime, but the gain is not an invariant property of the architecture. The matched-label-budget result and rolling-origin validation separate the effects of model design, access to later labeled data, and temporal regime. This distinction matters because the favorable final split should be interpreted as evidence of competitiveness rather than universal dominance. Tree ensembles were markedly more stable across temporal folds, while RCNH was more sensitive to calibration-period representativeness. The data audit is also central to interpretation:

28,296 exported rows correspond to 11,736 unique station-hours, of which only 4,848 form the active supervised sample. Reporting the true experimental unit prevents duplicated exports from artificially inflating sample size and apparent certainty. Feature analyses indicate that recent and rolling generation history is structurally important, but ablation and permutation results also reveal substantial predictor redundancy. The model can therefore be used as a forecasting component in a broader monitoring or digital-twin system, consistent with the role of digital-twin architectures described in prior work [8,9], but this study does not claim implementation of a complete digital twin.

These findings are consistent with prior photovoltaic-forecasting research showing that performance depends strongly on data characteristics, forecast horizon, and operating regime rather than on a single universally best model [3-7]. They also agree with broader evidence that tree-based ensembles remain strong benchmarks for medium-sized tabular prediction problems [13-17]. In this study, the additional contribution is to show that residual calibration can be useful on a favorable later regime while still losing temporal stability to tree ensembles when the regime changes.

The main limitations are the three-plant, 136-active-date scope; absence of hourly irradiance, cloud cover and temperature; unavailable issue timestamps for some operational indicators; and residual-calibrator sensitivity to temporal regime. Future work should extend validation across plants, seasons and years, incorporate verified meteorological and latency metadata, and examine online or drift-aware recalibration.

Conclusion

A leakage-controlled Residual-Calibrated Neural Hybrid was developed and evaluated for one-hour-ahead photovoltaic generation forecasting during active hours. The source audit reduced 28,296 raw rows to 11,736 unique station-hours and 4,848 active supervised observations, providing a defensible basis for chronological evaluation. Under the primary held-out protocol, RCNH achieved MAE = 2,591.80 kWh, RMSE = 3,906.53 kWh, and $R^2 = 0.908609$, with an 11.36% MAE reduction relative to the train-only MLP. However, a matched-label-budget MLP produced better final-test point metrics, and tree ensembles were more stable under rolling-origin validation. Therefore, RCNH should be considered a competitive regime-dependent residual-correction approach rather than a universally superior model. Its practical value is greatest when a recent, representative calibration window is available. The principal contribution is the combined evaluation protocol: unique station-hour auditing, strict prediction-time feature availability, chronological train-calibration-test separation, time-aware uncertainty quantification, matched information budgets, rolling-origin validation, and ablation analysis.

Data availability

The operational dataset contains industrial reporting and maintenance records and cannot be made publicly available because of confidentiality restrictions. The study reports the effective station-hour counts, chronological partitions, model specifications, uncertainty protocol, and aggregate results required to understand the experimental design. Raw source duplication was handled by deterministic station-hour consolidation of identical known values; conflicting known values would terminate the analytical pipeline.

Acknowledgment

This research was funded by the Ministry of Science and Higher Education of the Republic of Kazakhstan, Grant No. AP26105051.

References

[1] Inman R.H., Pedro H.T.C., Coimbra C.F.M. Solar forecasting methods for renewable energy integration // Progress in Energy and Combustion Science. 2013. Vol. 39, No. 6. P. 535-576. DOI: https://doi.org/10.1016/j.pecs.2013.06.002

[2] Diagne M., David M., Lauret P., Boland J., Schmutz N. Review of solar irradiance forecasting methods and a proposition for small-scale insular grids // Renewable and Sustainable Energy Reviews. 2013. Vol. 27. P. 65-76. DOI: https://doi.org/10.1016/j.rser.2013.06.042

[3] Antonanzas J., Osorio N., Escobar R., Urraca R., Martinez-de-Pison F.J., Antonanzas-Torres F. Review of photovoltaic power forecasting // Solar Energy. 2016. Vol. 136. P. 78-111. DOI: https://doi.org/10.1016/j.solener.2016.06.069

[4] Voyant C., Notton G., Kalogirou S., Nivet M.L., Paoli C., Motte F., Fouilloy A. Machine learning methods for solar radiation forecasting: A review // Renewable Energy. 2017. Vol. 105. P. 569-582. DOI: https://doi.org/10.1016/j.renene.2016.12.095

[5] Ahmed R., Sreeram V., Mishra Y., Arif M.D. A review and evaluation of the state-of-the-art in PV solar power forecasting: Techniques and optimization // Renewable and Sustainable Energy Reviews. 2020. Vol. 124. Article 109792. DOI: https://doi.org/10.1016/j.rser.2020.109792

[6] Iheanetu K.J. Solar photovoltaic power forecasting: A review // Sustainability. 2022. Vol. 14, No. 24. Article 17005. DOI: https://doi.org/10.3390/su142417005

[7] Gaboitaolelwe J., Zungeru A.M., Yahya A., Lebekwe C.K., Vinod D.N., Salau A.O. Machine learning based solar photovoltaic power forecasting: A review and comparison // IEEE Access. 2023. Vol. 11. P. 40820-40845. DOI: https://doi.org/10.1109/ACCESS.2023.3270041

[8] Fuller A., Fan Z., Day C., Barlow C. Digital twin: Enabling technologies, challenges and open research // IEEE Access. 2020. Vol. 8. P. 108952-108971. DOI: https://doi.org/10.1109/ACCESS.2020.2998358

[9] Angelova D.D. et al. A review on digital twins and its application in the modeling of photovoltaic installations // Energies. 2024. Vol. 17, No. 5. Article 1227. DOI: https://doi.org/10.3390/en17051227

[10] Kaufman S., Rosset S., Perlich C., Stitelman O. Leakage in data mining: Formulation, detection, and avoidance // ACM Transactions on Knowledge Discovery from Data. 2012. Vol. 6, No. 4. Article 15. DOI: https://doi.org/10.1145/2382577.2382579

[11] Hyndman R.J., Athanasopoulos G. Forecasting: Principles and Practice. 3rd ed. Melbourne: OTexts, 2021. URL: https://otexts.com/fpp3/

[12] Bergmeir C., Hyndman R.J., Koo B. A note on the validity of cross-validation for evaluating autoregressive time series prediction // Computational Statistics & Data Analysis. 2018. Vol. 120. P. 70-83. DOI: https://doi.org/10.1016/j.csda.2017.11.003

[13] Friedman J.H. Greedy function approximation: A gradient boosting machine // The Annals of Statistics. 2001. Vol. 29, No. 5. P. 1189-1232. DOI: https://doi.org/10.1214/aos/1013203451

[14] Breiman L. Random forests // Machine Learning. 2001. Vol. 45. P. 5-32. DOI: https://doi.org/10.1023/A:1010933404324

[15] Chen T., Guestrin C. XGBoost: A scalable tree boosting system // Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2016. P. 785-794. DOI: https://doi.org/10.1145/2939672.2939785

[16] Wolpert D.H. Stacked generalization // Neural Networks. 1992. Vol. 5, No. 2. P. 241-259. DOI: https://doi.org/10.1016/S0893-6080(05)80023-1

[17] Grinsztajn L., Oyallon E., Varoquaux G. Why do tree-based models still outperform deep learning on typical tabular data? // Advances in Neural Information Processing Systems. 2022. Vol. 35. P. 507-520. DOI: https://doi.org/10.52202/068431-0037

[18] Hyndman R.J., Koehler A.B. Another look at measures of forecast accuracy // International Journal of Forecasting. 2006. Vol. 22, No. 4. P. 679-688. DOI: https://doi.org/10.1016/j.ijforecast.2006.03.001

[19] Wilcoxon F. Individual comparisons by ranking methods // Biometrics Bulletin. 1945. Vol. 1, No. 6. P. 80-83. DOI: https://doi.org/10.2307/3001968

[20] Fisher A., Rudin C., Dominici F. All models are wrong, but many are useful: Learning a variable's importance by studying an entire class of prediction models simultaneously // Journal of Machine Learning Research. 2019. Vol. 20, No. 177. P. 1-81. URL: https://www.jmlr.org/papers/v20/18-760.html

[21] Kunsch H.R. The jackknife and the bootstrap for general stationary observations // The Annals of Statistics. 1989. Vol. 17, No. 3. P. 1217-1241. DOI: https://doi.org/10.1214/aos/1176347265

[22] Holm S. A simple sequentially rejective multiple test procedure // Scandinavian Journal of Statistics. 1979. Vol. 6, No. 2. P. 65-70. DOI: https://doi.org/10.2307/4615733

МРНТИ 28.23.29

https://doi.org/10.51889/2959-5894.2026.95.3.013

Абдрахманов Р.Е.[1,2], Алимова А.Н.[1*], Соколов С.В.[1], Чалова Н.Р.[1], Нурсеитов Д.Б.[1]

[1]Институт информационных и вычислительных технологий, г. Алматы, Казахстан
[2]Казахский национальный исследовательский технический университет имени К.И. Сатпаева, г. Алматы, Казахстан
[*]*e-mail: anic2002@gmail.com*

НАБОР ДАННЫХ И БЕНЧМАРК ДЛЯ ТОНКОЗЕРНИСТОЙ КЛАССИФИКАЦИИ ОХОТНИЧЬИХ ВОДОПЛАВАЮЩИХ ПТИЦ

Аннотация

Идентификация птиц по виду и полу имеет важное значение для учета добычи, мониторинга состояния охотничьих ресурсов и соблюдения установленных правил охоты. Существующие наборы данных и цифровые приложения преимущественно определяют вид, не учитывая половую принадлежность особей, что ограничивает их применение в охотничьем хозяйстве. В работе представлен специализированный набор данных GameWaterbird-41 для совместной классификации вида и пола охотничьих водоплавающих птиц и куликов, являющихся объектами охоты. Набор данных сформирован на основе материалов iNaturalist и включает 18 851 изображение 38 видов, распределенных по 41 классу. Проведен бенчмарк сверточных нейронных сетей и визуальных трансформеров в трех режимах: без дополнительного обучения, с обучением линейного классификатора и с полной тонкой настройкой. Рассмотрены модели общего назначения и модели, предварительно обученные на биологических изображениях. Наиболее высокое качество классификации получено при полной тонкой настройке на целевых данных. Набор данных и результаты бенчмарка могут быть использованы для совершенствования учета добычи и информационной поддержки управления охотничьими ресурсами.

Ключевые слова: тонкозернистое визуальное распознавание, классификация видов птиц, половой диморфизм, водоплавающие птицы, трансферное обучение, температурное масштабирование, аугментация при тестировании.

Абдрахманов Р.Е.[1,2], Алимова А.Н.[1*], Соколов С.В.[1], Чалова Н.Р.[1], Нурсеитов Д.Б.[1]
[1]Ақпараттық және есептеу технологиялары институты, Алматы қ., Қазақстан
[2]Қ.И. Сәтбаев атындағы Қазақ ұлттық техникалық зерттеу университеті, Алматы қ., Қазақстан

АҢШЫЛЫҚ СУ ҚҰСТАРЫНЫҢ ТҮРЛЕРІН ҰСАҚ ТҮЙІРШІКТІ ЖІКТЕУГЕ АРНАЛҒАН ДЕРЕКТЕР ЖИЫНТЫҒЫ ЖӘНЕ БЕНЧМАРК

Аңдатпа

Құстарды түрі мен жынысына қарай сәйкестендіру ауланған құстардың есебін жүргізу, аңшылық ресурстардың жағдайын мониторингтеу және белгіленген аңшылық ережелерінің сақталуын қамтамасыз ету үшін маңызды. Қолданыстағы деректер жиынтықтары мен цифрлық қосымшалар негізінен құстың түрін ғана анықтауға бағытталған, ал жеке құстардың жыныстық ерекшелігін ескермейді. Бұл олардың аңшылық шаруашылығында қолданылу мүмкіндігін шектейді. Бұл жұмыста аңшылық нысандары болып табылатын су құстары мен шалшықшылардың түрі мен жынысын бірге жіктеуге арналған GameWaterbird-41 мамандандырылған деректер жиынтығы ұсынылады. Деректер жиынтығы iNaturalist материалдары негізінде құрастырылған және 41 класқа бөлінген 38 түрге жататын 18 851 кескінді қамтиды. Конволюциялық нейрондық желілер мен визуалды трансформерлерге үш режимде: қосымша оқытусыз, сызықтық классификаторды оқыту арқылы және модельді толықтай дәлдеп оқыту арқылы бенчмарк-бағалау жүргізілді. Жалпы мақсаттағы модельдер және биологиялық кескіндер негізінде алдын ала оқытылған модельдер қарастырылды. Жіктеудің ең жоғары сапа көрсеткіштеріне мақсатты деректер жиынтығында модельді толықтай дәлдеп оқыту барысында қол жеткізілді. Деректер жиынтығы мен бенчмарк нәтижелері ауланған жануарлардың

есебін жүргізу жүйесін жетілдіру және аңшылық ресурстарды басқаруды ақпараттық қолдау мақсатында пайдаланылуы мүмкін.

Түйін сөздер: егжей-тегжейлі визуалды тану, құс түрлерін жіктеу, жыныстық диморфизм, су құстары, трансферлік оқыту, температуралық масштабтау, тестілеу кезіндегі аугментация.

R. Abdrakhmanov[1,2], A. Alimova[1*], S. Sokolov[1], N. Chalova[1], D. Nurseitov[1]
[1]The Institute of Information and Computational Technologies, Almaty, Kazakhstan
[2]K.I. Satbayev Kazakh National Research Technical University, Almaty, Kazakhstan

DATASET AND BENCHMARK FOR FINE-GRAINED CLASSIFICATION OF HUNTING WATERFOWL

Abstract

Identification of birds by species and sex is important for harvest accounting, monitoring of game resource state, and compliance with established hunting regulations. Existing datasets and digital applications predominantly identify species without accounting for the sex of individuals, which limits their applicability in game management. This paper presents a specialized dataset, GameWaterbird-41, for combined classification of species and sex of hunting waterfowl and waders that are game species. The dataset was composed based on materials from iNaturalist and includes 18,851 images of 38 species distributed across 41 classes. A benchmark of convolutional neural networks and vision transformers was conducted in such classification modes as zero-shot classification mode, classification mode with linear classifier training, and classification mode with full fine-tuning. General-purpose models and models pre-trained on biological images were considered. The highest classification quality was obtained with full fine-tuning on the target data. The dataset and benchmark results can be used to improve game harvest accounting and information support for game resource management.

Keywords: fine-grained visual recognition, bird species classification, sexual dimorphism, waterfowl, transfer learning, temperature scaling, test-time augmentation.

Введение

Идентификация вида и пола птицы по фотографии является классической задачей тонкозернистого визуального распознавания [1, 2]. В контексте охотничьего хозяйства данная задача имеет практическое значение, поскольку нормативные документы, регулирующие охоту, определяют не только перечень видов, разрешенных к добыче, но и половую принадлежность особей. Для многих видов уток характерен половой диморфизм. Например, самцы и самки кардинально различаются по окраске, что позволяет использовать их визуальные признаки для идентификации. У других видов, в частности у гусей, оба пола внешне неразличимы, однако сам вид может быть разрешен или запрещен к добыче. Надежный автоматический классификатор способен помочь охотникам правильно определять птицу в полевых условиях, а субъектам охотничьего хозяйства масштабировать мониторинг популяций на основе научных данных.

В настоящее время учет добываемых водоплавающих птиц в охотничьем хозяйстве преимущественно ведется по укрупненным таксономическим группам (утки, кулики) [3]. При этом нормативные документы регламентируют добычу конкретных видов, а в ряде случаев учитывают и половую принадлежность особей [4, 5]. В связи с этим разработка автоматизированных методов идентификации вида и пола птиц по фотографиям представляет практический интерес для повышения качества учета добычи, мониторинга охотничьих ресурсов и информационной поддержки принятия управленческих решений в охотничьем хозяйстве.

Существующие крупномасштабные наборы данных для распознавания птиц, такие как CUB-200-2011 [1], NABirds [2], Birdsnap [6] и наборы данных iNaturalist [7], ориентированы преимущественно на идентификацию вида среди сотен и тысяч классов. При этом CUB-200-2011, NABirds и Birdsnap сформированы главным образом на основе орнитофауны Северной Америки. Существующие цифровые приложения для идентификации птиц, включая Merlin Bird ID и iNaturalist, имеют более широкий географический охват, однако не ориентированы

специально на задачи учета охотничьих видов Казахстана и, как правило, не предусматривают определение пола особи. В отличие от них, разработанный в настоящем исследовании набор данных ориентирован на виды водоплавающих птиц и куликов, являющиеся объектами охоты и обитающие на территории Республики Казахстан или мигрирующих через нее, а для видов с выраженным половым диморфизмом предусматривает классификацию с учетом пола особи.

Предварительно обученные (фундаментальные) модели, такие как CLIP [8] и его биологическая модификация BioCLIP [9], показывают высокую точность при распознавании вида без дополнительного обучения (zero-shot). Такие модели предварительно обучаются на больших наборах изображений общего назначения и затем могут быть адаптированы для решения новых задач. Однако вопрос о том, способны ли такие модели решать более сложную задачу совместной классификации вида и пола без доменно-специфической тонкой настройки, остается открытым.

Несмотря на существование крупных наборов данных и современных моделей компьютерного зрения, специализированные наборы данных, ориентированные на совместную классификацию вида и пола охотничьих водоплавающих птиц, в настоящее время отсутствуют. Кроме того, недостаточно изучены возможности современных фундаментальных моделей при решении данной задачи. Целью настоящего исследования является разработка специализированного набора данных GameWaterbird-41 и проведение сравнительного анализа современных моделей компьютерного зрения для совместной классификации вида и пола охотничьих водоплавающих птиц как основы для создания цифрового определителя птиц.

Разработанный набор данных и результаты сравнительного анализа моделей могут быть использованы при создании цифровых систем поддержки охотничьего хозяйства, обеспечивающих автоматическое определение вида и пола птиц по фотографиям, полученным в полевых условиях. Использование подобных систем позволит повысить качество учета добычи, обеспечить получение объективной информации о видовом составе охотничьих ресурсов и расширить возможности их мониторинга.

Методология исследования

Тонкозернистое распознавание птиц. Набор данных CUB-200-2011 [1] является одним из используемых бенчмарков для тонкозернистого распознавания птиц по морфологическим признакам на уровне вида. Он включает 11 788 изображений 200 видов птиц Северной Америки. Набор данных NABirds [2] расширяет данный подход до 555 визуальных категорий, включая различия по полу и возрасту и содержит 48 000 изображений, собранных пользователями. Birdsnap [6] охватывает 500 видов птиц Северной Америки. Платформа iNaturalist [7] расширила область применения методов тонкозернистого распознавания до тысяч таксономических классов различных групп живых организмов. В отличие от перечисленных наборов данных, разработанный в настоящем исследовании набор данных включает значительно меньшее число видов и изображений, но ориентирован на решение специализированной регуляторно-экологической задачи классификации охотничьих водоплавающих птиц и куликов, являющихся объектами охоты, и дополнительно содержит информацию о половой принадлежности особей.

Фундаментальные модели для распознавания объектов дикой природы. Под моделью в работе понимается обученная на данных программа, которая по изображению оценивает вероятность принадлежности объекта к каждому из классов. Модель CLIP [8] показала, что визуально-языковое предобучение на 400 млн пар «изображение–текст» обеспечивает высокую эффективность классификации без дополнительного обучения на целевом наборе данных с использованием текстовых описаний классов. BioCLIP [9] представляет собой специализированную модификацию CLIP, адаптированную для биологических задач. Модель предварительно обучена на наборе данных TreeOfLife-10M, охватывающем 450 000 таксонов, и достигла высокой точности при распознавании биологических объектов. Модель

DINOv2 [10] основана на самонадзорном обучении и использует набор данных из 142 млн изображений без применения текстовых аннотаций. Получаемые признаки дают высокую эффективность по сравнению с другими подходами при решении задач тонкозернистой классификации как при использовании линейного классификатора, так и после тонкой настройки модели.

Визуальные трансформеры. Архитектура Vision Transformer (ViT) [11] показала, что трансформерные модели при масштабном предварительном обучении могут превосходить сверточные нейронные сети при анализе изображений. В работе [12] исследовано влияние аугментации данных и методов регуляризации на качество обучения моделей семейства ViT (AugReg). Аугментация данных представляет собой искусственное расширение набора данных за счет преобразований изображений, таких как сдвиг, отражение, поворот и изменение яркости. Полученные авторами контрольные точки ViT-B/16 показали, что модель хорошо обобщается при тонкой настройке.

Калибровка модели и TTA. Модель называют калиброванной, если заявленный ею уровень уверенности соответствует фактической доле правильных предсказаний. Например, среди предсказаний с вероятностью 90% правильными должны быть примерно девять из десяти. В исследовании [13] показано, что один обучаемый параметр температуры, применяемый к логитам, то есть выходным числовым значениям модели до их преобразования в вероятности, эффективно калибрует апостериорные вероятности без потери точности классификации. Такой подход называется температурным масштабированием и заключается в корректировке вероятностей нейронной сети с помощью параметра температуры, подбираемого на валидационных данных без переобучения самой модели. Направление ошибки калибровки зависит от используемой функции потерь. Модели, обученные со сглаживанием меток и функцией потерь Focal Loss [14], как правило, характеризуются заниженной уверенностью, при которой вероятности распределяются по классам более равномерно, чем это соответствует фактической достоверности предсказаний, и требуют применения температуры $T < 1$ для корректировки распределения вероятностей. Именно это наблюдается у всех дообученных моделей в настоящем исследовании, оптимальные значения температуры находятся в диапазоне от 0,49 до 0,69. Аугментация при тестировании (TTA) представляет собой применение преобразований изображения не на этапе обучения, а на этапе распознавания. При этом для одного исходного изображения формируется несколько аугментированных вариантов, а полученные для них предсказания усредняются. Такой подход снижает дисперсию предсказаний и повышает их устойчивость к изменениям изображения.

Классификация пола и вида птиц. Автоматическое определение пола птиц привлекло значительно меньше внимания, чем идентификация вида. Большинство исследователей либо рассматривают его как дополнительную задачу предсказания атрибута [2], либо ограничиваются мономорфными видами, исключая необходимость определения пола. Разработанный в настоящем исследовании набор данных включает пол в качестве основной цели классификации для диморфных видов, что делает его одним из немногих бенчмарков, ориентированных на совместную классификацию вида и пола птиц.

Набор данных GameWaterbird-41. Каталог видов сформирован на основе официального перечня животных, являющихся объектами регулируемой охоты в Республике Казахстан [4]. В него включены все водоплавающие и болотно-луговые птицы, являющиеся объектами охоты: речные и нырковые утки (семейство Anatidae), гуси (Anser spp.), крохали, пеганки и кулики (кроншнепы, веретенники, вальдшнеп и родственные виды). Всего в собранном наборе данных включено 38 видов. С точки зрения задачи классификации эти виды делятся на две группы:

– Полово-диморфные виды (25 видов). Самцы и самки имеют выраженные различия во внешнем облике и имеют разное оперение. В эту группу входят большинство речных уток (кряква, шилохвость, чирки, свиязь, широконоска и др.) и нырковых уток (чернети, нырки,

гоголь, морянка и др.). Для этих видов определение пола по фотографии возможно и имеет регуляторное значение.

– Мономорфные виды (13 видов). Самцы и самки практически не отличаются по внешнему виду и имеют одинаковое оперение. В эту группу входят все гуси рода *Anser*, а также вальдшнеп, кроншнепы и веретенники. Для этих видов классификация выполняется только на уровне вида.

Изображения получены из базы данных iNaturalist [15] с ограничением по уровню качества Research Grade (верифицированные наблюдения) и открытых лицензий (CC0, CC-BY, CC-BY-SA, CC-BY-NC, CC-BY-NC-SA). На рисунке 1 первые два изображения демонстрируют выраженный половой диморфизм у кряквы. Самец кряквы имеет яркое переливчатое оперение, тогда как для самки характерна бурая покровительственная окраска. Изображение белолобого гуся демонстрирует типичную проблему межвидового сходства среди мономорфных видов, затрудняющую их автоматическую классификацию. Загруженные изображения были обработаны детектором объектов YOLOv11x [16], предварительно обученным на наборе данных COCO.



Кряква ♂ Anas platyrhynchos | Кряква ♀ Anas platyrhynchos | Хохл. чернеть ♂ Aythya fuligula | Белолобый гусь Anser albifrons | Большой кроншнеп Numenius arquata

Рисунок 1. Примеры изображений из GameWaterbird-41, полученных из iNaturalist

В итоговый набор данных включались только изображения, которые удовлетворяли следующим критериям:

– на изображении присутствует только одна птица;

– ограничивающий прямоугольник птицы занимает не менее 20% площади изображения.

Применение этих критериев позволило исключить изображения с несколькими особями, удаленными объектами или частично перекрытыми птицами. Координаты ограничивающего прямоугольника сохраняются для каждого изображения и в дальнейшем используются на этапе инференса для автоматической обрезки области птицы.

Платформа iNaturalist позволяет пользователям указывать половую принадлежность особи («самец», «самка» или «не определен»). Для 25 диморфных видов в набор данных включаются только изображения с подтвержденной меткой пола «самец» или «самка». Изображения с меткой «не определен» или без указания пола исключались с целью снижения уровня шума в аннотациях. Для 13 мономорфных видов в набор данных включались все изображения, удовлетворяющие критериям качества, независимо от наличия информации о поле.

После сбора и фильтрации виды или половые группы, содержащие менее 50 изображений, объединялись в единый класс «прочие». В результате была сформирована следующая структура классов:

– 14 из 25 диморфных видов содержат не менее 50 изображений как самцов, так и самок. Для каждого такого вида сформировано два отдельных класса, что в сумме составляет 28 классов (14×2);

– 12 из 13 мономорфных видов содержат не менее 50 изображений и представлены 12 отдельными классами;

– оставшиеся 12 видов (11 диморфных и 1 мономорфный), не удовлетворяющие минимальному порогу по числу изображений, объединены в один класс «прочие».

Итоговая таксономия набора данных содержит 41 класс $(28 + 12 + 1)$, а общий объем набора данных составляет 18 851 изображение. Размер отдельных классов варьируется от 50 до 1 010 изображений, что отражает неравномерную представленность различных видов в базе данных iNaturalist. Набор данных был разделен на обучающую (75%), валидационную (15%) и тестовую (10%) выборки методом стратифицированной выборки по классам. В результате обучающая выборка содержит 14 182 изображения, валидационная 2 804 изображения, а тестовая 1 865 изображений. Сводные характеристики набора данных представлены в таблице 1, а распределение изображений по классам набора данных GameWaterbird-41 представлено на рисунке 2.

Таблица 1. Сводные характеристики набора данных GameWaterbird-41

Характеристика	*Значение*
Всего изображений	*18 851*
Всего видов	*38*
Полово-диморфные	*25*
Мономорфные	*13*
Всего классов	*41*
Половые классы (14×2*)*	*28*
Классы мономорфных видов	*12*
Прочие	*1*
Изображений в классе (мин / макс)	*50 / 1 010*
Обучение / Валидация / Тест	*14 182 / 2 804 / 1 865*
Источник изображений	*iNaturalist (глобально)*
Детектор	*YOLOv11x, bbox ≥20% кадра*

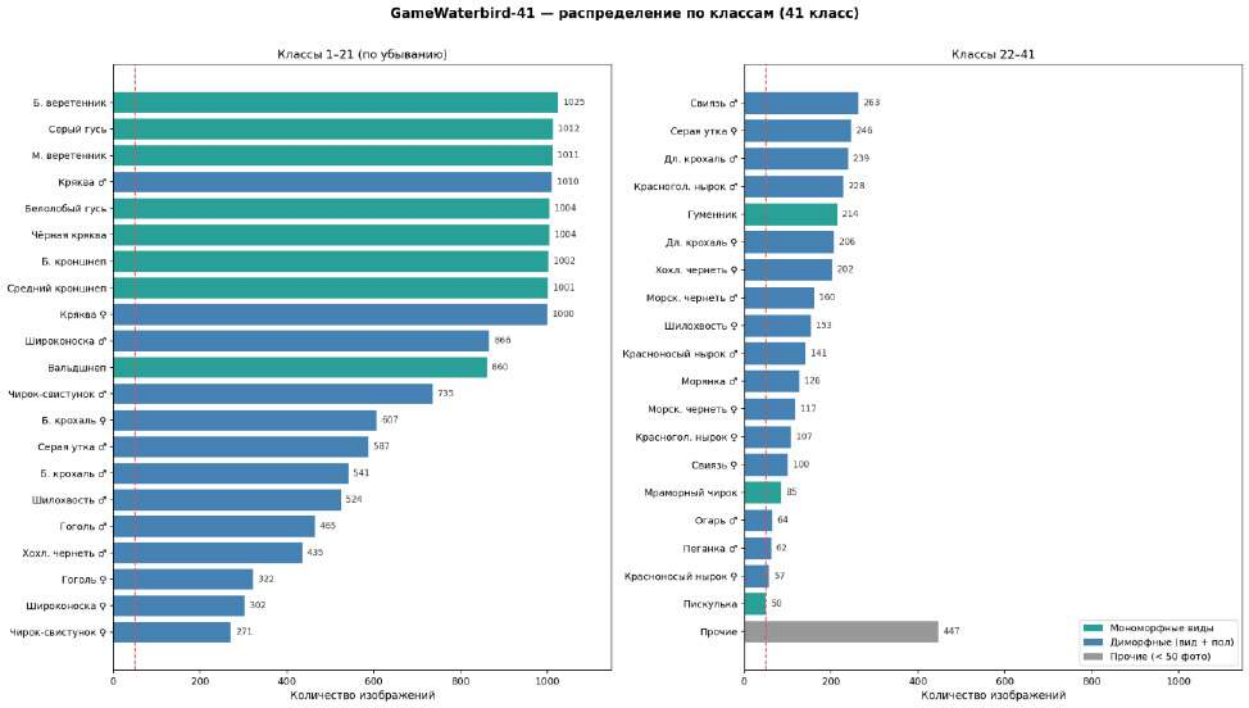


Рисунок 2. Распределение изображений по классам GameWaterbird-41

Для повышения читаемости диаграмма разделена на две панели, а классы отсортированы по убыванию количества изображений. Красная пунктирная линия означает порог в 50 изображений, используемый при формировании классов: виды и пол, содержащие меньшее число изображений, объединяются в класс «прочие». Зеленым цветом обозначены мономорфные виды, синим полово-диморфные классы («вид + пол»).

Постановка экспериментов

В данном исследовании оценивались шесть семейств моделей, охватывающих два типа архитектур (сверточные нейронные сети и визуальные трансформеры) и три подхода к предварительному обучению. Характеристики используемых моделей описаны в таблице 2.

Таблица 2. Модели, включенные в бенчмарк

Модель	*Архитектура*	*Количество параметров*	*Модель предобучена на датасете*	*Входное разрешение в пикселях*
ResNet-50 [17]	*CNN*	*25M*	*ImageNet-1K (надзор.)*	*224*
EfficientNet-B3 [18]	*CNN*	*12M*	*ImageNet-1K (надзор.)*	*224*
CLIP ViT-B/16 [8]	*ViT-B/16*	*150M*	*Изобр.–текст (400M)*	*224*
BioCLIP ViT-B/16 [9]	*ViT-B/16*	*150M*	*Биологич. фото (10M)*	*224*
ViT-B/16 AugReg [12]	*ViT-B/16*	*86M*	*ImageNet-21K (надзор.)*	*224*
DINOv2 ViT-B/14 [10]	*ViT-B/14*	*87M*	*Самонадзор. (142M)*	*518*

Для оценки возможностей различных подходов рассматривались три режима обучения. Модели CLIP и BioCLIP оценивались без доменно-специфического обучения. В этом режиме сравниваются эмбеддинги (векторы признаков) изображения и текстовых описаний классов. Эмбеддинг представляет собой числовое представление изображения, формируемое моделью, при котором визуально похожие изображения имеют близкие векторы признаков. Были протестированы различные варианты текстовых подсказок (научное название, русское название и их комбинации), после чего для каждой модели фиксировался наилучший результат. Поскольку данные модели не имеют механизма предсказания пола, оценка проводилась только на уровне вида. Предсказание считалось верным, если вид был определен правильно, вне зависимости от пола особи.

Для моделей CLIP и BioCLIP также исследовался режим обучения линейного классификатора. При этом визуальный кодировщик использовался с замороженными весами, то есть его параметры не изменялись, а обучался только добавленный поверх него линейный классификатор. Линейный классификатор обучался для распознавания всех 41 классов, при этом веса базовой сети оставались неизменными. Такой режим позволяет оценить, насколько хорошо замороженное признаковое пространство разделяет классы «вид + пол» без полной адаптации модели.

На заключительном этапе шесть моделей дообучались в режиме полной тонкой настройки для решения задачи классификации 41 класса. Поверх каждой базовой модели добавлялась классификационная головка, состоящая из слоя LayerNorm и полносвязного слоя Linear(feat_dim, 41). Параметры базовой сети и классификационной головы обновлялись совместно с дифференциальными скоростями обучения. Параметры обучения, приведенные ниже, необходимы для воспроизведения экспериментов, однако для интерпретации полученных результатов их детальное понимание не требуется.

Все модели обучались с использованием единой конфигурации. В качестве оптимизатора применялся AdamW с коэффициентом затухания весов 0,05. Скорость обучения для классификационной головы составляла 10^{-3}, а для базовой сети 10^{-5}. Обучение проводилось в течение 25 эпох с использованием косинусного отжига и 2-эпохным линейным прогревом. Для компенсации дисбаланса классов (от 50 до 1 010 изображений на класс) применялись WeightedRandomSampler и функция потерь [14] с параметром $\gamma = 2{,}0$. Весовые коэффициенты классов задавались обратно пропорционально их частоте, а для дополнительного повышения устойчивости обучения использовались сглаживание меток со значением 0,1.

Аугментация обучающих данных включала случайное масштабированное кадрирование (масштаб 0,7–1,0), случайное горизонтальное отражение, поворот изображения в диапазоне $\pm 15^{\circ}$ и случайное изменение яркости, контрастности и насыщенности в диапазоне $\pm 0{,}2$. Все преобразования применялись только к области изображения, ограниченной прямоугольником, содержащим птицу. На этапе тестирования использовалась единственная центральная обрезка изображения.

Все модели, кроме DINOv2 ViT-B/14, использовали входное разрешение 224×224 пикселя. Для модели DINOv2 ViT-B/14 использовалось разрешение 518×518 пикселей, соответствующее ее исходной архитектуре ViT-B/14 с размером патча 14×14 и решеткой признаков 37×37. Обучение всех моделей проводилось на одной видеокарте NVIDIA RTX 4080 SUPER с объемом 16 ГБ.

После полной тонкой настройки моделей выполнялось температурное масштабирование. Для этого все логиты модели делились на скалярный параметр температуры, после чего к ним применяется нормирующее преобразование softmax:

$$p_i = \frac{\exp\,(z_i/T)}{\sum_j \exp\,(z_j/T)}.$$

Значение температуры подбиралось на валидационной выборке путем минимизации отрицательного логарифмического правдоподобия. При этом веса модели не изменялись, поскольку процедура выполнялась после завершения обучения. Температурное масштабирование не изменяет класс с максимальной вероятностью, поэтому точность Top-1 остается неизменной, но можно скорректировать уровень уверенности модели в предсказаниях. Во всех проведенных экспериментах оптимальное значение температуры оказалось меньше единицы. Это указывает на заниженную уверенность моделей, и характерной особенностью обучения с использованием сглаживания меток и функции потерь. В отличие от типичной для нейросетевых классификаторов избыточной уверенности, которая требует температуры больше единицы для сглаживания распределения вероятностей, здесь наблюдается обратный эффект: температура меньше единицы делает распределение вероятностей более выраженным.

На этапе тестирования дополнительно исследовалось влияние аугментации. Для каждого тестового изображения формировалось восемь детерминированных вариантов, включающих центральные обрезки трех масштабов, их горизонтальные отражения, вариант с изменением цветовых характеристик и вариант с поворотом изображения. Итоговые предсказания получали путем усреднения логитов всех вариантов изображения. При совместном использовании методов усреднение логитов выполнялось перед применением температурного масштабирования.

Качество моделей оценивалось по показателям точности Top-1 и Top-5 на тестовой выборке, содержащей 1 865 изображений. Для оценки качества калибровки вероятностных предсказаний использовалась ожидаемая ошибка калибровки (ECE), вычисляемая по 15 равным интервалам [13]. По смыслу ECE представляет собой средневзвешенное расхождение между заявленной уверенностью модели и фактической долей правильных ответов. Для её

вычисления предсказания разбиваются на интервалы по уровню уверенности, после чего в каждом интервале заявленная уверенность сравнивается с фактической точностью. Нулевое значение ECE соответствует идеальной калибровке. Для режима без дополнительного обучения и режима обучения только линейного классификатора значение ECE не рассчитывалось, поскольку показатели калибровки несопоставимы между режимами.

Метрики качества. Основным показателем качества классификации является точность Top-1, определяемая как доля изображений, для которых верный класс получил наибольшую оценку среди всех 41 класса. Дополнительно рассчитывается точность Top-5 – доля изображений, для которых верный класс вошел в число пяти наиболее вероятных. Второй показатель показывает, насколько близко модель была к правильному ответу в тех случаях, когда все же ошиблась.

$$\text{Top} - 1 = \frac{\text{M}_1}{\text{N}}, \qquad \text{Top} - 5 = \frac{\text{M}_5}{\text{N}},$$

где N – размер тестовой выборки (1 865 изображений), M_1 – число изображений, для которых верный класс получил наибольшую оценку, M_5 – число изображений, для которых верный класс попал в первую пятерку.

Для анализа качества распознавания отдельных классов используются две метрики: полнота (Recall) и точность (Precision). При расчете этих метрик каждый класс k рассматривается независимо от остальных, и все тестовые изображения распределяются по одной из четырех категорий:

– TP (истинно-положительные) изображения класса k, правильно отнесенные моделью к классу k;

– FN (ложно-отрицательные) изображения класса k, ошибочно отнесенные моделью к другому классу;

– FP (ложно-положительные) изображения других классов, ошибочно отнесенные моделью к классу k;

– TN (истинно-отрицательные) изображения других классов, которые модель к классу k и не относила.

На основе этих категорий рассчитываются Recall и Precision:

$$Recall = \frac{TP}{(TP + FN)}, \qquad Precision = \frac{TP}{(TP + FP)}.$$

Recall отвечает на вопрос: «Какую долю изображений птиц данного класса модель распознала правильно?», тогда как Precision отвечает на вопрос: «Какая доля изображений птиц, отнесенных моделью к данному классу, действительно принадлежит этому классу?». Обе метрики определяются непосредственно по матрице ошибок, где Recall вычисляется по строке соответствующего класса, а Precision по его столбцу.

Например, для самки серой утки (Mareca strepera) из 24 тестовых изображений модель верно распознала 19 (TP), пять отнесла к другим классам (FN), а еще девять изображений других видов ошибочно отнесла к этому классу (FP). В результате Recall составила 19/24=79%, а Precision 19/28=68%. Такое расхождение свидетельствует о том, что к классу самки серой утки ошибочно относятся самки других видов, прежде всего кряквы, что согласуется с наблюдением о криптической окраске самок.

Термин «точность» используется в статье в двух значениях: точность Top-1 – доля правильных классификаций на всей тестовой выборке, а Precision – доля правильно классифицированных изображений среди всех изображений, отнесенных моделью к данному классу. Во избежание разночтений далее для второй метрики сохраняется английский термин Precision, а для доли правильно распознанных изображений данного класса используется термин «Полнота» (recall).

Результаты экспериментов

В таблице 3 представлены результаты сравнительного анализа (бенчмарка) всех моделей на тестовой выборке, содержащей 1 865 изображений, относящихся к 41 классу. Для моделей, прошедших полную тонкую настройку, приведены значения точности Top-1, Top-5 и ожидаемой ошибки калибровки (ECE), полученные при совместном использовании температурного масштабирования и восьмикратной аугментации на этапе тестирования (TTA×8), обеспечивших наилучшее качество классификации. В режиме без дополнительного обучения (zero-shot) оценка проводилась только на уровне вида: модель предсказывала вид по текстовым подсказкам без учета пола особи, а предсказание считалось верным при корректном определении вида вне зависимости от пола. Это делает результаты несопоставимыми напрямую с дообученными моделями, которые выполняют классификацию по полному набору классов «вид + пол». Значение ожидаемой ошибки калибровки (ECE) указывается только для дообученных моделей после полной тонкой настройки.

Строки таблицы сгруппированы по режимам обучения. Колонки Top-1 и Top-5 отражают долю правильно классифицированных изображений: чем выше значение, тем лучше качество модели. Колонка ECE характеризует расхождение между уровнем уверенности модели и фактической точностью её предсказаний: чем меньше значение, тем лучше калибровка модели.

Таблица 3. Результаты бенчмарка моделей на тестовой выборке GameWaterbird-41

Модель	*Режим*	*Top-1 (%)*	*Top-5 (%)*	*ECE*
CLIP ViT-B/16	*Zero-shot*	*47,51*	*87,02*	–
BioCLIP ViT-B/16	*Zero-shot*	*67,99*	*90,83*	–
CLIP ViT-B/16	*Линейный классификатор*	*31,80*	*69,71*	–
BioCLIP ViT-B/16	*Линейный классификатор*	*54,10*	*82,52*	–
ResNet-50	*Тонкая настройка (T+TTA)*	*74,64*	*93,46*	*0,051*
EfficientNet-B3	*Тонкая настройка (T+TTA)*	*81,34*	*95,55*	*0,040*
BioCLIP ViT-B/16	*Тонкая настройка (T+TTA)*	*90,99*	*97,16*	*0,011*
CLIP ViT-B/16	*Тонкая настройка (T+TTA)*	*91,90*	*97,43*	*0,023*
ViT-B/16 AugReg	*Тонкая настройка (T+TTA)*	*93,78*	*98,39*	*0,013*
DINOv2 ViT-B/14	*Тонкая настройка (T+TTA)*	*95,34*	*98,55*	*0,010*

Полученные результаты показывают, что тонкая настройка (дообучение), то есть продолжение обучения предварительно обученной модели на собственных данных, является необходимым этапом для достижения высокой точности классификации. Лучший результат в режиме без дополнительного обучения демонстрирует модель BioCLIP (67,99%), однако ее точность уступает на 27 процентных пунктов по сравнению с лучшей дообученной моделью DINOv2 (95,34%). Результаты свидетельствуют о том, что предобучение на общих биологических наборах данных улучшает качество извлекаемых признаков, но не заменяет доменную адаптацию при совместной классификации вида и пола.

Для модели CLIP обучение только линейного классификатора (31,80%) привело к снижению точности по сравнению с режимом zero-shot (47,51%). Данный результат обусловлен различием постановки задачи, в режиме zero-shot классификация выполнялась только на уровне вида (без учета пола), тогда как линейный классификатор обучался различать 41 класс, включая половое разделение. При использовании замороженного пространства признаков CLIP линейный классификатор не разделяет полные классы «вид + пол» без сквозной адаптации параметров модели. Для BioCLIP снижение точности менее выраженное (с 67,99% до 54,10%), что указывает на более информативное признаковое пространство, сформированное в процессе предобучения на биологических изображениях.

Наилучшие результаты среди всех дообученных моделей продемонстрировала модель DINOv2 ViT-B/14, достигшая точности Top-1, равное 95,34% и Top-5, равной 98,55%. За ней

следуют ViT-B/16 AugReg (93,78%) и CLIP ViT-B/16 (91,90%). Модель DINOv2 использует исходное разрешение изображения 518×518 пикселей вместо 224×224 и самонадзорное предобучение, основанное на извлечении плотных признаков, что обеспечивает преимущество при тонкозернистой классификации.

Сверточные нейронные сети (CNN) показывают более низкую точность по сравнению с визуальными трансформерами. Модели ResNet-50 и EfficientNet-B3, несмотря на высокие результаты на наборе данных ImageNet, достигли точности лишь 74,64% и 81,34% соответственно. Это может быть связано с тем, что механизм глобального внимания визуальных трансформеров лучше учитывает тонкие различия в окрасе и структуре оперения, характерные для задач тонкозернистой классификации.

Оценка калибровки модели. Калибровка модели является необходимым условием для практического применения модели в полевых условиях. Охотник, консультирующийся с моделью, ожидает, что заявленный уровень уверенности соответствует фактической точности. Без калибровки исходные оценки уверенности DINOv2 систематически занижены. На рисунке 3 приведен пример классификации самца чирка-свистунка (Anas crecca), где без калибровки (T=1,0) правильному классу присваивается лишь 51,8% вероятности, хотя он лидирует среди всех альтернатив. После использования температурного масштабирования (T=0,494) вероятность правильного класса возрастает до 95,9%, при этом ранжирование классов не меняется, а корректируются только абсолютные значения оценок уверенности. В таблице 4 представлены результаты оценки влияния температурного масштабирования и TTA на показатели качества модели DINOv2 ViT-B/14. Значение параметра температуры $T = 0{,}494$ было подобрано на валидационной выборке.

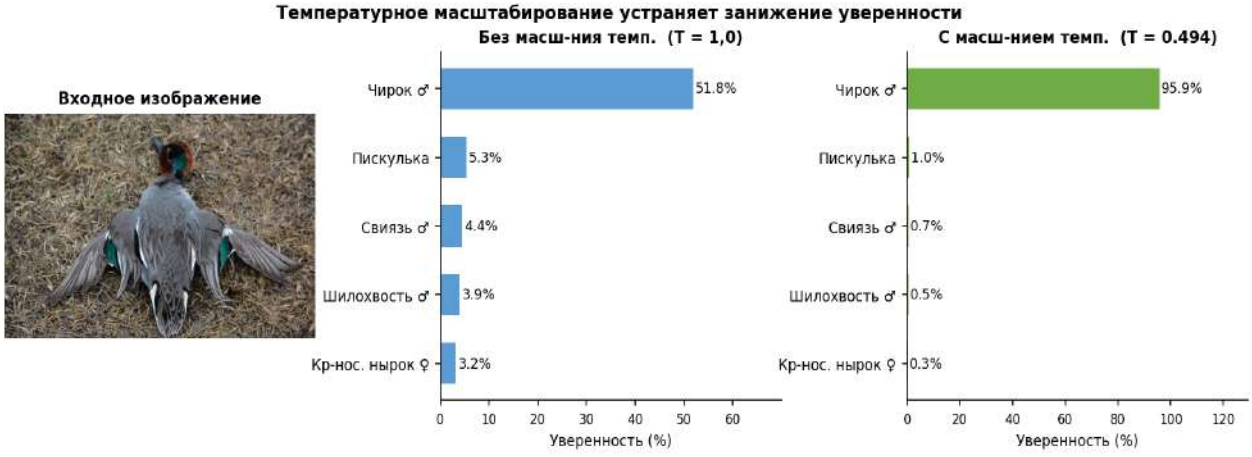


Рисунок 3. Влияние температурного масштабирования на уверенность модели для самца чирка-свистунка (Anas crecca)

Таблица 4. Влияние температурного масштабирования и TTA на DINOv2 ViT-B/14

Конфигурация	*Top-1*	*Top-5*	*ECE*
Базовая ($T = 1{,}0$, без TTA)	*95,17*	*98,61*	*0,308*
+ Темп. ($T = 0{,}494$)	*95,17*	*98,61*	*0,013*
+ TTA×8 ($T = 1{,}0$)	*95,34*	*98,55*	*0,310*
+ Темп. + TTA×8	*95,34*	*98,55*	*0,010*

Температурное масштабирование снижает значение ECE с 0,308 до 0,013 без изменения точности Top-1, что подтверждает корректность ранжирования классов моделью при заниженном уровне уверенности исходных предсказаний. Иными словами, до калибровки заявленная моделью уверенность в среднем расходилась с фактической точностью примерно

на 31 процентный пункт, тогда как после калибровки это расхождение сократилось до 1,3 пункта. Применение TTA повышает точность Top-1 на 0,17 процентного пункта по сравнению с использованием одного варианта изображения. Совместное применение обеих техник дает лучшее значение ECE, равное 0,010, при сохранении прироста точности, полученного за счет TTA.

Анализ по группам видов. В таблице 5 представлены значения точности Top-1 модели DINOv2 ViT-B/14 с использованием температурного масштабирования и аугментации на этапе тестирования (T+TTA) для различных групп видов.

Таблица 5. Точность Top-1 модели DINOv2 ViT-B/14 (T+TTA) по группам видов

Группа видов	*Top-1, %*
Кулики (Numenius, Limosa, Scolopax)	*95–100*
Речные утки, самцы (6 классов)	*93–100*
Речные утки, самки (6 классов)	*79–100*
Речные утки, мономорфные (Anas zonorhyncha)	*95*
Нырковые утки, самцы (6 классов)	*93–100*
Нырковые утки, самки (5 классов)	*80–100*
Нырковые утки, мономорфные (Marmaronetta angustirostris)	*88*
Крохали (Mergus spp., 4 класса)	*91–97*
Пеганки (Tadorna spp., 2 класса)	*100*
Гуси (Anser spp., 4 вида)	*93–100*

Как видно из таблицы 5, наиболее высокие значения точности достигнуты для куликов. Для вальдшнепа (*Scolopax rusticola*) и большого веретенника (*Limosa limosa*) точность достигает 100%, ни для одного вида куликов точность не снижается ниже 95%. Это обусловлено характерными морфологическими признаками этих видов, такими как особенности облика и размеров, а также отсутствием близких визуально сходных классов в наборе данных.

Для полово-диморфных видов точность классификации самцов выше, чем у самок. Самцы уток, как правило, имеют яркое контрастное брачное оперение, тогда как самки большинства видов окрашены в однотонные бурые тона. Это приводит к более низкой и вариабельной точности по классификации самок (79–100%) относительно самцов (91–100%). Наиболее сложными для распознавания оказались самка серой утки (*Mareca strepera*, 79%) и самка красноголового нырка (*Aythya ferina*, 80%), что обусловлено их криптической окраской и высоким сходством с другими видами в пределах своих таксономических групп.

Для гусей основная часть ошибок связана с взаимной путаницей между близкими видами. Точность Top-1 варьируется от 93% для белолобого гуся (*Anser albifrons*) до 96% для серого гуся (*Anser anser*). Чаще ошибки возникают между *A. albifrons* и пискулькой (*A. erythropus*). 5 из 100 тестовых снимков *A. albifrons* были классифицированы как *A. erythropus*. Это связано с тем, что диагностические признаки, такие как размер белого пятна на лбу или наличие желтого глазного кольца, часто неразличимы на исходных фотографиях, что ограничивает точность классификации по визуальным данным. При интерпретации результатов для отдельных классов следует учитывать объем тестовой выборки. Для ряда редких видов тестовая выборка содержит не более пяти изображений, например, для *Sibirionetta formosa*, *Oxyura leucocephala*, самка *Mergellus albellus*. В таких случаях даже одна ошибка классификации может изменить значение точности на 20-100 процентных пунктов. Поэтому диапазоны, приведенные в таблице 5, отражают общую тенденцию, тогда как оценки для редких классов ограничены малым объемом тестовой выборки. Полная нормализованная по строкам матрица ошибок модели DINOv2 ViT-B/14 (TTA×8) на тестовой выборке GameWaterbird-41, содержащей 1 865 изображений, представлена на рисунках 4–6. Поскольку при 41 классе единая матрица становится трудночитаемой в пределах ширины страницы, она разделена на три блока в соответствии с таксономической принадлежностью классов: речные

утки (Рисунок 4), нырковые утки (Рисунок 5), а также крохали, пеганки, гуси, кулики и класс «прочие» (Рисунок 6).

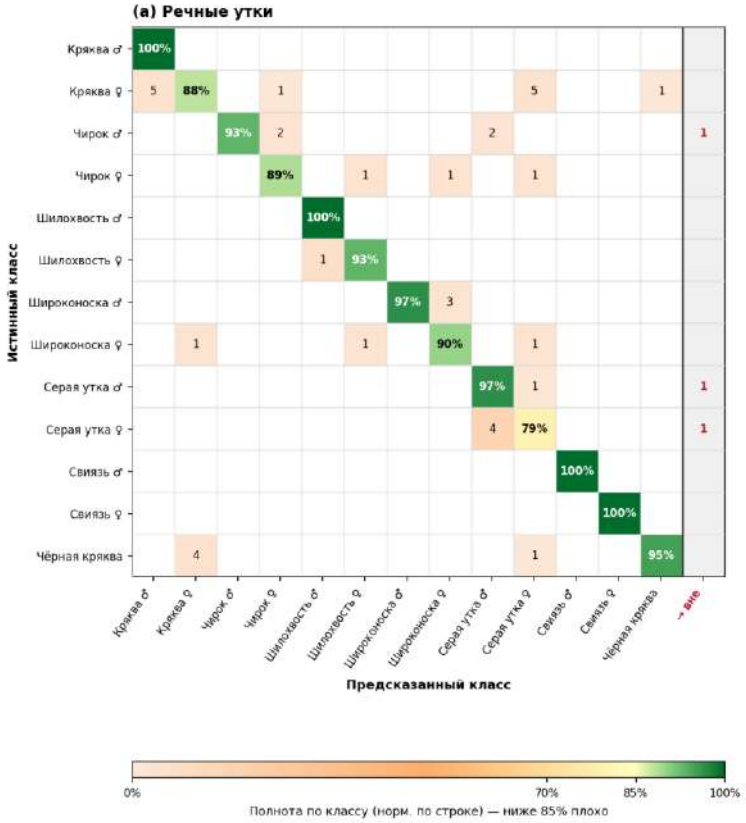


Рисунок 4. Матрица ошибок (блок 1 из 3), речные утки

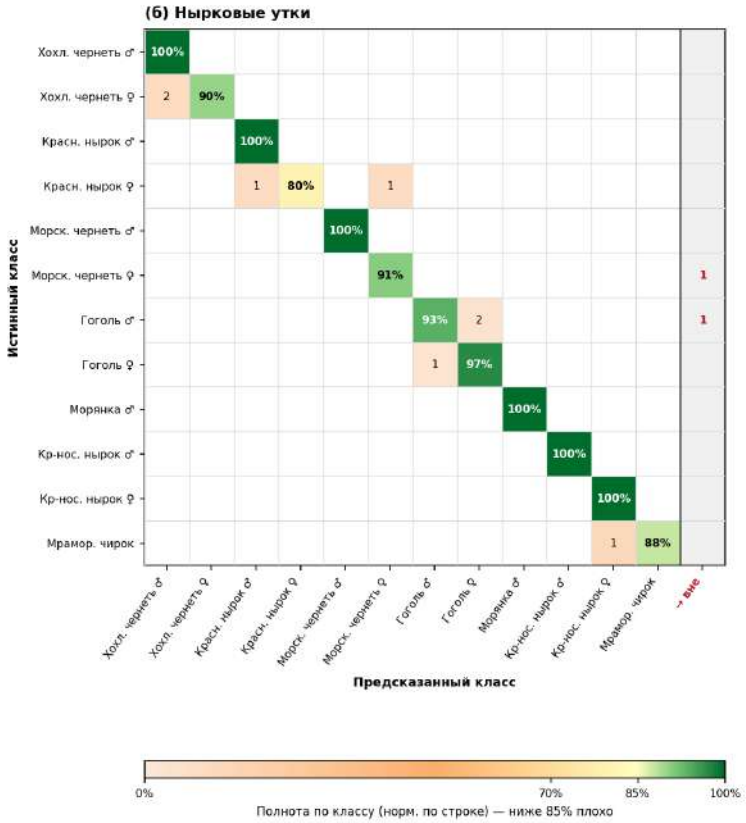


Рисунок 5. Матрица ошибок (блок 2 из 3), нырковые утки

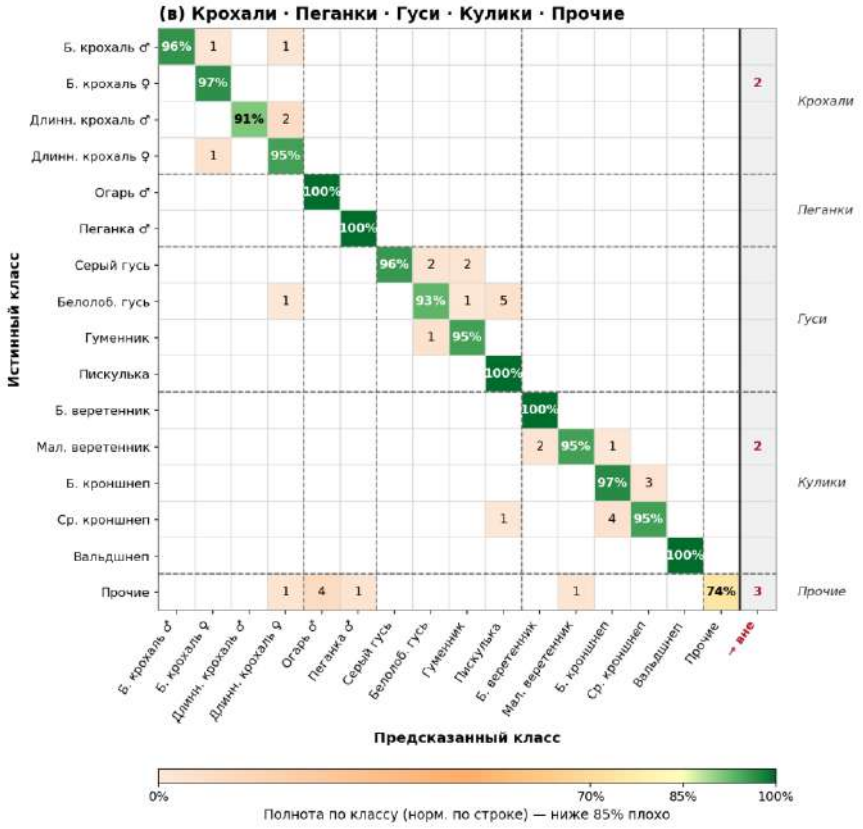


Рисунок 6. Матрица ошибок (блок 3 из 3), крохали, пеганки, гуси, кулики, прочие

На диагонали представлены значения полноты (recall) для каждого класса, вне диагонали количество ошибок классификации (≥1). Белые ячейки соответствуют отсутствию ошибок классификации, а пунктирные линии разделяют таксономические группы внутри рисунка 6. Строки матрицы нормализованы по всем 41 классу, а не внутри блока. Ошибки, выходящие за пределы блока, собраны в отдельном столбце «→ вне», поэтому сумма значений в каждой строке по-прежнему составляет 100%. Доля таких ошибок составляет 12 из 91 (13%), то есть основная часть ошибок приходится на классы внутри одного таксономического блока. Путаница наблюдается внутри группы гусей (Anser albifrons ↔ A. erythropus), а также между самками уток со сходной покровной окраской.

Дискуссия

Полученные результаты показывают, что при решении задачи совместной классификации вида и пола птиц определяющим фактором является не специализация предобучения, а дообучение на целевых данных. Разрыв между лучшим результатом в режиме без дополнительного обучения (BioCLIP, 67,99%) и лучшей дообученной моделью (DINOv2 ViT-B/14, 95,34%) превышает 27 процентных пунктов. При этом BioCLIP, предобученная на биологических изображениях, существенно превосходит CLIP общего назначения в режиме без дообучения (67,99% против 47,51%), тогда как после полной тонкой настройки разрыв между ними сокращается примерно до одного процентного пункта. Наилучший результат показывает модель DINOv2, обученная без разметки и без какой-либо биологической специализации. Это указывает на то, что признаки, необходимые для различения пола и близкородственных видов, извлекаются преимущественно из целевой выборки, а не наследуются из доменного предобучения.

Второй результат касается природы оставшихся ошибок, которые в большей степени связаны с определением пола, чем вида. Полнота для самцов лежит в диапазоне 91–100%,

тогда как для самок 79–100%. Анализ отдельных классов показывает, что самки не только хуже распознаются, но чаще становятся источником ложноположительных классификаций. Так, для самки серой утки полнота составляет 79%, а precision всего лишь 68%, поскольку к данному классу модель ошибочно относит самок кряквы и других речных уток. Это согласуется с биологическими особенностями видов: самки большинства уток обладают покровительственной окраской, тогда как брачное оперение самцов отличается контрастностью и выраженной видоспецифичностью. Для практических задач это означает, что наибольшую сложность представляет определение пола особей птиц с неяркой окраской.

Третий результат позволяет определить пределы применимости предложенного подхода. Для гусей рода Anser ошибки носят межвидовой характер и сосредоточены в паре Anser albifrons и A. erythropus, где 5 из 100 тестовых изображений белолобого гуся ошибочно классифицированы как пискулька. Диагностические признаки этих видов, такие как размер белого пятна на лбу и наличие желтого окологлазного кольца, на любительских фотографиях часто неразличимы. Это указывает на то, что ограничение обусловлено не столько особенностями модели, сколько недостаточной информативностью исходных изображений. Необходимые диагностические признаки отсутствуют или выражены недостаточно, поэтому увеличение объема выборки или усложнение модели не гарантирует устранения подобных ошибок. Схожая закономерность отмечается и в других прикладных исследованиях по распознаванию биологических объектов, где точность, достигнутая на курируемых наборах данных, снижается при переходе к изображениям, полученным в естественных условиях.

Отдельного внимания заслуживает калибровка модели. Сырые выходы дообученных моделей систематически занижают уверенность. Ожидаемая ошибка калибровки составляет 0,308, а температурное масштабирование снижает её до 0,013, без изменения предсказаний. Для рассматриваемой задачи это имеет практическое значение, поскольку решение о добыче принимается по пороговому значению уверенности. Некалиброванная модель не позволяет корректно интерпретировать этот показатель, так как заявленные 50% могут соответствовать и 30%, так и 95% вероятности правильной классификации. Практическая ценность работы состоит не только в достигнутой точности, но и в том, что показатель уверенности становится интерпретируемым и пригодным для построения правил принятия решения.

Ограничения работы определяют направления дальнейших исследований. Изображения получены с платформы iNaturalist и сняты преимущественно в благоприятных условиях, тогда как в полевой обстановке птицы нередко наблюдаются на большом расстоянии, против света или в полете. Для ряда редких классов тестовая выборка содержит менее пяти изображений, что не позволяет делать надежных выводов об их распознавании. Набор данных не различает переходное (эклипсное) оперение, при котором самцы уток внешне сближаются с самками, что характерно для части охотничьего сезона. Наконец, для мономорфных видов определение пола по фотографии невозможно в принципе, поэтому данный случай сознательно выведен за рамки постановки задачи. Дальнейшая работа предполагает расширение набора данных изображениями, полученными в полевых условиях, введение классов эклипсного оперения, а также проверку устойчивости моделей на данных, независимых от платформы iNaturalist.

Заключение

В настоящей работе представлен набор данных GameWaterbird-41, содержащий 18 851 изображение 38 охотничьих видов водоплавающих птиц и куликов, являющихся объектами охоты, объединенных в 41 класс с учетом вида и пола. Также проведен сравнительный анализ шести семейств моделей компьютерного зрения в режимах без дополнительного обучения (zero-shot), линейного классификатора и полной тонкой настройки.

Полученные результаты демонстрируют, что дообученные визуальные трансформеры существенно превосходят CNN-модели и фундаментальные модели, используемые без дополнительного обучения. Наилучший результат показывает модель DINOv2 ViT-B/14, достигнувшая точности 95,34% по показателю Top-1 и 98,55% по показателю Top-5 при

совместном применении температурного масштабирования и аугментации на этапе тестирования (TTA). Обучение с использованием сглаживания меток и функции потерь Focal Loss приводит к заниженной оценке уверенности модели, что эффективно компенсируется температурным масштабированием.

Сложной группой для классификации остаются гуси рода *Anser,* для которых высокое межвидовое сходство ограничивает точность распознавания по фотографическим данным. Дальнейшие исследования будут направлены на расширение набора данных за счет малочисленных классов, а также на систематическую ручную проверку разметки каждого класса, прежде всего для визуально схожих пар видов, аннотации которых наиболее подвержены ошибкам идентификации.

Благодарность

Исследование финансируется Комитетом науки Министерства науки и высшего образования Республики Казахстан в рамках проекта «Разработка цифровой системы управления охотничьими хозяйствами Республики Казахстан с применением методов машинного обучения» (ИРН AP26104615).

Список использованных источников

[1] Wah C., Branson S., Welinder P., Perona P., Belongie S. The Caltech-UCSD Birds-200-2011 Dataset. – 2011. – Technical Report № CNS-TR-2011-001. – 8 p. – URL: https://authors.library.caltech.edu/records/cvm3y-5hh21.

[2] Van Horn G., Branson S., Farrell R., Haber S., Barry J., Ipeirotis P., Perona P., Belongie S. Building a Bird Recognition App and Large Scale Dataset with Citizen Scientists: The Fine Print in Fine-Grained Dataset Collection // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). – 2015. – P. 595–604. – DOI:10.1109/CVPR.2015.7298658.

[3] Республики Казахстан. Законы. Об охране, воспроизводстве и использовании животного мира: нормативные правовые акты: принят Парламентом Республики Казахстан 9 июля 2004 года № 593. – URL: https://adilet.zan.kz/rus/docs/Z040000593_ (дата обращения: 13.07.2026).

[4] Правила охоты: утв. приказом и.о. Министра сельского хозяйства Республики Казахстан от 27 февраля 2015 года № 18-03/157: зарегистрирован в Министерстве юстиции Республики Казахстан 20 мая 2015 года № 11091. – URL: https://adilet.zan.kz/rus/docs/V1500011091 (дата обращения: 13.07.2026).

[5] Налоговый кодекс Республики Казахстан: Кодекс Республики Казахстан от 18 июля 2025 года № 214-VIII ЗРК: по состоянию на 01.07.2026. – URL: https://adilet.zan.kz/rus/docs/K2500000214 (дата обращения: 13.07.2026).

[6] Berg T., Liu J., Lee S.W., Alexander M.L., Jacobs D.W., Belhumeur P.N. Birdsnap: Large-Scale Fine-Grained Visual Categorization of Birds // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). – 2014. – P. 2019–2026. DOI: 10.1109/CVPR.2014.259.

[7] Van Horn G., Cole E., Beery S., Wilber K., Belongie S., Mac Aodha O. Benchmarking Representation Learning for Natural World Image Collections // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). – 2021. – P. 12884–12893. DOI: 10.1109/CVPR46437.2021.01269.

[8] Radford A., Kim J.W., Hallacy C., Ramesh A., Goh G., Agarwal S., Sastry G., Askell A., Mishkin P., Clark J., Krueger G., Sutskever I. Learning Transferable Visual Models from Natural Language Supervision // Proceedings of the 38th International Conference on Machine Learning. – 2021. – Vol. 139. – P. 8748–8763. – URL: https://proceedings.mlr.press/v139/radford21a.html (дата обращения: 13.07.2026). DOI:10.48550/arXiv.2103.00020.

[9] Stevens S., Wu J., Thompson M.J., Campolongo E.G., Song C.H., Carlyn D.E., Dong L., Dahdul W.M., Stewart C., Berger-Wolf T., Chao W.-L., Su Y. BioCLIP: A Vision Foundation Model for the Tree of Life // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). – 2024. – P. 19412–19424. – DOI: 10.1109/CVPR52733.2024.01836.

[10] Oquab M., Darcet T., Moutakanni T., Vo H. V., Szafraniec M., Khalidov V., Fernandez P., Haziza D., Massa F., El-Nouby A., Assran M., Ballas N., Galuba W., Howes R., Huang P.-Y., Li S.- W., Misra I., Rabbat M., Sharma V., Synnaeve G., Xu H., Jégou H., Mairal J., Labatut P., Joulin A., Bojanowski P. DINOv2:

Learning Robust Visual Features without Supervision // Transactions on Machine Learning Research. – 2024. – 32 p. – URL: https://openreview.net/forum?id=a68SUt6zFt (дата обращения: 13.07.2026). DOI: 10.48550/arxiv.2304.07193.

[11] Dosovitskiy A., Beyer L., Kolesnikov A., Weissenborn D., Zhai X., Unterthiner T., Dehghani M., Minderer M., Heigold G., Gelly S., Uszkoreit J., Houlsby N. An Image Is Worth 16×16 Words: Transformers for Image Recognition at Scale // Proceedings of the International Conference on Learning Representations (ICLR). – 2021. – 21 p. –URL: https://openreview.net/pdf?id=YicbFdNTTy (дата обращения: 13.07.2026). DOI:10.48550/arXiv.2010.11929.

[12] Steiner A., Kolesnikov A., Zhai X., Wightman R., Uszkoreit J., Beyer L. How to Train Your ViT? Data, Augmentation, and Regularization in Vision Transformers // Published in Transactions on Machine Learning Research. – 2022. – 16 p. – URL: https://openreview.net/pdf?id=4nPswr1KcP (дата обращения: 13.07.2026). DOI: 10.48550/arXiv.2106.10270.

[13] Guo C., Pleiss G., Sun Y., Weinberger K. Q. On Calibration of Modern Neural Networks // Proceedings of the 34th International Conference on Machine Learning. – 2017. – Vol. 70. – P. 1321–1330.– URL: https://proceedings.mlr.press/v70/guo17a.html (дата обращения: 13.07.2026).

[14] Lin T.-Y., Goyal P., Girshick R., He K., Dollar P. Focal Loss for Dense Object Detection // Proceedings of the IEEE International Conference on Computer Vision. – 2017. – P. 2999–3007. DOI: 10.1109/ICCV.2017.324.

[15] iNaturalist. – URL: https://www.inaturalist.org (дата обращения: 13.07.2026).

[16] Jocher G. Ultralytics YOLO. – URL: https://github.com/ultralytics/ultralytics (дата обращения: 13.07.2026).

[17] He K., Zhang X., Ren S., Sun J. Deep Residual Learning for Image Recognition // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. – 2016. – P. 770–778. DOI: 10.1109/CVPR.2016.90.

[18] Tan M., Le Q.V. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks // Proceedings of the 36th International Conference on Machine Learning. – 2019. – Vol. 97. – P. 6105-6114. URL: https://proceedings.mlr.press/v97/tan19a.html (дата обращения: 17.08.2026).

References

[1] Wah C., Branson S., Welinder P., Perona P., Belongie S. The Caltech-UCSD Birds-200-2011 Dataset. – 2011. – Technical Report № CNS-TR-2011-001. – 8 p. – URL: https://authors.library.caltech.edu/records/cvm3y-5hh21.

[2] Van Horn G., Branson S., Farrell R., Haber S., Barry J., Ipeirotis P., Perona P., Belongie S. Building a Bird Recognition App and Large Scale Dataset with Citizen Scientists: The Fine Print in Fine-Grained Dataset Collection // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). – 2015. – P. 595–604. – DOI:10.1109/CVPR.2015.7298658.

[3] Respublika Kazakhstan. Ob okhrane, vosproizvodstve i ispol'zovanii zhivotnogo mira [On the Protection, Reproduction and Use of Wildlife]: Law of the Republic of Kazakhstan No. 593 dated July 9, 2004. Available at: https://adilet.zan.kz/rus/docs/Z040000593_ (accessed July 13, 2026). (In Russian).

[4] Pravila okhoty [Hunting Rules]: approved by Order of the Acting Minister of Agriculture of the Republic of Kazakhstan No. 18-03/157 dated February 27, 2015; registered with the Ministry of Justice of the Republic of Kazakhstan on May 20, 2015, No. 11091. Available at: https://adilet.zan.kz/rus/docs/V1500011091 (accessed July 13, 2026). (In Russian).

[5] Nalogovyi kodeks Respubliki Kazakhstan [Tax Code of the Republic of Kazakhstan]: Code of the Republic of Kazakhstan No. 214-VIII ZRK dated July 18, 2025, as amended on July 1, 2026. Available at: https://adilet.zan.kz/rus/docs/K2500000214 (accessed July 13, 2026). (In Russian).

[6] Berg T., Liu J., Lee S.W., Alexander M.L., Jacobs D.W., Belhumeur P.N. Birdsnap: Large-Scale Fine-Grained Visual Categorization of Birds // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). – 2014. – P. 2019–2026. DOI: 10.1109/CVPR.2014.259.

[7] Van Horn G., Cole E., Beery S., Wilber K., Belongie S., Mac Aodha O. Benchmarking Representation Learning for Natural World Image Collections // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). – 2021. – P. 12884–12893. DOI: 10.1109/CVPR46437.2021.01269.

[8] Radford A., Kim J.W., Hallacy C., Ramesh A., Goh G., Agarwal S., Sastry G., Askell A., Mishkin P., Clark J., Krueger G., Sutskever I. Learning Transferable Visual Models from Natural Language Supervision

// Proceedings of the 38th International Conference on Machine Learning. – 2021. – Vol. 139. – P. 8748–8763. – URL: https://proceedings.mlr.press/v139/radford21a.html (accessed July 13, 2026). DOI:10.48550/arXiv.2103.00020.

[9] Stevens S., Wu J., Thompson M.J., Campolongo E.G., Song C.H., Carlyn D.E., Dong L., Dahdul W.M., Stewart C., Berger-Wolf T., Chao W.-L., Su Y. BioCLIP: A Vision Foundation Model for the Tree of Life // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). – 2024. – P. 19412–19424. – DOI: 10.1109/CVPR52733.2024.01836.

[10] Oquab M., Darcet T., Moutakanni T., Vo H. V., Szafraniec M., Khalidov V., Fernandez P., Haziza D., Massa F., El-Nouby A., Assran M., Ballas N., Galuba W., Howes R., Huang P.-Y., Li S.- W., Misra I., Rabbat M., Sharma V., Synnaeve G., Xu H., Jégou H., Mairal J., Labatut P., Joulin A., Bojanowski P. DINOv2: Learning Robust Visual Features without Supervision // Transactions on Machine Learning Research. – 2024. – 32 p. – URL: https://openreview.net/forum?id=a68SUt6zFt (accessed July 13, 2026). DOI: 10.48550/arxiv.2304.07193.

[11] Dosovitskiy A., Beyer L., Kolesnikov A., Weissenborn D., Zhai X., Unterthiner T., Dehghani M., Minderer M., Heigold G., Gelly S., Uszkoreit J., Houlsby N. An Image Is Worth 16×16 Words: Transformers for Image Recognition at Scale // Proceedings of the International Conference on Learning Representations (ICLR). – 2021. – 21 p. –URL: https://openreview.net/pdf?id=YicbFdNTTy (accessed July 13, 2026). DOI:10.48550/arXiv.2010.11929.

[12] Steiner A., Kolesnikov A., Zhai X., Wightman R., Uszkoreit J., Beyer L. How to Train Your ViT? Data, Augmentation, and Regularization in Vision Transformers // Published in Transactions on Machine Learning Research. – 2022. – 16 p. – URL: https://openreview.net/pdf?id=4nPswr1KcP (accessed July 13, 2026). DOI: 10.48550/arXiv.2106.10270.

[13] Guo C., Pleiss G., Sun Y., Weinberger K. Q. On Calibration of Modern Neural Networks // Proceedings of the 34th International Conference on Machine Learning. – 2017. – Vol. 70. – P. 1321–1330.– URL: https://proceedings.mlr.press/v70/guo17a.html (accessed July 13, 2026).

[14] Lin T.-Y., Goyal P., Girshick R., He K., Dollar P. Focal Loss for Dense Object Detection // Proceedings of the IEEE International Conference on Computer Vision. – 2017. – P. 2999–3007. DOI: 10.1109/ICCV.2017.324.

[15] iNaturalist. – URL: https://www.inaturalist.org (accessed July 13, 2026).

[16] Jocher G. Ultralytics YOLO. – URL: https://github.com/ultralytics/ultralytics (accessed July 13, 2026).

[17] He K., Zhang X., Ren S., Sun J. Deep Residual Learning for Image Recognition // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. – 2016. – P. 770–778. DOI: 10.1109/CVPR.2016.90.

[18] Tan M., Le Q.V. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks // Proceedings of the 36th International Conference on Machine Learning. – 2019. – Vol. 97. – P. 6105-6114. URL: https://proceedings.mlr.press/v97/tan19a.html (accessed August 17, 2026).

IRSTI 28.23.25

https://doi.org/10.51889/2959-5894.2026.95.3.014

S.Zh. Aibagarov[1*], A.A. Mukhanbet[1]

[1] Al-Farabi Kazakh National University, Almaty, Kazakhstan
[*]*e-mail: awer1307dot@gmail.com*

BENCHMARKING HANDOVER CONTROL IN MULTI-USER 5G NETWORKS

Abstract

Handover management is an important issue in mobile fifth-generation (5G) networks. A controller must decide when each piece of user equipment (UE) should switch its serving base station to maximize throughput while minimizing handovers, ping-pong switching, and connection failures. This paper presents a reproducible simulation environment and a systematic review of handover strategies in a multi-user scenario with four base stations and ten mobile UEs, using a Third Generation Partnership Project (3GPP) Urban Macro channel model with log-normal shadowing. We evaluate four controllers over 100 independent episodes: the 3GPP A3-event algorithm with hysteresis and time-to-trigger (TTT), an ideal Strongest-Cell rule, a No-Handover lower bound, and a learned controller trained with Proximal Policy Optimization (PPO). The results show that the standardized A3-event policy provides the best overall balance. It achieves the highest mean system throughput at 89.1 Mbit/s, the fewest handovers per episode at 21.7, and the lowest ping-pong rate at 0.098 among all methods assessed. The Strongest-Cell rule offers similar throughput but results in about 2.6 times more handovers. The learned controller shows stable training but does not reach the performance of the tuned A3 baseline based on the current reward and observation design. We discuss why the hysteresis-and-TTT logic remains a strong baseline and outline the conditions where a learned policy could outperform it. The entire environment and evaluation process are available for others to reproduce.

Keywords: 5G, handover management, mobility management, A3-event, ping-pong handover, reinforcement learning, benchmark, radio resource management.

С.Ж. Айбагаров[1], А.А. Мұханбет[1]

[1] Әл-Фараби атындағы Қазақ ұлттық университеті, Алматы қ., Қазақстан

КӨП ПАЙДАЛАНУШЫЛЫ 5G ЖЕЛІЛЕРІНДЕ ХЭНДОВЕРДІ БАСҚАРУДЫ САЛЫСТЫРМАЛЫ БАҒАЛАУ

Аңдатпа

Хэндоверді басқару мобильді 5G желілеріндегі негізгі мәселелердің бірі болып табылады: контроллер әрбір пайдаланушы жабдығының (UE) өткізу қабілетін барынша арттыру үшін қай уақытта қызмет көрсетуші базалық станцияны ауыстыруы керектігін анықтауы тиіс, сонымен қатар шамадан тыс хэндоверлерді, ping-pong ауысуларды және байланыс ақауларын болдырмауы қажет. Бұл жұмыста төрт базалық станция және он мобильді UE бар көп пайдаланушылы сценарий үшін лог-нормалды көлеңкеленуі бар 3GPP Urban Macro арна моделі жағдайында хэндовер стратегияларының қайта жаңғыртылатын симуляциялық ортасы және жүйелі бенчмаркі ұсынылады. Біз 100 тәуелсіз эпизод бойынша төрт контроллерді бағалаймыз: гистерезис және іске қосылу уақыты бар 3GPP A3-event алгоритмі, идеалдандырылған Strongest-Cell ережесі, No-Handover төменгі шекарасы және Proximal Policy Optimization (PPO) әдісімен оқытылған контроллер. Нәтижелер стандартталған A3-event саясаты жалпы тепе-теңдік бойынша ең жақсы көрсеткішке қол жеткізетінін көрсетеді: ол бағаланған барлық әдістер арасында ең жоғары орташа жүйелік өткізу қабілетін (89,1 Мбит/с), бір эпизодтағы ең аз хэндовер санын (21,7) және ең төмен ping-pong хэндовер деңгейін (0,098) қамтамасыз етеді. Strongest-Cell ережесі салыстырмалы өткізу қабілетіне жеткенімен, шамамен 2,6 есе көп хэндовер тудырады, ал оқытылған контроллер тұрақты оқытуға қарамастан, қазіргі сыйақы функциясы мен бақылау кеңістігі дизайны жағдайында бапталған A3 базалық стратегиясымен теңесе алмайды. Біз гистерезис және іске қосылу уақыты логикасының неліктен күшті базалық тәсіл болып қалатынын талқылаймыз және оқытылған саясат оны қандай жағдайларда жақсарта алатынын сипаттаймыз. Толық орта мен бағалау конвейері қайта жаңғыртуды қамтамасыз ету үшін жарияланған.

Түйін сөздер: 5G, хэндоверді басқару, мобильділікті басқару, A3-event, ping-pong хэндовер, нығайтумен оқыту, бенчмарк, радиоресурстарды басқару.

С.Ж. Айбагаров[1], А.А. Мұханбет[1]
[1] Казахский национальный университет имени аль-Фараби, г.Алматы, Казахстан

СРАВНИТЕЛЬНАЯ ОЦЕНКА УПРАВЛЕНИЯ ХЭНДОВЕРОМ В МНОГОПОЛЬЗОВАТЕЛЬСКИХ СЕТЯХ

Аннотация

Хэндовер это головная боль любой сети 5G. Контроллеру надо угадать момент, когда абонента (UE) пора передать другой базовой станции: слишком поздно связь просядет, слишком рано и слишком часто начнутся ping-pong-переключения и обрывы. Мы собрали симуляционную среду, которую легко повторить, и прогнали в ней несколько стратегий хэндовера. Сценарий: четыре базовые станции, десять движущихся UE, канал по модели 3GPP Urban Macro с логнормальным затенением. Сравнивали четыре контроллера на 100 независимых эпизодах 3GPP A3-event с гистерезисом и временем срабатывания, идеализированный Strongest-Cell, вариант вообще без хэндоверов и обученный контроллер на PPO. Победил старый добрый A3-event. Он вытянул 89,1 Мбит/с средней пропускной способности, 21,7 хэндовера за эпизод и ping-pong-rate 0,098 лучшие цифры по всем трём метрикам сразу. Strongest-Cell по скорости почти не отстаёт, но переключается примерно в 2,6 раза чаще. А обученный контроллер, хотя учился ровно и без срывов, обойти настроенный A3 так и не смог, по крайней мере с той функцией вознаграждения и тем набором наблюдений, что были у нас. Дальше в статье разбираемся, почему связка гистерезиса и time-to-trigger оказывается такой крепкой базой и что должно измениться, чтобы обученная политика всё-таки вышла вперёд. Всю среду и конвейер оценки мы выложили можно проверить самим.

Ключевые слова: 5G, управление хэндовером, управление мобильностью, A3-event, ping-pong-хэндовер, обучение с подкреплением, бенчмарк, управление радиоресурсами.

Introduction

Key Provisions

This study presents a reproducible simulation environment for comparing handover-control strategies in a multi-user 5G network. Four approaches were evaluated: the 3GPP A3-event algorithm, Strongest-Cell, No-Handover, and a controller trained using Proximal Policy Optimization. The A3-event algorithm achieved the best overall performance, with the highest mean throughput of 89.1 Mbit/s, the lowest number of handovers among the active controllers, and a low ping-pong rate. The results demonstrate that a well-tuned standardized handover mechanism can outperform a reinforcement-learning controller when the latter has limited observations and an insufficiently optimized reward function. Mobility is one of the main characteristics of cellular networks. As a user equipment (UE) moves within the coverage area, the network has to hand over its connection from one base station (BS) to another periodically. This process, called handover, is important for continuity of service in Long-Term Evolution (LTE) and New Radio (NR) cellular systems [1], but it is also costly. Each handover causes a short interruption of the data plane, consumes signaling resources and, if triggered too aggressively, can cause the UE to oscillate between two cells, a pathology known as the ping-pong effect [2], [3]. With the densification of fifth-generation (5G) networks and the reduction of the cell size, the number of handovers increases and the trade-off between the connection quality and the handover overhead becomes more and more delicate [4].

The Third Generation Partnership Project (3GPP) specifies a family of measurement-event-triggered handovers, of which the most commonly deployed is event A3, where handover is triggered if the reference signal received power (RSRP) of a neighbour cell is higher than that of the serving cell for a specified time duration for a given hysteresis [5], [6]. Hysteresis and time-to-trigger (TTT) parameters are used to prevent unnecessary handovers due to rapid fading; however, their fixed values prevent adaptiveness to UE speed and load, and this has stimulated many works on data-driven mobility management [7]. Learning-based methods have been proposed to adapt or substitute handover logic, most notably reinforcement learning (RL) [8].

Following the success of deep reinforcement learning in high-dimensional control tasks [9], deep RL has been applied for learning handover protocol [10], predictive handover for sixth-generation (6G) networks [11], proportional fairness mobility management [12] and traffic steering in Open

Radio Access Network (O-RAN) architectures [13], [14]. Survey studies have cataloged a rapid uptake of learning-based mobility management over LTE and NR [4], and a vast range of prior work targets specific applications such as learning for sliced networks [15], A3-offset and TTT optimisation via Bayesian optimisation and deep reinforcement learning [14], learning of deep deterministic policy gradient (DDPG) parameters [16], based on the DDPG framework [17], and graph-convolutional models for vehicular scenarios [18]. Learning-based handover schemes often report a significant reduction in unwanted handovers and radio link failures [19], [20] and ongoing online learning research attempts to provide guarantees for smooth operation [21]. The vast range of learning techniques have however been evaluated using different simulators, mobility models and metrics, making direct comparison of results problematic and difficult to judge how a learned strategy fares against the default one. This literature often highlights the fundamental trade-off between two key mobility-management failure modes.

Conservative handover parameters (large hysteresis, long TTT) tend to cause high radio link failure due to excessive attachment to weakened cells, whilst overly aggressive settings increase the ping-pong rate and the overall handover overhead [7]. It is possible to navigate this trade-off with A3-event's static configuration parameters, but this leads to static mobility management. Many of the learned solutions in [6, 10, 14, 16] can be seen as attempting to learn a function mapping observations to this trade-off, to provide data-driven, adaptive mobility management. This has been demonstrated with various reported gains, though these are often highly sensitive to the training simulation, mobility model, and the choice of reward signal, with several authors noting that only time-dependent observation and a well-defined churn reward yield gains over simple baselines [10], [21].

This leads to a clear need for benchmarking where a learning approach is directly compared to the default setting on a fair basis. We frame our work around this benchmarking premise. We introduce one comprehensive, reproducible, multi-user environment, run a selection of handover schemes through this benchmark under identical conditions, and conduct a discussion of our results to explain the advantages of both standardized algorithms and learning-based approaches.

Our key contributions are threefold:

1) We propose and implement a fully reproducible, Gymnasium-compatible, multi-user handover benchmark environment, following the OpenAI Gym-style interface for reinforcement-learning environments [22], consisting of four cells, 10 mobile users, a 3GPP Urban Macro pathloss model with log-normal shadowing, and explicit handover interruption time.

2) We benchmark four different handover controllers: the standardized 3GPP A3-event algorithm, an idealized Strongest Cell (SC) algorithm, a No-Handover lower bound, and a Proximal Policy Optimization (PPO)-trained controller, using a collection of 100 independent episodes, measuring five key performance indicators: total system throughput, handover count, ping-pong rate, connection failure rate, and spectral efficiency.

3) We discuss the factors behind the continued performance of the standardized A3-event policy, and the necessary conditions for a learning controller to effectively outperform it. All associated code and trained models have been made publicly available.

Research methodology

System Model

Network and Mobility. The setup is a 5G downlink serving a 1000x1000 m2 area with 4 base stations at (250, 250), (750, 250), (250, 750), and (750, 750) m. 10 UEs traverse randomly using the random-walk model with step sizes between 1-30 m, aiming for a mix of pedestrian and vehicular mobility. A single episode runs for 500 transmission time intervals (TTIs). Handovers result in 10 TTIs of downtime for the affected UE.

Channel Model

The Propagation uses the 3GPP Urban Macro (UMa) Non-line-of-sight path-loss model [23]. For a UE located distance d km away from base station, path loss is calculated as shown in Equation (1):

$$L_p(d) = 128.1 + 37.6\ \log_{10}(d_{km}) \quad [\text{dB}] \tag{1}$$

The RSRP received by UE i from base station j combines path loss with a per-link log-normal shadowing term that is fixed within an episode, as defined in Equation (2):

$$\text{RSRP}_{i,j} = -L_p(d_{i,j}) + X_{i,j}, \qquad X_{i,j} \sim \mathcal{N}(0, \sigma^2),\ \sigma = 8\ \text{dB} \tag{2}$$

where the standard deviation of the shadowing is σ = 8 dB. A UE is declared as a connection failure whenever the RSRP of its serving cell falls below the 3GPP coverage threshold of −110 dBm.

Throughput

Given a serving RSRP and an effective noise floor $N_0 = -100$ dBm, the instantaneous Shannon capacity delivered to UE i over the system bandwidth B = 9 MHz is given by Equation (3):

$$\gamma_i = \frac{\text{RSRP}_{i,b_i} - N_0}{10}, \qquad C_i = B\ \log_2(1 + 10^{\gamma_i}) \tag{3}$$

where b_i denotes the serving base station of UE i and γ_i is the corresponding signal-to-noise ratio in linear scale. The capacity is set to zero for any UE currently undergoing a handover. The system throughput at each step is the sum of the per-UE capacities.

Handover Strategies

We benchmark four controllers, all acting on the same per-step observation of serving and neighbor RSRP, UE speed, and time since the last handover.

A3-event (3GPP standard). Following the 3GPP measurement-event framework, a handover from the serving cell to a neighbor j is triggered when the condition in Equation (4) is satisfied:

$$\text{RSRP}_{i,j} > \text{RSRP}_{i,b_i} + H \quad \text{for } T_{TTT} \text{ consecutive steps} \tag{4}$$

holds continuously for the time-to-trigger window. We use a hysteresis margin H = 3 dB and a time-to-trigger of 40 steps, standard values that suppress fast-fading-induced spurious handovers. When several neighbors qualify, the one with the highest RSRP is selected.

Strongest Cell (idealized). At every step each UE connects to the base station with the instantaneously highest RSRP, with no hysteresis or dwell time. This is an unrealistic upper reference for signal quality that deliberately ignores handover cost.

No Handover (lower bound). Each UE remains attached to its initial cell for the entire episode. This isolates the throughput attainable with zero mobility signaling and serves as a floor.

PPO (learned). This agent trained using Proximal Policy Optimization (PPO) [7], an on-policy actor-critic algorithm, available from Stable-Baselines3 [24]. It takes as input a 50D vector, composed of 5 features for each of the 10 UE and it outputs a MultiDiscrete action determining for each UE whether to stick to the current cell or to handover to one of the other four cells. The agent is rewarded on a combined objective, that consists of the normalized throughput on the positive term, and the handover rate, the number of UEs in interruption, and low connection quality on the negative term, as specified in Equation (5):

$$r = 0.6\,\widehat{C}\ -\ 0.5\,\eta_{HO}\ -\ 0.1\,\phi_{HO}\ +\ 0.1\,Q \tag{5}$$

where Ĉ is normalized system throughput, η_HO the handover rate, φ_HO the fraction of UEs in handover, and Q the mean connection quality. The agent is trained for one million environment steps across three random seeds.

Experimental Setup

All four controllers are evaluated in the exact same environment for 100 independent episodes of 500 steps each; episodes seeds are fixed across all methods so that every controller faces the same mobility and shadowing realizations. We record 5 metrics per method: mean system throughput (Mbit/s), number of handovers per episode, ping-pong rate (fraction of completed handovers that bring a UE back to its previous cell within 100 steps), connection failure rate (fraction of UE-steps with serving RSRP < 110 dBm) and spectral efficiency (bit/s/Hz). PPO controller is trained using a standard Stable-Baselines3 configuration; we average over 3 random seeds for the learning curve.

The environment, trained agents and evaluation scripts are publicly available. These 5 metrics are selected to describe the full mobility management cost structure as opposed to just the system throughput: system throughput and spectral efficiency reflect the utility; handover count the overhead; ping-pong rate the pathological back and forth handovers; and connection failure rate how frequently we leave a UE with a useless RSRP. A method with good system throughput and bad number of handovers or ping pong is operationally useless since every handover consumes a network resource and disrupts the user. A single value is not enough to describe the performance; the combination of these 5 metrics under fixed episode seeds is crucial to show the trade-offs between methods as illustrated in Figures 2 and 3.

Results of the study

Aggregate Comparison

Table 1 reports the mean and standard deviation of each metric over the 100 evaluation episodes. The 3GPP A3-event policy attains the highest mean throughput at 89.1 Mbit/s, exceeding the idealized Strongest-Cell rule by about 11% while triggering only 21.7 handovers per episode - roughly 2.6× fewer than Strongest Cell (56.2). A3-event also yields the lowest ping-pong rate (0.098) and the lowest connection-failure rate (0.427) among the active controllers. The No-Handover bound, by construction, incurs no handovers or ping-pong but sacrifices throughput (71.1 Mbit/s) and suffers frequent connection failures as UEs drift away from their initial cell. The PPO controller, under the present reward and observation design, does not match the tuned A3 baseline on the operational metrics.

Table 1. Handover-management benchmark over 100 evaluation episodes (mean ± std)

Method	*Throughput (Mbit/s)*	*Handovers*	*Ping-pong rate*	*Conn. failure*	*Spec. eff. (b/s/Hz)*
A3-event	*89.1 ± 25.2*	*21.7 ± 4.2*	*0.098 ± 0.070*	*0.427 ± 0.092*	*9.90 ± 2.80*
Strongest Cell	*80.3 ± 23.4*	*56.2 ± 6.2*	*0.325 ± 0.063*	*0.402 ± 0.080*	*8.92 ± 2.60*
No Handover	*71.1 ± 24.9*	*0.0 ± 0.0*	*0.000 ± 0.000*	*0.542 ± 0.102*	*7.90 ± 2.76*
PPO (learned)	*33.3 ± 14.8*	*71.2 ± 7.0*	*0.346 ± 0.072*	*0.602 ± 0.064*	*3.70 ± 1.64*

The throughput, along with 95% confidence intervals, for all four schemes are illustrated in Figure 1. It is visually and statistically clear, the four algorithms can be separated. The ranking, from high throughput to low, is: A3-event, Strongest Cell, No Handover and Learned Controller.

This demonstrates that the edge A3 has over Strongest Cell can be explained. Although they are both aimed at identifying the strong cells for tracking, they approach it in slightly different manner. Hysteresis in A3 is responsible to keep it within a good cell whereas the learned algorithm attempts to anticipate when best to handover between the currently perceived good cells.

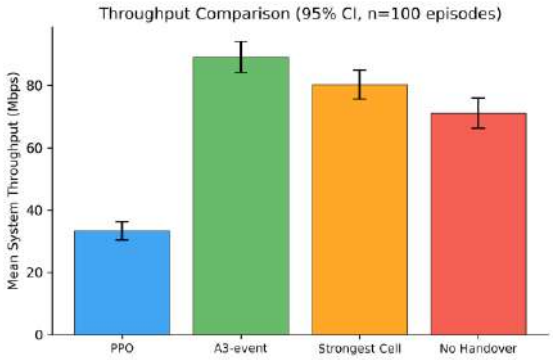


Figure 1. Mean system throughput by method, with 95% confidence intervals over 100 episodes

In Figure 2 we position each method on the throughput and handover plane, with the region we are most interested in being on the top left (good throughput, few handovers). A3-event can be seen on its own in that region, clearly showing that it has the most effective balance between operational parameters. Strongest Cell is at a similar throughput to A3-event but at much higher handover, with no Handover lying on the axis at the lower handover but with considerably lower throughput.

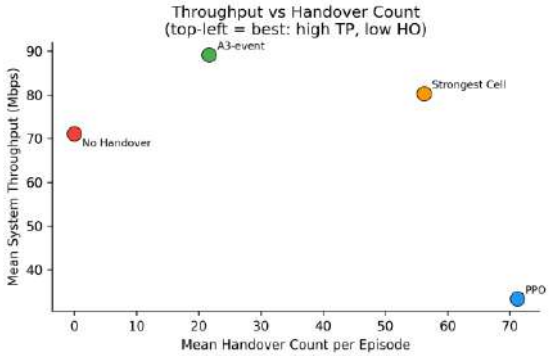


Figure 2. Throughput versus handover count. The upper-left region denotes the best trade-off

Figure 3 compares ping-pong rates. A3-event keeps ping-pong low (0.098) thanks to its dwell-time logic, whereas Strongest Cell, which reacts to every RSRP fluctuation, exhibits a high ping-pong rate (0.325) despite competitive throughput. No Handover trivially has zero ping-pong. This metric isolates the practical value of the hysteresis-and-TTT mechanism that the A3 event encodes.

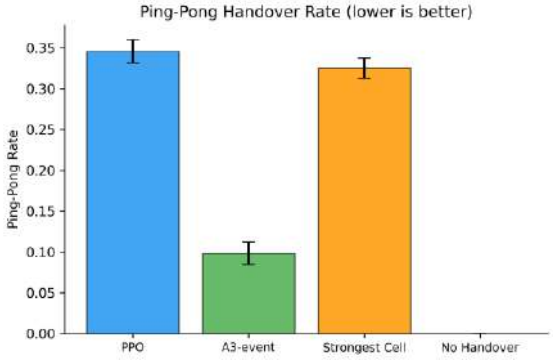


Figure 3. Ping-pong handover rate by method (lower is better)

On the Learned Controller

Figure 4 shows the PPO training curve averaged over three seeds. The mean episode reward rises from roughly −44 to about +23 over one million steps, confirming that the agent does learn to improve its own reward signal and that the poor operational scores are not an artifact of undertraining. The

gap between rising reward and weak operational metrics points instead to the difficulty of the control problem as posed: with a 50-dimensional observation, a joint MultiDiscrete action over ten UEs, and a handover that is penalized both directly and through its interruption cost, the agent converges to a conservative-yet-churning policy that neither matches A3's selectivity nor its timing. This is a known challenge in learning-based mobility management - the reward must be carefully shaped, and partial observability of fast fading limits what an instantaneous policy can achieve [10], [12]. The honest reading of these results is that a well-tuned standardized baseline is hard to beat, and that surpassing it requires either richer temporal observations, a factored or per-UE action design, or explicit reward shaping around the ping-pong and interruption costs.

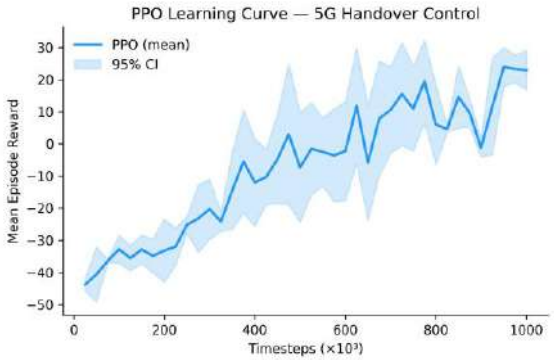


Figure 4. PPO training reward versus timesteps, averaged over three seeds (shaded area: 95% confidence interval)

Discussion

The central, well-supported finding of this benchmark is that the 3GPP A3-event algorithm provides the best throughput-handover balance of the strategies evaluated, combining the highest throughput with the fewest handovers and the lowest ping-pong rate. The comparison with Strongest Cell quantifies precisely what the A3 hysteresis-and-TTT logic buys: nearly the same signal tracking at a fraction of the handover and ping-pong cost. The No-Handover bound shows that some mobility signaling is necessary - staying put degrades both throughput and connection reliability as UEs move. Finally, the learned controller serves as a cautionary data point: reinforcement learning is not automatically superior to a tuned standard, and a fair benchmark must include such a baseline before claiming improvement. These observations align with recent literature reporting that the benefit of learned mobility management depends strongly on observation design and reward engineering [10], [14], [16].

Conclusion

We provided a reproducible benchmark of handover-management strategies for a multi-user 5G network - the 3GPP A3-event algorithm, an idealized Strongest-Cell rule, a No-Handover lower bound, and a PPO-trained controller - in one Gymnasium environment under a 3GPP Urban Macro channel model. The standard A3-event policy achieves the best compromise across 100 episodes of evaluation, exhibiting the best throughput (89.1 Mbit/s), the fewest number of handovers (21.7 per episode) and the smallest rate of ping-pong handovers (0.098). The benchmark reveals the advantages of hysteresis-and-TTT logic and furnishes a realistic baseline on which a learned controller could be evaluated. The environment will be extended with interference-aware signal-to-interference-plus-noise ratio (SINR), realistic correlated mobility traces, more flexible learned controllers (with richer state spaces, factorized actions, and penalties on interruption and ping-pong), to study under which conditions a controller learned policy could overcome the performance of the A3 policy.

Acknowledgment

This research was funded by the Committee of Science of the Ministry of Science and Higher Education of the Republic of Kazakhstan (Grant No.BR24993211).

References

[1] S. Sesia, I. Toufik, and M. Baker, LTE The UMTS Long Term Evolution: From Theory to Practice, 2nd ed. Wiley, 2011. URL: https://onlinelibrary.wiley.com/doi/book/10.1002/9780470978504.

[2] M. Tayyab, X. Gelabert, and R. Jäntti, "A survey on handover management: From LTE to NR," IEEE Access, vol. 7, pp. 118907-118930, 2019. DOI: 10.1109/ACCESS.2019.2937405.

[3] D. López-Pérez, I. Güvenç, and X. Chu, "Mobility management challenges in 3GPP heterogeneous networks," IEEE Communications Magazine, vol. 50, no. 12, pp. 70-78, 2012. DOI: 10.1109/MCOM.2012.6384454.

[4] M. Mollel et al., "A survey of machine learning applications to handover management in 5G and beyond," IEEE Access, vol. 9, pp. 45770-45802, 2021. DOI: 10.1109/ACCESS.2021.3067503.

[5] 3GPP, "NR; Overall description; Stage-2," 3GPP TS 38.300, V18.0.0, 2024. URL: https://www.3gpp.org/ftp/Specs/archive/38_series/38.300/.

[6] 3GPP, "Evolved Universal Terrestrial Radio Access (E-UTRA) and Evolved Universal Terrestrial Radio Access Network (E-UTRAN); Overall description," 3GPP TS 36.300, V17.0.0, 2022. URL: https://www.etsi.org/deliver/etsi_ts/136300_136399/136300/17.00.00_60/ts_136300v170000p.pdf.

[7] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "Proximal policy optimization algorithms," arXiv preprint arXiv:1707.06347, 2017. DOI: 10.48550/arXiv.1707.06347.

[8] R. S. Sutton and A. G. Barto, Reinforcement Learning: An Introduction, 2nd ed. MIT Press, 2018. URL: https://mitpress.mit.edu/9780262039246/reinforcement-learning/.

[9] V. Mnih et al., "Human-level control through deep reinforcement learning," Nature, vol. 518, no. 7540, pp. 529-533, 2015. DOI: 10.1038/nature14236.

[10] C. Lee and J. Schulman et al., "A deep reinforcement learning-based approach for adaptive handover protocols in mobile networks," arXiv preprint arXiv:2401.14823, 2024. DOI: 10.48550/arXiv.2401.14823.

[11] I. Panitsas, et al., "Predictive handover strategy in 6G and beyond: A deep and transfer learning approach," in Proc. IEEE GLOBECOM, 2024. URL: https://arxiv.org/abs/2404.08113.

[12] A. Prado, et al., "Enabling proportionally-fair mobility management with reinforcement learning in 5G networks," IEEE Journal on Selected Areas in Communications, vol. 41, no. 6, pp. 1681-1696, 2023. DOI: 10.1109/JSAC.2023.3273705.

[13] A. Lacava, et al., "Programmable and customized intelligence for traffic steering in 5G networks using Open RAN architectures," IEEE Transactions on Mobile Computing, vol. 23, no. 4, pp. 2882-2897, 2024. DOI: 10.1109/TMC.2023.3266642.

[14] Intelligent handover management enabled by O-RAN and deep reinforcement learning, ResearchGate preprint, 2024. DOI: 10.1109/VTC2024-FALL63153.2024.10757909.

[15] Handover management in a sliced 5G network using deep reinforcement learning, ResearchGate preprint, 2024. DOI: 10.1109/IWCMC61514.2024.10592506.

[16] Advanced handover optimization using deep reinforcement learning in 5G networks, Journal of King Saud University Computer and Information Sciences, 2025. DOI: 10.1007/s44443-025-00124-0.

[17] T. P. Lillicrap et al., "Continuous control with deep reinforcement learning," in Proc. Int. Conf. on Learning Representations (ICLR), 2016. URL: https://arxiv.org/abs/1509.02971.

[18] GCN-based throughput-oriented handover management in dense 5G vehicular networks, arXiv preprint arXiv:2505.04894, 2025. DOI: 10.48550/arXiv.2505.04894.

[19] Hybrid artificial intelligence scheme for vertical handover in heterogeneous networks, in Proc. 8th Int. Conf. on Graphics and Signal Processing (ICGSP), 2024. DOI: 10.1145/3694875.3694884.

[20] Predictive handover mechanism for seamless mobility in 5G and beyond networks, ResearchGate preprint, 2025. DOI: 10.1049/cmu2.12878.

[21] Smooth handovers via smoothed online learning, arXiv preprint arXiv:2501.08099, 2025. DOI: 10.48550/arXiv.2501.08099.

[22] G. Brockman et al., "OpenAI Gym," arXiv preprint arXiv:1606.01540, 2016. DOI: 10.48550/arXiv.1606.01540.

[23] 3GPP, "Study on channel model for frequencies from 0.5 to 100 GHz," 3GPP TR 38.901, V17.0.0, 2022. URL: https://www.3gpp.org/ftp/Specs/archive/38_series/38.901/.

[24] A. Raffin, A. Hill, A. Gleave, A. Kanervisto, M. Ernestus, and N. Dormann, "Stable-Baselines3: Reliable reinforcement learning implementations," Journal of Machine Learning Research, vol. 22, no. 268, pp. 1-8, 2021. URL: https://jmlr.org/papers/v22/20-1364.html.

МРНТИ 28.23.29

https://doi.org/10.51889/2959-5894.2026.95.3.015

О.К. Бегимбетов[1], А.Т. Нурсеитова[2*], Т.Т. Болтайханова[3], К.Б. Тармышов[4], Г.Т. Тургумбаева[5]

[1]РГП «Институт информационных и вычислительных технологий» КН МНВО РК, г.Алматы, Казахстан
[2]ТОО Astana IT University, г.Астана, Казахстан
[3]ТОО ҚазМұнайГаз Инжиниринг, г.Астана, Казахстан
[4]Независимый исследователь, г.Мюнхен, Германия
[5]НАО Карагандинский Индустриальный Университет, г.Темиртау, Казахстан
[*]*e-mail: a.nurseitova@astanait.edu.kz*

СТРАТЕГИЯ МАКСИМИЗАЦИИ ПАССАЖИРООБОРОТА НА СОЦИАЛЬНО ЗНАЧИМЫХ ЖЕЛЕЗНОДОРОЖНЫХ МАРШРУТАХ В УСЛОВИЯХ ДЕТЕРМИНИРОВАННЫХ ТАРИФОВ

Аннотация

В статье рассматриваются методы повышения эффективности использования провозной емкости пассажирских поездов дальнего следования, функционирующих в условиях государственного регулирования тарифов. В отличие от классических систем управления доходами, основанных на динамическом ценообразовании, социально значимые железнодорожные перевозки ограничены фиксированными тарифами и не могут использовать ценовые механизмы для регулирования спроса. В этих условиях основным инструментом повышения эффективности становится интеллектуальное управление распределением мест между различными корреспонденциями Origin–Destination (O&D). Предлагается комплексная методология управления емкостью, основанная на классификации поездов по уровню населенности, сегментации маршрутов по их вкладу в пассажирооборот, применении виртуальной вложенности (Virtual Nesting), восстановлении скрытого спроса и технологии маршрутного сшивания (Route Splicing). Разработаны алгоритмы управления для пяти категорий поездов в зависимости от прогнозируемого коэффициента населенности. Проведен эмпирический анализ на реальных данных АО «Пассажирские перевозки» (Казахстан) за период 2022–2025 гг. и имитационное моделирование, демонстрирующее потенциальное увеличение пассажирооборота на 6–12% по сравнению с базовым методом статического квотирования. Показано, что максимизация пассажирооборота может позволить повысить эффективность использования субсидируемых перевозок без изменения тарифной политики.

Ключевые слова: пассажирские перевозки, железнодорожный транспорт, управление доходами, пассажирооборот, O&D-модель, провозная емкость, Virtual Nesting, Route Splicing, социально значимые маршруты.

О.К Бегимбетов[1], А.Т. Нурсеитова[2*], Т.Т. Болтайханова[3], К.Б. Тармышов[4], Г.Т. Тургумбаева[5]
[1]ҚР ҒЖБМ ҒК Ақпараттық және есептеуіш технологиялар институты, Алматы, Қазақстан
[2]Astana IT University ЖШС, Астана қ,, Қазақстан
[3]ҚазМұнайГаз Инжиниринг ЖШС, Астана қ., Қазақстан
[4]Тәуелсіз зерттеуші, Мюнхен қ,, Германия
[5]Қарағанды индустриялық университеті КЕАҚ, Теміртау қ,, Қазақстан

БЕЛГІЛЕНГЕН ТАРИФТЕР ЖАҒДАЙЫНДА ӘЛЕУМЕТТІК МАҢЫЗЫ БАР ТЕМІРЖОЛ БАҒЫТТАРЫНДАҒЫ ЖОЛАУШЫЛАР АЙНАЛЫМЫН БАРЫНША МАКСИМИЗАЦИЯЛАУ СТРАТЕГИЯСЫ

Аңдатпа

Мақалада мемлекеттік тарифтік реттеу жағдайында жұмыс істейтін алыс қашықтыққа қатынайтын жолаушылар пойыздарының тасымалдау сыйымдылығын пайдалану тиімділігін арттыру әдістері қарастырылады. Динамикалық баға белгілеуге негізделген дәстүрлі кірістерді басқару жүйелерінен

(Revenue Management) айырмашылығы – әлеуметтік маңызы бар теміржол тасымалдары белгіленген (бекітілген) тарифтермен шектеледі, сондықтан сұранысты реттеу үшін баға механизмдерін қолдана алмайды. Бұл жағдайда тиімділікті арттырудың негізгі құралы – түрлі жөнелту-межелі пункттер (Origin–Destination / O&D) арасында орындарды бөлуді интеллектуалды басқару болып табылады. Жұмыста пойыздарды толығу деңгейі бойынша жіктеуге, бағыттарды олардың жолаушылар айналымына қосатын үлесіне қарай сегменттеуге, виртуальды ұяшықтау (Virtual Nesting) әдісін қолдануға, жасырын сұранысты қалпына келтіруге және бағыттарды түйістіру (Route Splicing) технологиясына негізделген сыйымдылықты басқарудың кешенді әдістемесі ұсынылады. Жоспарланатын толығу коэффициентіне байланысты пойыздардың бес санаты үшін басқару алгоритмдері әзірленді. «Жолаушылар тасымалы» АҚ-ның (Қазақстан) 2022–2025 жылдар аралығындағы нақты деректері негізінде эмпирикалық талдау және имитациялық модельдеу жүргізілді. Нәтижесінде, статикалық квоталаудың базалық әдісімен салыстырғанда, жолаушылар айналымын 6–12%-ға арттырудың потенциалды мүмкіндігі көрсетілді. Жолаушылар айналымын барынша арттыру (максимизациялау) тарифтік саясатты өзгертпей-ақ, субсидияланатын тасымалдарды пайдалану тиімділігін жоғарылатуға мүмкіндік беретіні дәлелденді.

Түйін сөздер: жолаушылар тасымалы, теміржол көлігі, кірістерді басқару, жолаушылар айналымы, O&D-модель, тасымалдау сыйымдылығы, виртуальды ұяшықтау (Virtual Nesting), бағыттарды түйістіру (Route Splicing), әлеуметтік маңызы бар бағыттар.

O.K. Begimbetov[1], A.T. Nurseitova[2], T.T. Boltaikhanova[3], K.B. Tarmyshov[4] , G.T. Turgumbayeva[5]
[1]International Information Technology University, Almaty, Kazakhstan, Almaty, Kazakhstan
[2]Astana IT University LLC, Astana, Kazakhstan
[3]KazMunayGas Engeneering LLP, Astana, Kazakhstan
[4]Independent Researcher, Munich, Germany
[5]Karaganda Industrial University NJSC, Temirtau, Kazakhstan

PASSENGER TURNOVER MAXIMIZATION STRATEGY FOR SOCIALLY SIGNIFICANT RAILWAY ROUTES UNDER DETERMINISTIC TARIFFS

Abstract

This article examines methods for enhancing the capacity utilization efficiency of long-distance passenger trains operating under state-regulated tariffs. Unlike traditional revenue management systems based on dynamic pricing, socially significant passenger rail services are constrained by fixed tariffs and cannot employ price mechanisms to regulate demand. Under these conditions, intelligent management of seat allocation across various Origin–Destination (O&D) city pairs becomes the primary instrument for increasing operational efficiency. We propose a comprehensive capacity management methodology based on train classification by occupancy level, route segmentation by contribution to passenger turnover, application of Virtual Nesting, unconstrained demand estimation, and Route Splicing technology. Control algorithms were developed for five train categories based on predicted load factors. An empirical analysis using real-world operational data from JSC "Passenger Transportation" (Kazakhstan) for the 2022–2025 period, combined with simulation modeling, demonstrates a potential 6–12% increase in passenger turnover compared to the baseline static bucket allocation method. The findings show that maximizing passenger turnover can improve the efficiency of subsidized rail operations without altering existing tariff policies.

Keywords: passenger transportation, railway transport, revenue management, passenger turnover, O&D model, carrying capacity, Virtual Nesting, Route Splicing, socially significant routes.

Введение

Управление доходами (Revenue Management) является одним из ключевых инструментов повышения эффективности транспортных компаний. Концепция управления доходами исторически сформировалась в авиационной индустрии и базировалась на трех фундаментальных столпах: динамическом ценообразовании, распределении емкости и структурных решениях сети [1]. Однако экстраполяция классических моделей на сектор пассажирских железнодорожных перевозок требует существенной концептуальной адаптации, особенно когда речь идет о социально значимых перевозках, субсидируемых государством в рамках обязательств по общественному обслуживанию (Public Service

Obligation, PSO) [2]. В подобных условиях железнодорожный оператор оперирует в парадигме жестко зафиксированного тарифа, который не может выступать инструментом регулирования спроса [3]. В контексте протяженных маршрутов (до 50 станций и времени в пути до 48 часов), характерных для географии Казахстана, ключевой операционной проблемой становится фрагментация провозной емкости. Железнодорожный транспорт обладает уникальной сетевой топологией: один рейс представляет собой сложную матрицу возможных корреспонденций Origin & Destination (O&D). В поезде, проходящем 50 станций, формируется теоретическое множество из 1225 возможных пар станций отправления и назначения. Каждая такая пара представляет собой отдельную корреспонденцию Origin–Destination, конкурирующую за ограниченную провозную емкость поезда. Пассажир, приобретающий билет на короткий промежуточный отрезок, физически изымает место из инвентарной базы на всем протяжении маршрута поезда. Если спрос на дальние поездки, пересекающие данный короткий участок, высок, оператор сталкивается с феноменом вытеснения или упущенного спроса (Spill). Одновременно на участках до посадки локального пассажира и после его высадки возникает эффект потери емкости (Spoilage) — места следуют абсолютно пустыми из-за искусственной блокировки сквозного проезда [4].

В данной работе проверяются следующие гипотезы:

1. При фиксированных тарифах географическая вложенность эффективнее временных квот.
2. Восстановление нецензурированного спроса значимо повышает пассажирооборот (RPK).
3. Технология маршрутного сшивания эффективна при уровне загрузки Network Load Factor (NLF– сетевой коэффициент населенности) > 80%.

Методология исследования

Международный опыт управления провозной емкостью

Для формирования жизнеспособной модели оптимизации в условиях ручного управления необходимо детально проанализировать эволюцию систем резервирования.

Специфика и ограничения наследия СНГ: Казахстан и Россия. Железнодорожные администрации государств-участников Содружества Независимых Государств (СНГ), включая Казахстан и Российскую Федерацию, исторически оперируют в едином технологическом стандарте, заданном регламентами Организации сотрудничества железных дорог (ОСЖД) и межправительственными соглашениями [5]. Технологическим ядром управления пассажирскими перевозками выступают автоматизированные системы управления (АСУ) семейства «Экспресс». Традиционная парадигма управления емкостью в этих системах базируется на механизме так называемых «станционных трафаретов» или жестких физических квот. Суть этого подхода заключается в априорном разделении общей вместимости поезда на фиксированные блоки, которые жестко закрепляются за крупными узловыми станциями по маршруту следования [5]. Фундаментальный недостаток системы жесткого квотирования заключается в ее неспособности адаптироваться к стохастической природе спроса. Если спрос на станции отправления превышает квоту, пассажиры получают отказ (Spill), в то время как квота промежуточной станции может оставаться невыкупленной. В результате эти места следуют пустыми от начальной станции до транзитной (Spoilage) [4]. Хотя современные итерации систем бронирования в СНГ внедряют элементы логической доступности и переменного трафарета, управление сложными социально значимыми маршрутами с 40–50 остановками по-прежнему требует интенсивного ручного вмешательства аналитиков [5].

Опыт операторов Европейского Союза. Европейский рынок пассажирских перевозок представляет собой сложный конгломерат коммерческих (дерегулированных) и социально значимых (субсидируемых) маршрутов. Как показывает анализ шведского опыта, на дерегулированных рынках распределение пропускной способности инфраструктуры между операторами может базироваться на аукционных механизмах и оценке готовности платить, однако для субсидируемых сегментов основным критерием остается социальный анализ

затрат и выгод [6]. Оценка экономической ценности конкретного поезда или маршрута представляет значительную сложность даже для самой железнодорожной компании [7]. Наиболее релевантным примером технологического преодоления проблемы фиксированных тарифов является внедрение Системы управления доходами (Yield Management System) итальянским национальным оператором Trenitalia в сотрудничестве с корпорацией IBM [8]. Система, управляющая инвентарем для более чем 10 миллионов пассажиров и обрабатывающая свыше 120 миллиардов решений об аллокации мест, базируется на двухэтапной стохастической модели [8]. Ввиду наличия кластерной структуры фиксированных семейств тарифов система Trenitalia не может бесконечно варьировать цены. Вместо этого она прогнозирует неконсолидированный (unconstrained) спрос с использованием аддитивных методов коррекции и оптимизирует распределение провозной емкости по каждой паре O&D на протяжении всего горизонта бронирования. Важнейшим элементом является использование техники частичной вложенности (partial nesting) O&D [8].

Североамериканский опыт. Американский оператор дальнего следования Amtrak оперирует в условиях, схожих с казахстанскими: огромные расстояния, множество промежуточных остановок и жесткая конкуренция с другими видами транспорта. Управление инвентарем в Amtrak сфокусировано на минимизации двух упомянутых ранее рисков: потери емкости и упущенного спроса [4]. 3Ключевым аналитическим сдвигом в практике Amtrak стал отказ от использования традиционного коэффициента загрузки (Load Factor, процент проданных мест) в качестве главной метрики успеха. Данный показатель признан обманчивым, так как полная посадка на коротком плече маскирует пустоту на остальном протяжении маршрута. Вместо этого фокус смещен на метрику T-RASM (Ticket Revenue per Available Seat-Mile) – выручку от билетов на доступную кресло-милю [4]. В контексте неизменного тарифа, привязанного к расстоянию, максимизация T-RASM математически эквивалентна максимизации произведенных пассажиро-километров [4].

Азиатский вектор. Опыт японских высокоскоростных железных дорог (Shinkansen) демонстрирует альтернативный подход, при котором доходность максимизируется не только управлением продажами, но и управлением самой физической емкостью. В отличие от авиации, где размер самолета фиксирован, железнодорожная отрасль позволяет добавлять емкость дискретными «кусками» (вагонами) в зависимости от кривой накопления спроса [9]. Исследования показывают, что совместная оптимизация ценообразования и пропускной способности имеет принципиальное значение. В условиях, когда изменение цен невозможно, японские операторы применяют политику гибкой емкости (flexible capacity), адаптируя длину состава под прогнозируемые транзитные потоки [1]. Разработанные математические модели показывают, что такие стратегии асимптотически оптимальны и масштабируются пропорционально росту спроса [9].

Прогнозирование спроса на железнодорожные билеты в Казахстане. Важным направлением исследований в области управления пассажирскими перевозками является прогнозирование спроса на железнодорожные билеты. В работе [10] проведен сравнительный анализ методов прогнозирования спроса на основе исторических данных о продажах билетов в Казахстане за период 2021–2023 гг. Исследование охватывает 55 корреспонденций Origin-Destination (O&D) и использует данные, предоставленные билетным агентом Nur-Kassa LLP. Авторы рассматривают четыре класса моделей: ARIMA (AutoRegressive Integrated Moving Average) – классический статистический метод для прогнозирования временных рядов; SARIMAX(Seasonal ARIMA with eXogenousvariables) – расширение ARIMA, учитывающее сезонность и внешние факторы; LSTM (Long Short-Term Memory) – рекуррентная нейронная сеть, способная моделировать долгосрочные зависимости в данных; CNN (Convolutional Neural Network) — сверточная нейронная сеть, применяемая для прогнозирования всей O&D-матрицы как единого 3D-массива, где каждый элемент представляет пассажиропоток между станцией отправления и станцией назначения. Ключевой результат исследования заключается в том, что CNN-модель, учитывающая пространственные взаимосвязи между различными O&D-парами (общие маршруты, транзитные узлы), демонстрирует наилучшую точность прогнозирования. Значения средней абсолютной ошибки (MAE) для CNN составили 3.583 в первом тестовом сценарии (ноябрь–декабрь 2022 г.) и 3.111 во втором (сентябрь 2022 г. – август 2023 г.), что существенно ниже показателей ARIMA, SARIMAX и LSTM [10]. Полученные результаты имеют прямое отношение к методологии, предлагаемой в настоящей статье. Во-первых, прогнозирование спроса на уровне отдельных O&D пар является необходимым условием для установления уровней защиты (Protection

Levels) в рамках виртуальной вложенности (Virtual Nesting). Во-вторых, способность CNN-модели учитывать взаимосвязи между корреспонденциями открывает возможности для построения более точных прогнозов секционной населенности (SLF) и, следовательно, более эффективного применения технологии сшивания маршрутов (Route Splicing). В-третьих, горизонтальная ось прогнозирования (количество дней до отправления) в работе [10] коррелирует с хронологией управленческих решений (D-45, D-30, D-14, D-7, D-3, D-1). Здесь и далее нотация D-N означает количество дней до даты отправления поезда: например, D-45 соответствует 45-му дню до отправления, а D-0 — дню отправления. Это позволяет интегрировать прогнозные модели в разработанный алгоритм управления провозной емкостью.

Концепция максимизации пассажирооборота

При фиксированных тарифах эффективность работы поезда определяется не количеством перевезенных пассажиров, а объемом выполненной транспортной работы.

Целевая функция. В качестве основной целевой функции предлагается использовать показатель пассажирооборота (Revenue Passenger Kilometers, RPK)[4]:

$$RPK = \sum_{i=1}^{n} Pax_i \cdot Distance_i \qquad (1)$$

где Pax_i – количество пассажиров по *i*-й корреспонденции, $Distance_i$ – расстояние поездки.

Максимизация показателя *RPK* является главной целью управления.

Доступная провозная емкость. Доступная провозная емкость (Available Seat Kilometers, ASK) представляет собой меру предложенного рынку объема услуг с учетом физической длины маршрута [4]:

$$ASK = TotalSeats \times TotalRouteDistance, \qquad (2)$$

где *Total Seats* — общая вместимость поезда (количество мест), *TotalRouteDistance* — протяжённость всего маршрута (км).

Сетевой коэффициент населенности. Сетевой коэффициент населенности (NLF) отражает пространственную утилизацию поезда в отличие от обманчивого процента проданных билетов [4]:

$$NLF = \frac{RPK}{ASK}, \qquad (3)$$

где *RPK* и *ASK* — показатели, определённые формулами (1) и (2) соответственно.

Секционная населенность. Для анализа распределения пассажиропотока по отдельным участкам маршрута используется показатель секционной населенности (Sectional Load Factor, SLF). В отличие от NLF, характеризующего использование провозной емкости на протяжении всего маршрута, показатель SLF отражает степень заполнения поезда на каждом отдельном перегоне между соседними станциями.

SLF для каждого перегона рассчитывается как отношение количества пассажиров, следующих по данному перегону, к общей вместимости поезда:

$$SLF_{leg} = \frac{Pax_{on_leg}}{Seats_{total}}, \qquad (4)$$

где Pax_{on_leg} – количество пассажиров, следующих через данный перегон,

$Seats_{total}$ – общая вместимость поезда.

Классификация корреспонденций.

Для управления спросом все корреспонденции предлагается разделять на три группы на основе их вклада в общий пассажирооборот (Таблица 1).

Таблица 1. Классификация O&D корреспонденций

Категория	*Определение*	*Стратегическое значение*
Магистральные (Long-haul O&D)	*>75% длины маршрута*	*Фундаментальное ядро генерации RPK, наивысший приоритет, устанавливаются жесткие уровни защиты (Protection Levels)*
Среднемагистральные (Mid-haul O&D	*35%–75% длины маршрута*	*Формируют базовую экономическую загрузку поезда*
Локальные (Shorthaul O&D)	*<35% длины маршрута*	*Критическая угроза для NLF, подлежат жесткому квотированию на ранних этапах продаж*

Обоснование порогов классификации рейсов. Пороги классификации рейсов по NLF (50%, 65%, 80%, 90%) являются экспертными значениями, основанными на следующем эмпирическом анализе исторических данных:

1. Анализ распределения NLF: на основе данных по 2190 рейсам за 2022–2025 гг. построено распределение NLF, выявившее естественные кластеры в окрестностях указанных порогов. Распределение имеет колоколообразную форму с пиком в диапазоне 65–80%.

2. Анализ кривых бронирования: Для каждого порогового значения проанализирована форма кривой накопления продаж. Рейсы с NLF < 50% демонстрируют аномально низкие темпы продаж на всем горизонте бронирования. Рейсы с NLF > 90% характеризуются практически линейным выкупом с ранним насыщением.

3. Сопоставление с международной практикой: Используемые пороги согласуются с классификациями, применяемыми в системах управления доходами Amtrak и Trenitalia [4, 8].

4. Практическая управляемость: Количество категорий (5) и их границы выбраны так, чтобы обеспечить различный набор управляющих воздействий для каждой категории, избегая избыточной детализации.

Методология управления емкостью

Виртуальная вложенность (Virtual Nesting). Основой предлагаемого подхода является концепция виртуальной вложенности, адаптированная к условиям фиксированных тарифов. При управлении поездами аналитик должен реализовывать логику «ограждения» инвентаря (fare fencing), переложенную с ценовых классов на географические расстояния [8]. Магистральные плечи формируют высший Класс 1. Среднемагистральные отнесены к Классу 2. Локальные образуют Класс 3. Принцип виртуальной вложенности диктует иерархию доступа: провозная емкость, выделенная для Класса 3, также прозрачно доступна для Классов 1 и 2. Однако емкость, защищенная для Класса 1 (Protection Level), абсолютно недоступна для пассажиров Классов 2 и 3 [8].

Пример распределения мест по виртуальным классам бронирования и соответствующая динамика емкостей приведены в таблице 2.

Таблица 2. Пример распределения мест при виртуальной вложенности

Класс	*Базовый пул*	*Защищенный уровень*	*Доступность для классов*
Класс 1 (Long-haul)	*200*	*80*	*Только Класс 1*
Класс 2 (Mid-haul)	*180*	*40*	*Классы 1, 2*
Класс 3 (Short-haul)	*120*	*-*	*Классы 1, 2, 3*

В условиях устаревших систем типа «Экспресс» эта логика эмулируется через физическое выделение определенных вагонов исключительно под транзитный поток.

Адаптация эвристики ожидаемого предельного пассажирооборота. В классической модели EMSR (Expected Marginal Seat Revenue) сравниваются доходы. В нашей модели сравниваются ожидаемые приросты пассажиро-километров:

$$E(\Delta RPK_{long}) = Distance_{long} \times P(D_{long} > x) \quad (5)$$

где $Distance_{long}$ – физическая длина магистрального маршрута, $P(D_{long} > x)$ – вероятность того, что спрос на магистральный маршрут превысит количество уже защищенных мест x, D_{long} – случайная величина спроса на магистральный маршрут. Продажа на локальное плечо должна быть заблокирована, пока справедливо неравенство:

$$Distance_{short} < Distance_{long} \times P(D_{long} > x) \quad (6)$$

где $Distance_{short}$ – физическая длина локального (short-haul) участка маршрута.

Поскольку длина полного маршрута в поездах на 40–50 станций часто в 15–20 раз превосходит локальный перегон, математика неумолимо требует жесткой защиты мест для дальних поездок даже при крайне низкой вероятности их реализации [8].

Восстановление истинного спроса. Фундаментальной ошибкой аналитиков является отождествление исторических данных о продажах с реальным историческим спросом. Если в прошлом году все места в поезде были проданы, система покажет спрос, равный вместимости, тогда как реальный спрос мог быть существенно выше. Попытка базировать уровни защиты на цензурированных (обрезанных вместимостью) данных приводит к занижению квот для дальних пассажиров и эффекту нисходящей спирали (spiral-down effect) [8]. Для корректной работы необходимо восстанавливать неконсолидированный спрос с использованием методов децензурирования: мониторинг отказов в системе резервирования, анализ кривой бронирования (Booking Curve), экстраполяция темпов продаж до даты отправления [8].

Технология маршрутного сшивания (Route Splicing). Одной из самых продвинутых техник ручного управления при высоком NLF является контроль идеальных стыковок. Если аналитический мониторинг показывает формирование устойчивого разрыва спроса на средней станции маршрута, аналитик применяет сшивание. На маршруте от станции A до станции Z при исторически сильном спросе на участках A–M и M–Z аналитик разрешает продажу комбинации A–M и M–Z как двух независимых бронирований на одно физическое место. Количественный пример работы технологии маршрутного сшивания: рассмотрим поезд вместимостью 500 мест на маршруте A–Z (1190 км) с точкой разрыва M на 600 км от A. Тогда расчёт RPK для базового сценария и после сшивания будут соответственно:

RPK = 200 × 1190 + 50 × 600 + 30 × 590 = 238000 + 30000 + 17700 = 285700
RPK* = 200 × 1190 + 100 × 600 + 80 × 590 = 238000 + 60000 + 47200 = 345200

Результаты расчета эффекта от внедрения технологии маршрутного сшивания приведены в таблице 3.

Таблица 3. Пример расчета эффекта технологии маршрутного сшивания

Показатель	*Базовый сценарий*	*После сшивания*	*Изменение*
Пассажиры A–Z	*200*	*200*	*0*
Пассажиры A–M	*50*	*100*	*+50*
Пассажиры M–Z	*30*	*80*	*+50*
Итого пассажиров	*280*	*380*	*+100*
RPK, км	*285 700*	*345 200*	*+59 500*

Прирост RPK составляет 20.8% при сохранении неизменной вместимости поезда.

Алгоритм управления. Предлагаемый алгоритм управления провозной емкостью представлен в виде псевдокода на алгоритме 1.

Алгоритм 1. Алгоритм управления провозной емкостью

Algorithm 1 Алгоритм управления провозной емкостью

Require: *train* — информация о поезде, *forecast_nlf* — прогноз NLF
Ensure: Распределение мест по O&D-классам
category ← *ClassifyByNLF*(*forecast_nlf*)
protection_levels ← ∅
splicing_candidates ← ∅
if *category* = 1 **then** ▷ NLF < 50%
    Разрешить продажу на все O&D без ограничений
    Рассмотреть отцепку вагонов
else if *category* = 2 **then** ▷ 50% ≤ NLF < 65%
    Блокировать продажи на участки < 3 станций в радиусе 100 км от крупных городов
    Построить SLF-профиль на D-20
    Закрыть локальные O&D на дефицитных перегонах на D-10
else if *category* = 3 **then** ▷ 65% ≤ NLF < 80%
    protection_levels ← *CalculateProtectionLevels*(*forecast_demand*)
    Применить виртуальную вложенность
    if *booking_pace* > *historical_norm* на D-25 **then**
        Расширить пул защиты
    end if
    Сбросить нереализованные остатки на D-4..D-1
else if *category* = 4 **then** ▷ 80% ≤ NLF < 90%
    Блокировать все локальные перегоны (< 35% маршрута)
    protection_levels ← *CalculateProtectionLevels*(*forecast_demand*)
    Открыть изолированные перегоны на D-30
    Постепенно открывать mid-haul на D-9..D-6
else if *category* = 5 **then** ▷ NLF ≥ 90%
    Запросить дополнительные вагоны на D-45
    Установить Protection Level 75–80% для long-haul
    splicing_candidates ← *FindSplicingCandidates*(*SLF_profile*)
    Применить технологию маршрутного сшивания на D-14
end if
return (*protection_levels*, *splicing_candidates*)

Дифференцированное управление в зависимости от уровня загрузки

Исторический и прогнозируемый уровни загрузки определяют агрессивность стратегии ограждения инвентаря (fencing). Предлагается выделять пять категорий поездов по прогнозируемому коэффициенту населенности NLF.

Категория 1: Депрессивная населенность (NLF < 50%). Хронический дефицит пассажиров на всем протяжении маршрута, отсутствие выраженных пиков. Приоритет – минимизация потери емкости. Полное отсутствие логических ограничений по O&D. Цель – захват любого доступного платежеспособного спроса. Рассмотрение вопроса о физическом сокращении составности (отцепке вагонов) [4]. Хронология действий аналитика:

1. D-45 (Открытие продаж): Полное снятие всех системных ограничений, трафаретов и квот. Продажа разрешена на любые мыслимые комбинации O&D.

2. D-15 до D-8: Оценка кривой накопления бронирований. При низких темпах продаж вмешательство запрещено.

3. D-7 до D-2: Пассивное наблюдение. При превышении 70% выкупа на любом O&D – перевод в Категорию 2.

Категория 2: Разбалансированная населенность (NLF = 50%– 65%). Поезд располагает достаточным общим резервом мест, однако на некоторых перегонах вблизи крупных агломераций формируются локальные пики спроса. Хронология действий:

1. D-45: Наложение превентивных мягких ограничений. Блокировка продаж на короткие пригородные участки (менее 3 станций) в радиусе 100 км от крупных городов.

2. D-20: Построение графика секционной населенности (SLF) по каждому перегону.

3. D-10: Закрытие продаж на локальные O&D в зоне дефицитного перегона.

4. D-3: Полное снятие всех ограничений. Оставшаяся емкость отдается локальному спросу.

Категория 3. Стабильная населенность (NLF =65%–80%). Устойчивый спрос с регулярными периодами конкуренции между пассажирами различной дальности следования. Классический полигон для виртуальной вложенности. Хронология действий:

1. D-45: Защита мест (Protection Level) для запросов с протяженностью более 65% маршрута на основе восстановленного спроса [8].

2. D-25: Анализ скорости выкупа магистральных плеч. При опережении нормы — расширение пула защиты.

3. D-10: Блокировка продаж на участки короче 10 станций, пересекающих центры транзитного потока. 4. D-4 до D-1: Сброс нереализованных остатков из защищенного пула в общий инвентарь.

Категория 4: Дефицитная населенность (NLF = 80%90%). Состояние постоянной нехватки провозной способности. Риск вытеснения высокодоходных длинных O&D достигает критических значений. Хронология действий:

1. D-45: Экстремально жесткое разделение инвентаря. Локальные перегоны (короче 35% маршрута) тотально блокируются.

2. D-30: Ревизия образовавшихся «дыр» в инвентаре. Открытие продаж на изолированные пустые перегоны.

3. D-20 до D-10: Расширение пула защиты при опережении темпов продаж.

4. D-9 до D-6: Постепенное открытие продаж на среднемагистральные участки.

5. D-5 до D-3: Абсолютное снятие всех блокировок («Правило последних 24 часов») [4].

Категория 5: Экстремальная населенность (NLF > 90%). Сверх пиковый спрос (национальные праздники, летние отпуска). Поезд гарантированно будет продан на 100%. Хронология действий:

1. D-45: Запрос на физическое масштабирование состава (прицепка дополнительных вагонов) [1]. Установление режима диктатуры магистральных продаж (ограждение до 75–80%).

2. D-14: Продажа непроданных мест исключительно через технологию маршрутного сшивания.

3. D-3: Финальная распродажа изолированных коротких перегонов локальным пассажирам.

Категоризация поездов и соответствующие им стратегии управления детально систематизированы в таблице 4.

Таблица 4. Сводная таблица категорий поездов и стратегий управления

№	Категория	NLF	Базовая парадигма управления емкостью
1	Депрессивная	<50%	Приоритет минимизации потери емкости, полное отсутствие ограничений по O&D
2	Разбалансированная	50%–65%	Точечная интервенция, легкие ограничения на локальных перегонах
3	Стабильная	65%–80%	Классическая балансировка, виртуальная вложенность
4	Дефицитная	80%–90%	Агрессивное ограждение, жесткая блокировка локальных продаж
5	Экстремальная	>90%	Стратегическое планирование, технология маршрутного сшивания

Результаты исследования

Эмпирический анализ исторических данных. Для проверки предпосылок предложенной методологии проведен анализ исторических данных АО «Пассажирские перевозки» (Казахстан) по четырем поездам дальнего следования: магистральным 109Х/110Х (маршрут Актобе - Мангистау, протяженность 1190 км) и сезонным 631Х/632Т (Атырау - Дины Нурпеисовой, 283 км). Период покрытия данных: 2022-2025 гг. Общий объем обработанных записей – 1063 192 строки продаж. Разработана система материализованных представлений (materialized views) для поэтапной очистки и агрегации данных.

Очистка данных и восстановление реального пассажиропотока. Критической проблемой исходных данных является наличие у одного билета как confirmed-строки (при покупке), так и refunded-строки (при возврате). Простой фильтр WHERE status_id = 2 оставлял 168 530 «зомби-подтверждений» – билетов, учтенных в пассажирообороте, но фактически возвращенных. Разработан слой данных mv_sales_traveled, реализующий правило: билет считается проехавшим, если существует confirmed-строка и отсутствует refunded-строка по тому же ticket_number:

```
SELECT s.*
FROM mv_sales_clean s
WHERE s.status_id = 2
    AND NOT EXISTS (
            SELECT 1 FROM mv_sales_clean r
            WHERE r.ticket_number = s.ticket_number AND r.status_id = 1
    );
```

Результаты очистки представлены в таблице 5.

Таблица 5. Очистка данных

Слой	*Строк*	*Что удалено*
sales (raw)	*1063192*	
mv_sales_clean	*995465*	*67727 ETL-дублей*
mv_sales_traveled	*658361*	*337104 зомби-строк*

Физическая валидация очистки проведена через проверку: на одном физическом месте за один рейс сумма distance_km всех проехавших билетов не должна превышать route_km. Результат: 99.984% мест прошли проверку (88 нарушений на 555 511 рейсовых мест).

Базовые показатели. Расчет *RPK* выполнен по формуле (1). Расчет ASK выполнен с учетом per-departure составности по формуле (2). Ключевое решение: состав поезда меняется день-ко-дню (от 1 до 13 вагонов на магистралях), поэтому ASK рассчитывается для каждого рейса индивидуально. NLF рассчитан по формуле (3). Результаты по поездам представлены в таблице 6.

Таблица 6. Базовые показатели по поездам

Поезд	*Рейсов*	*NLF,%*	*RPK,млн*	*ASK,млн*
109Х	*995*	*66.7*	*235.4*	*352.9*
110Х	*957*	*68.9*	*244.6*	*355.2*
631Х	*139*	*76.3*	*4.7*	*6.1*
632Х	*139*	*77.8*	*4.8*	*6.1*

Классификация рейсов по 5 категориям. На основе расчетного NLF рейсы распределены по пяти категориям согласно порогам, определенным в разделе 5. Результаты представлены в таблице 7.

Таблица 7. Распределение рейсов по категориям NLF

№	Категория	NLF	Рейсов	Доля,%	RPK,млн
1	Депрессивная	<50%	364	16.6	39.8
2	Разбалансированная	50%–65%	491	22.4	97.4
3	Стабильная	65%–80%	690	31.5	181.0
4	Дефицитная	80%–90%	521	23.8	120.6
5	Экстремальная	>90%	124	5.7	50.6

Распределение имеет колоколообразную форму с центром в «Стабильной» категории, что соответствует типичному для социально значимых маршрутов характеру спроса.

6.4 Показатель секционной населенности и потеря емкости. Для расчета SLF на каждом перегоне выполнена реконструкция последовательности станций. В базе данных отсутствует поле «порядок станций», поэтому последовательность восстановлена через MIN (distance_km) от каждой станции до терминала. Всего восстановлено 64 станции (по 22 на 109X/110X, по10 на 631X/632T). Сумма длин перегонов точно совпадает с длиной маршрута (1190/1190/283/283км). Каждый билет занимает место на всех перегонах между origin и destination. SLF на перегоне рассчитан по формуле (4). Анализ SLF-профиля поезда 109X выявил две четкие точки разрыва спроса (Таблица 8).

Таблица 8. SLF-профиль поезда 109X

Участок	СреднийSLF	Значение
Актобе-1→Кульсары(legs1–12)	*0.68→0.78*	*Основной“горб” загрузки*
Кульсары→Толтыр(leg13)	*Падение до 0.63*	***Первая точка разрыва***
Толтыр→Бейнеу(legs14–16)	*0.61–0.63*	*Средний участок*
Бейнеу→Устюрт(leg17)	*падениедо0.53*	***Вторая точка разрыва***
Устюрт→Шетпе(legs18–20)	*0.52–0.53*	*Западная половина*
Шетпе→Мангистау(leg21)	***0.47***	*Самый недозагруженный*

На рисунке 1 представлен SLF-профиль поезда 109X (Актобе–Мангистау) как функция номера перегона. На графике отчетливо видны два характерных участка:

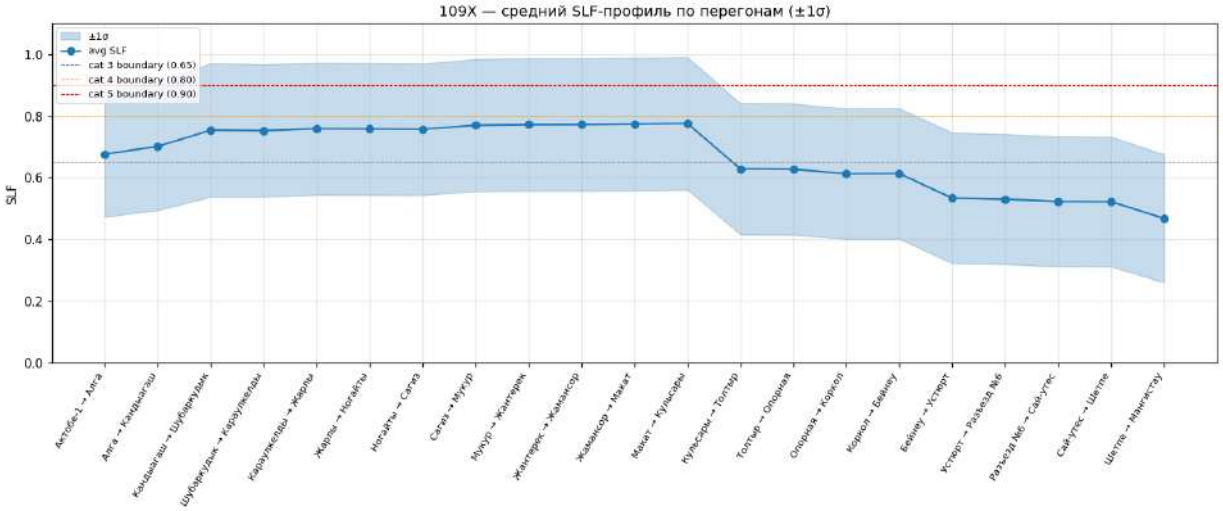


Рисунок 1. SLF-профиль поезда 109X (Актобе– Мангистау)

1. Восточная половина маршрута (legs 1–12): SLF монотонно возрастает от 0.68 до 0.78, формируя устойчивый «горб» загрузки на участке Актобе-1 → Кульсары. Это свидетельствует о стабильном спросе на восточном направлении.

2. Две точки разрыва:

• Кульсары → Толтыр (leg 13): SLF падает до 0.63 — первая точка разрыва спроса;

• Бейнеу → Устюрт (leg 17): SLFпадает до 0.53 — вторая точка разрыва.

3. Западная половина маршрута (legs 18–21): SLF продолжает снижаться до 0.47 на перегоне Шетпе → Мангистау, что указывает на существенно более низкий спрос на западном плече маршрута.

Штриховыми линиями показан доверительный интервал ±1σ, демонстрирующий устойчивость выявленных закономерностей. Точки разрыва – Кульсары и Бейнеу – являются прямыми кандидатами для применения технологии маршрутного сшивания (раздел *Технология маршрутного сшивания Route Splicing).* Для визуализации временной динамики загрузки перегонов построена тепловая карта SLF поезда 109X (Рис. 2). По оси абсцисс отложены перегоны от Актобе-1 до Мангистау, по оси ординат — даты отправления в хронологическом порядке. Цветовая шкала отражает значение SLF: от темно-синего (SLF = 0) до темно-красного (SLF ≥ 0.8).

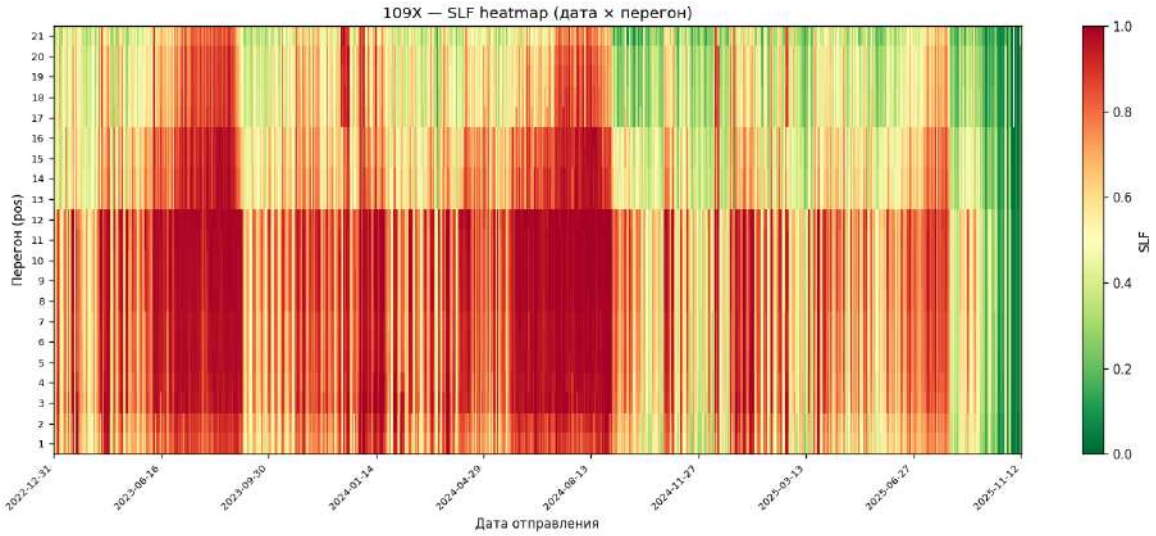


Рисунок 2. Тепловая карта SLF поезда 109X в разрезе дата отправления × номер перегона

Анализ тепловой карты позволяет выделить следующие закономерности:

1. Пространственная структура: Восточная половина маршрута (legs 1-12) на протяжении всего периода наблюдений демонстрирует устойчиво высокую загрузку (красные и оранжевые тона). Западная половина (legs 13-21) характеризуется более низкой и нестабильной загрузкой, особенно на конечных перегонах.

2. Сезонная динамика: В летние месяцы (июль-август 2024) наблюдается пиковая загрузка практически на всем протяжении маршрута. В межсезонье западная половина маршрута переходит в зеленую зону (SLF 0.4-0.6).

3. Тренд 2025 года: Конец 2025 года демонстрирует заметный спад загрузки даже на восточной половине маршрута, что может свидетельствовать о начальных проявлениях эффекта нисходящей спирали.

Тепловая карта подтверждает устойчивость выявленных точек разрыва спроса (legs 13 и 17) во времени и позволяет идентифицировать периоды, когда применение технологии маршрутного сшивания могло бы быть наиболее эффективным. Точки разрыва – Кульсары и Бейнеу – являются прямыми кандидатами для применения технологии маршрутного сшивания. Классификация O&D-корреспонденций по haul-классам представлена в таблице 9. Long-haul доминирует на всех маршрутах, что свидетельствует о фундаментальной роли магистральных корреспонденций в генерации пассажирооборота.

Таблица 9: Распределение RPK по haul-классам

Поезд	*Long-haul (>75%)*	*Mid-haul (35%–75%)*	*Short-haul (<35%)*
109X	*57.7*	*36.8*	*5.5*
110X	*55.6*	*39.7*	*4.7*
631X	*91.5*	*8.3*	*0.2*
632T	*93.1*	*6.7*	*0.2*

Расчет потери емкости по категориям NLF (Таблица 10) показывает, что 89% общего эффекта потери емкости сосредоточено в категориях 1-3 (NLF < 80%).

Таблица10. Потеря провозной емкости (Spoilage) по категориям NLF

№	*Категория*	*Средний NLF*	*Доля Spoilage*	*Млн кресло-км*	*Доля от общего, %*
1	*Депрессивная*	*0.366*	*0.635*	*67.3*	*29.2*
2	*Разбалансированная*	*0.582*	*0.418*	*69.4*	*30.1*
3	*Стабильная*	*0.725*	*0.275*	*68.6*	*29.7*
4	*Дефицитная*	*0.845*	*0.155*	*21.8*	*9.4*
5	*Экстремальная*	*0.929*	*0.071*	*3.9*	*1.7*

Кривая бронирования. Диагностика горизонта продаж показала резкий разрыв на D-45: бакет D-46..60 в 30 раз меньше D-31..45 (0.50% vs1 4.47% выборки). Это соответствует системному горизонту продаж в АСУ «Экспресс» – 45 дней, что подтверждено продакт-менеджером. Для каждого рейса построены две кривые:

- netcurve – доля активных в инвентаре билетов на момент D-X;
- gross curve – доля билетов, когда-либо вошедших в инвентарь (без вычета возвратов).

Кривые выкупа по категориям NLF представлены в таблице 11 и на рисунке 3.

Таблица 11. Кривая выкупа по категориям NLF (доля net от количества фактически перевезенных пассажиров), %

№	*Категория*	*D-45*	*D-30*	*D-21*	*D-14*	*D-10*	*D-7*	*D-3*	*D-1*	*D-0*
1	*Депрессивная*	*0.9*	*12.7*	*23.5*	*34.9*	*44.2*	*52.7*	*69.8*	*88.0*	*100*
2	*Разбалансированная*	*4.3*	*16.5*	*27.8*	*39.8*	*49.1*	*57.4*	*73.1*	*89.0*	*100*
3	*Стабильная*	*2.3*	*16.9*	*30.1*	*44.0*	*54.0*	*62.8*	*77.9*	*91.4*	*100*
4	*Дефицитная*	*2.1*	*20.7*	*35.4*	*50.2*	*59.9*	*68.3*	*82.6*	*93.5*	*100*
5	*Экстремальная*	*2.9*	*30.1*	*47.5*	*69.8*	*69.8*	*76.1*	*85.4*	*93.7*	*100*

Кривые выкупа по категориям NLF представлены на рисунке 3. Наблюдается четкое разделение «веером»: на D-14 разрыв между категорией 1 и категорией 5 составляет 27 процентных пунктов (35% vs 62%). Категория 5 уже на D-30 имеет 30% выкупа против 12% у категории 1 – это дает возможность ранней классификации рейса по темпу выкупа. Сплошные линии (net) демонстрируют долю билетов, остающихся активными в инвентаре на каждый момент времени, тогда как пунктирные линии(gross) показывают совокупный спрос без учета возвратов. Разрыв между этими кривыми отражает теневой спрос (shadow demand) – билеты, которые были куплены, но впоследствии возвращены. Кривые по haul-классам представлены в таблице 12 и на рисунке 4.

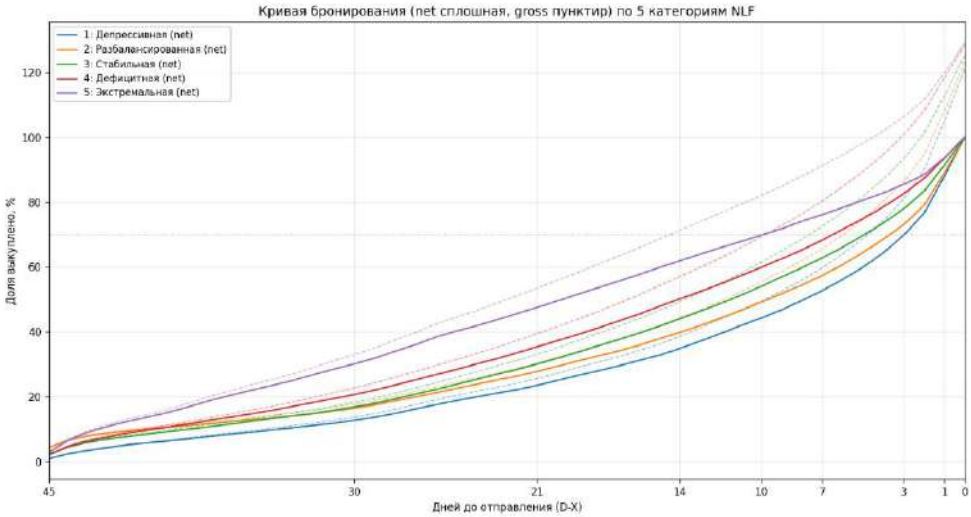


Рисунок 3. Кривая бронирования по категориям NLF: сплошные линии – net, пунктир gross

Таблица 12: Кривая выкупа по haul-классам, %

Класс	*D-30*	*D-21*	*D-14*	*D-7*	*D-3*	*D-1*	*D-0*
Long-haul	*21.3*	*37.4*	*53.4*	*73.0*	*87.4*	*98.0*	*100*
Mid-haul	*20.4*	*34.5*	*48.6*	*67.5*	*82.5*	*94.4*	*100*
Short-haul	*4.7*	*8.2*	*12.6*	*21.9*	*36.4*	*62.3*	*100*

Графическая визуализация этих данных содержится на рисунке 4.

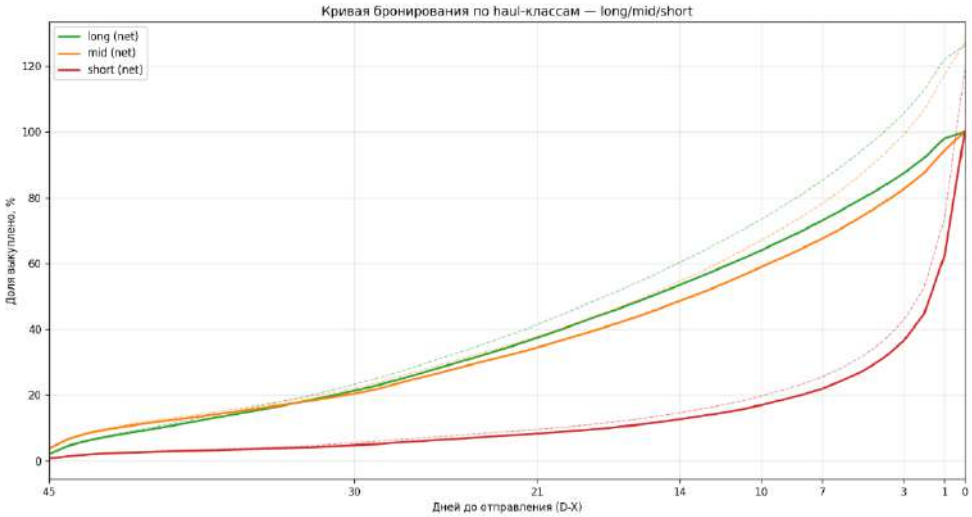


Рисунок 4. Кривая бронирования по haul-классам: long, mid, short

Анализ кривых по haul-классам демонстрирует принципиальное различие в поведении пассажиров разных категорий дальности. Long и mid-haul идут вместе на всем горизонте бронирования: на D-30 они составляют 21.3% и 20.4% соответственно, на D-7 - 73.0% и 67.5%. Short-haul, напротив, представляет собой отдельный трек: пассажиры начинают бронировать значительно позже. На D-3 выкуплено только 36.4% short-haul мест против 87.4% long-haul. В последние три дня short-haul почти утраивается (с 36.4% до 100%).

Это наблюдение является прямым эмпирическим аргументом в пользу виртуальной вложенности (Virtual Nesting) и защиты мест для long-haul корреспонденций (Protection Levels), описанных в разделах *Виртуальная вложенность (Virtual Nesting)* и *Адаптация эвристики ожидаемого предельного пассажирооборота*. Long-haul спрос можно ограждать на ранних этапах, так как short-haul пассажир физически не появляется до последних 2-3 дней.

«Жертва short ради long» из формулы (5) статьи в нашей выборке имеет очень узкое окно реализации – преимущественно D-3..D-1. Анализ разрыва между gross и net кривыми (Таблица 13) позволяет количественно оценить теневой спрос.

Таблица 13. Теневой спрос (полная и чистая кривые бронирования),%

№	Категория	netD-0	gross D-0	Теневой спрос
1	Депрессивная	100	121.0	+21.0
2	Разбалансированная	100	123.2	+23.2
3	Стабильная	100	125.2	+25.2
4	Дефицитная	100	128.4	+28.4
5	Экстремальная	100	129.1	+29.1

Через инвентарь рейса прошло на 21–29% больше билетов, чем итоговое число пассажиров. На пиковых рейсах (категория 5) теневой спрос максимален и составляет 29.1%. Это эмпирическая количественная оценка нецензурированного спроса из раздела *Восстановление истинного спроса* статьи – для категоризированного алгоритма это означает, что наблюдаемый финальный NLF систематически занижает истинный пиковый спрос. Анализ временного распределения возвратов (refundpace) показал, что 63.4% всех возвратов происходят в последние 7 дней до отправления (Таблица14*).*

Таблица 14. Распределение возвратов по дням до отправления

Окно	Возвратов	Доля,%
D-0	24856	14.75
D-1	23757	14.10
D-2–3	27721	16.45
D-4–7	30445	18.07
D-8–14	26723	15.86
D-15–21	14479	8.59
D-22–30	10968	6.51
D-31–45	9581	5.69

Анализ возвратов показал, что 63.4% всех возвратов приходится на последние 7 дней до отправления поезда. Данный результат свидетельствует о существенном перераспределении доступной провозной емкости на заключительном этапе продаж и подтверждает необходимость динамического управления инвентарем в преддверии отправления. Следовательно, при принятии решений о снятии ограничений и открытии продаж на ранее защищенные O&D-корреспонденции необходимо учитывать ожидаемый объем возвратов, что позволяет эффективнее использовать освобождающиеся места и снизить потери пассажирооборота.

Сравнение кривых выкупа по годам в сопоставимом окне март-июль (Таблица 15) позволяет оценить динамику покупательского поведения

Таблица 15. Год-к-году динамика кривой выкупа (март-июль), %

Поезд	Год	D-30	D-21	D-14	D-7	D-3	D-0
109X	2023	15.6	29.5	44.1	62.1	76.4	100
109X	2024	17.6	31.6	45.2	61.6	74.6	100
109X	2025	15.1	28.2	42.8	61.2	76.7	100
110X	2023	23.5	40.5	57.1	74.6	85.1	100
110X	2024	29.3	45.9	59.4	74.1	83.9	100
110X	2025	18.8	34.6	49.1	66.9	82.6	100

У поезда 109X динамика бронирования по годам остается стабильной: различия между годовыми кривыми не превышают 2 процентных пунктов. Для поезда 110X, напротив, на горизонте D-30 в 2025 году доля выкупленных мест составляет 18.8% против 29.3% в 2024 году, что соответствует замедлению темпов раннего бронирования примерно на 35%. При этом к моменту отправления (D-0) различия практически исчезают, а итоговый NLF остается сопоставимым. Полученный результат свидетельствует о смещении спроса в сторону более позднего бронирования и показывает, что использование только ранних данных о продажах может привести к недооценке ожидаемого спроса при принятии решений о распределении провозной емкости.

Управление системными рисками и механизмами тигации

Имплементация агрессивных методов управления O&D в среде социально значимых перевозок сопровождается комплексом институциональных, технологических и социальных рисков.

Социально-политический конфликт. Субсидируемые государством пассажирские перевозки обладают двойственной природой. С одной стороны, от оператора требуют коммерческой эффективности и сокращения размера субсидий; с другой – поезда выполняют безальтернативную социальную функцию обеспечения территориальной связности [2]. Жесткая блокировка продаж на короткие перегоны с целью роста RPK неизбежно приведет к социальной напряженности. Для митигации данного риска используется концепция мультимодальной синхронизации и императивное «Правило последних 24 часов», гарантирующее, что ни одно место не уйдет пустым [4].

Искажение данных и проблема No-Show. Опора аналитиков исключительно на исторические данные продаж таит в себе ловушку. Исторические ряды часто сильно загрязнены факторами массовых возвратов, групповых отмен бронирований и неявки пассажиров (no-shows) [1]. Феномен нисходящей спирали заставляет аналитиков из года в год занижать квоты для длинных поездок [8]. Решением является обязательное внедрение процедур очистки данных: корректировка исторических рядов на коэффициент упущенного спроса в пиковые периоды и непрерывный мониторинг текущих темпов бронирования.

Инфраструктурные барьеры динамической емкости. Стратегия Категории 5 опирается на добавление вагонов. На практике этот механизм жестко лимитирован длиной пассажирских платформ на периферийных станциях и тяговыми расчетами [7].

Технологическая ригидность устаревших систем. Фундаментальным ограничением выступает архитектура систем резервирования, используемых в СНГ (АСУ «Экспресс»). Эти системы проектировались в эпоху плановой экономики и ориентированы на статичное квотирование [5]. Они зачастую не поддерживают сложную математику виртуальной вложенности O&D «из коробки» [8].

Дискуссия

Ограничения исследования Представленное исследование имеет следующие ограничения:

1. Отсутствие данных о упущенном спросе: В системе АСУ «Экспресс» не фиксируются попытки бронирования, завершившиеся отказом. Восстановление истинного спроса выполнено косвенными методами (анализ кривой бронирования, gross vs net).

2. Неполнота данных о составе поезда: Отсутствует информация о вагонах без продаж (вагоны, в которых отсутствовали зарегистрированные продажи) и о прицепке/отцепке вагонов на промежуточных станциях. Это может приводить к систематическим ошибкам в расчете ASK на отдельных леггах.

3. Ограниченный временной горизонт: Данные покрывают период 2022–2025 гг. с неравномерным покрытием месяцев (отсутствуют январь 2023, ноябрь декабрь 2025).

4. Ограниченная выборка поездов: Анализ выполнен только по четырем поездам (два магистральных, два сезонных), что может ограничивать обобщаемость результатов на другие маршруты.

5. Отсутствие данных о no-show: Используемые данные не позволяют идентифицировать пассажиров, не явившихся на посадку (no-show), поскольку ин формация о фактической посадке отсутствует.

6. Статичность тарифов: Исследование исходит из предположения о фиксированных тарифах, что соответствует текущей практике, но не позволяет оценить совместный эффект ценообразования и управления емкостью.

Выявленные ограничения определяют перспективные направления дальнейших исследований. Приоритетом является расширение выборки маршрутов и восстановление непрерывных временных рядов для повышения обобщаемости результатов. Кроме того, планируется уточнить расчет *ASK* с учетом фактора *no-show* и схем формирования поездов, а также перейти от статичной тарифной модели к совместной оптимизации управления емкостью (*capacity management*) и динамического ценообразования.

Заключение

В условиях детерминированных тарифов максимизация пассажирооборота становит ся основным направлением повышения эффективности социально значимых желез нодорожных перевозок. Предложенная методология позволяет перейти от статического распределения мест к интеллектуальному управлению провозной емкостью на основе прогнозирования спроса и анализа O&D-потоков. Основные результаты работы:

1. Разработана математическая модель управления емкостью с целевой функцией максимизации RPK.

2. Предложена классификация поездов по 5 категориям в зависимости от прогнозируемого NLF.

3. Разработаны пошаговые алгоритмы управления для каждой категории.

4. Проведен эмпирический анализ на реальных данных АО «Пассажирские пере возки» (2 190 рейсов, 658 361 проехавших пассажиров), показавший:

- Наличие четких точек разрыва SLF (Кульсары и Бейнеу на 109X) — кандидатов для технологии маршрутного сшивания;
- Теневой спрос 21–29% от количества фактически перевезенных пассажи ров;
- 63.4% возвратов в последние 7 дней;
- Ранние сигналы изменения покупательского поведения на поезде 110X (падение темпа выкупа на D-30 на 35%).

5. Выявлены и систематизированы риски внедрения и предложены механизмы их митигации.

Практическое применение разработанного подхода способно повысить сетевую населенность поездов дальнего следования и сократить потребность в государственном субсидировании без изменения тарифной политики. Несмотря на полученные результаты, предложенная методология требует дальнейшего развития. Дальнейшие исследования могут быть направлены на разработку методов автоматической классификации на основе машинного обучения и интеграцию с системами динамического ценообразования в случае либерализации тарифной политики.

Благодарность

Данное исследование выполнено при финансовой поддержке Комитета науки Министерства науки и высшего образования Республики Казахстан в рамках грантового финансирования AP26103651 «Разработка и улучшение модели управления доходами железнодорожных пассажирских перевозок на основе методов машинного обучения и искусственного интеллекта». Авторы выражают благодарность сотрудникам АО «Пассажирские перевозки» за предоставленные данные и ценные консультации.

References

[1] Kamandanipour K., Haji Yakhchali S., Tavakkoli-Moghaddam R. Dynamic revenue management in a passenger rail network under price and fleet management decisions // Annals of Operations Research. — 2023. — 6 Apr. — P. 1–25. — DOI: 10.1007/s10479-023-05296-4. — PMID: 37361095. — PMCID: PMC10078050.

[2] Fraser J., Thompson L. S. Purchasing Power Parity Adjusted Railway Performance [Электронный ресурс] // World Bank Infrastructure Note. — 1994. — No. RW-9. — December. — Режим доступа: https://documents1.worldbank.org/curated/en/242121468327616446/pdf/360080Purchasing0parity0td1rw9.pdf (дата обращения: 03.06.2026).

[3] Wang Y., Li Z., Zhang X. A Literature Review of Railway Pricing Based on Revenue Management // Mathematics. — 2023. — Vol. 11, No. 4. — P. 85. — DOI: 10.3390/math11040085.

[4] Pricing and Revenue Management [Электронный ресурс] // American Association of State Highway and Transportation Officials (AASHTO). — Режим доступа: https://transportation.org/rail/wp-content/uploads/sites/24/2024/07/AASHTO v4.ppt (дата обращения: 03.06.2026).

[5] Economic Commission for Europe, Inland Transport Committee, Working Party on Rail Transport, Group of Experts on International Railway Passenger Hubs. Handbook of Instruments Governing International Rail Passenger Transport [Элек тронный ресурс] : Analysis of the legal framework governing international passenger transportation / transmitted by the Russian Federation. — Second session, Geneva, 15–17 December 2021. — Agenda item 6. — Approved at the meeting of the OSJD Commission on Transport Law on 24–26 November 2020. — Режим доступа: https://unece.org/sites/default/files/2022-07/ECE-TRANS-SC.2 HUBS-2021-07e.pdf (дата обращения: 03.06.2026).

[6] Broman E., Eliasson J., Aronsson M. Efficient capacity allocation on deregulated railway markets // Journal of Rail Transport Planning & Management. — 2022. — Vol. 21. — P. 100294. — DOI: 10.1016/j.jrtpm.2021.100294.

[7] Klabes S. G. Algorithmic Railway Capacity Allocation in a Competitive European Railway Market [Электронный ресурс] : Dissertation zur Erlangung des akademischen Grades eines Doktors der Ingenieurwissenschaften / S. G. Klabes ; Rheinisch-Westf¨alische Technische Hochschule Aachen. — Gutachter: Prof. Dr.-Ing. E. Wendler, Prof. Dr.rer.nat. B. V¨ocking. — Tag der m¨undlichen Pr¨ufung: 14. Januar 2010. — Режим доступа: http://publications.rwth aachen.de/record/51543/files/Klabes_Sebastian.pdf (дата обращения: 03.06.2026).

[8] Berto A. Rail Yield Management. Trenitalia Case [Электронный ресурс] : PhD Thesis in Operations Research / A. Berto ; Roma Tre University. — Advisors: Prof. A. Pacifici, Prof. G. Nicosia. — Year 2019. — Режим доступа: https://arcadia.sba.uniroma3.it/bitstream/2307/40637/1/Tesi_PhD_Berto_2019.pdf (дата обращения: 03.06.2026).

[9] Manchiraju C. Dynamic Pricing and Capacity Optimization in Railways [Электрон ный ресурс] // Mitsubishi Electric Research Laboratories. — 2025. — No. TR2025 007. — Режим доступа: https://www.merl.com/publications/docs/TR2025-007.pdf (дата обращения: 03.06.2026).

[10] Boltaikhanova T., Bostanbekov K., Nurseitov D., Begimbetov O. (2024). Forecasting Railway Ticket Demand. 10.20944/preprints202411.0098.v1

ҒТАХР 20.53.15

https://doi.org/10.51889/2959-5894.2026.95.3.016

Ж. Дүйсенбекқызы[1*], А. Тұрсынхан[2], Г. Газиз[3], Б. Ілесова[1]

[1]Алматы технологиялық университеті, Алматы қ., Қазақстан
[2]Халықаралық инженерлік-технологиялық университеті, Алматы қ., Қазақстан
[3]Абай атындағы Қазақ ұлттық педагогикалық университеті, Алматы қ., Қазақстан
[*]*e-mail: duisenbekkyzy.zh@kaznu.kz*

ТРАНСФЕРЛІК ОҚЫТУ ӘДІСІ НЕГІЗІНДЕ ҚАЗАҚ ТІЛДІ БАЛАЛАРДЫҢ СӨЙЛЕУІН АВТОМАТТЫ ТАНУ ЖҮЙЕСІН МОДИФИКАЦИЯЛАУ

Аңдатпа

Зерттеу жұмысы қазақ тілінің акустикалық және фонетикалық ерекшеліктерін ескере отырып, 5–12 жас аралығындағы балалардың сөйлеуін автоматты тану тиімділігін арттыруға бағытталған. Зерттеудің мақсаты – OpenAI ұсынған Whisper-tiny архитектурасын балалар сөйлеуінің мамандандырылған аудиодеректері негізінде қайта оқыту арқылы қазақ тіліне бейімдеу. Жұмыста екі кезеңді fine-tuning стратегиясы қолданылды: бірінші кезеңде модель 235 аудиоүлгімен, ал екінші кезеңде 1936 үлгіге дейін кеңейтілген корпуспен оқытылды. Аудиожазбалардың дискреттеу жиілігі 16 кГц-ке келтіріліп, мәтіндік транскрипциялар біріздендірілді. Модель сапасы Word Error Rate (WER) және Accuracy метрикалары арқылы бағаланды. Эксперимент нәтижесінде WER көрсеткіші 0,80-нен 0,42-ге дейін төмендеп, салыстырмалы жақсару 47,5%-ды құрады, ал дәлдік 20,43%-дан 57,95%-ға дейін артты. Нәтижелер деректер көлемін ұлғайту мен доменге бейімделген қайта оқытудың балалар сөйлеуін тану сапасына оң әсер ететінін көрсетті. Ұсынылған тәсілді аз ресурсты тілдерге арналған білім беру және интеллектуалды сөйлеу жүйелерін әзірлеуде қолдануға болады.

Түйін сөздер: автоматты сөйлеуді тану (ASR), Whisper модельдері, қазақ тілі, балалар сөйлеуі, fine-tuning, трансферлік оқыту, WER метрикасы, нейрондық желілер.

Ж. Дүйсенбекқызы[1], А. Тұрсынхан[2], Г. Газиз [3], Б. Ілесова[1]
[1]Алматинский технологический университет, г. Алматы, Казахстан
[2]Международный инженерно-технологический университет, г. Алматы, Казахстан
[3]Казахский национальный педагогический университет имени Абая, г. Алматы, Казахстан

МОДИФИКАЦИЯ СИСТЕМЫ АВТОМАТИЧЕСКОГО РАСПОЗНАВАНИЯ ДЕТСКОЙ РЕЧИ НА КАЗАХСКОМ ЯЗЫКЕ ПОСРЕДСТВОМ ТРАНСФЕРНОГО ОБУЧЕНИЯ

Аннотация

Данная работа посвящена повышению эффективности автоматического распознавания речи детей в возрасте 5–12 лет с учетом акустических и фонетических особенностей казахского языка. Цель исследования – адаптировать предложенную OpenAI архитектуру Whisper-tiny к казахскому языку посредством дообучения на специализированном наборе аудиоданных детской речи. В работе применена двухэтапная стратегия fine-tuning: на первом этапе модель обучалась на 235 аудиообразцах, а на втором – на расширенном корпусе из 1936 образцов. Аудиозаписи были приведены к частоте дискретизации 16 кГц, а текстовые транскрипции унифицированы. Качество модели оценивалось с использованием метрик Word Error Rate (WER) и Accuracy. В результате эксперимента показатель WER снизился с 0,80 до 0,42, что соответствует относительному улучшению на 47,5%, а точность увеличилась с 20,43% до 57,95%. Результаты показали, что увеличение объема данных и доменно-ориентированное дообучение положительно влияют на распознавание детской речи. Предложенный подход может применяться при разработке образовательных и интеллектуальных речевых систем для языков с ограниченными ресурсами.

Ключевые слова: автоматическое распознавание речи (ASR), модели Whisper, казахский язык, детская речь, fine-tuning, трансферное обучение, метрика WER, нейронные сети.

Zh. Duisenbekkyzy[1], A. Tursynkhan[2], G. Gaziz[3], B. Ilessova[1]
[1]Almaty Technological University, Almaty, Kazakhstan
[2]International Engineering and Technology University, Almaty, Kazakhstan
[3]Abai Kazakh National Pedagogical University, Almaty, Kazakhstan

ENHANCING AUTOMATIC SPEECH RECOGNITION FOR KAZAKH PEDIATRIC SPEECH THROUGH TRANSFER LEARNING AND FINE-TUNING

Abstract

This study aims to improve automatic speech recognition for children aged 5–12 by accounting for the acoustic and phonetic characteristics of the Kazakh language. The objective is to adapt OpenAI's Whisper-tiny architecture to Kazakh through fine-tuning on a specialized children's speech dataset. A two-stage fine-tuning strategy was applied: the model was first trained on 235 audio samples and then on an expanded corpus of 1,936 samples. The audio recordings were standardized to a sampling rate of 16 kHz, and the text transcriptions were normalized. Model performance was evaluated using Word Error Rate (WER) and Accuracy. The experimental results showed that WER decreased from 0.80 to 0.42, representing a relative improvement of 47.5%, while Accuracy increased from 20.43% to 57.95%. These findings demonstrate that increasing the amount of data and applying domain-specific fine-tuning improve the recognition of children's speech. The proposed approach can support the development of educational and intelligent speech systems for low-resource languages.

Keywords: Automatic Speech Recognition (ASR), Whisper models, Kazakh language, pediatric speech, fine-tuning, transfer learning, WER metric, neural networks.

Кіріспе

Негізгі ережелер. Зерттеу жұмысы 5-12 жас аралығындағы қазақ тілді балалардың сөйлеуін автоматты тануды жетілдіруге бағытталған. OpenAI Whisper-tiny архитектурасы екі кезеңді fine-tuning стратегиясы арқылы мамандандырылған аудиодеректер жиынтығында бейімделді. Корпус бастапқыда 235 аудиоүлгіден тұрып, кейіннен 1936 сэмплге дейін кеңейтілді. Нәтижесінде WER (Word Error Rate) 0.80-ден 0.42-ге дейін төмендеп, жүйенің дәлдігі 20,43%-дан 57,95%-ға дейін артты. Алынған нәтижелер ресурсы шектеулі тілдер үшін балалар сөйлеуін тану жүйелерін әзірлеуде доменге бейімделген қосымша оқытудың тиімді екенін көрсетті. Сөйлеуді автоматты тану (ASR) технологиялары дүние жүзінде қарқынды дамып, білім беру, денсаулық сақтау, интеллектуалды интерфейстер және қолжетімділік салаларында кеңінен қолданылуда. Соңғы жылдары end-to-end архитектуралар мен transformer негізіндегі жүйелер үлкен көлемдегі деректерді пайдалана отырып, ASR дәлдігін айтарлықтай жақсартты. Алайда ғылыми әдебиеттер ASR модельдерінің өнімділігі олардың оқытылған тіліне және доменіне тікелей байланысты екенін көрсетеді [1]. Осылайша, аз ресурсты тілдер үшін сөйлеуді тану жүйелерін әзірлеу негізгі ғылыми міндеттердің бірі болып қала береді.

Аз ресурсты тілдер арасында қазақ тілі үшін ASR зерттеулері әлі де жеткіліксіз. Қазақ тілінің агглютинативті морфологиясы, дауысты үндестік, флексиясыз құрылым және фонетикалық тұрақсыздық сөйлеуді тану модельдері үшін айтарлықтай қиындықтар туғызады [2]. Бұл лингвистикалық ерекшеліктер алдын ала оқытылған көптілді модельдерді қазақ тіліне тікелей қолдануды шектейді.

Балалар сөйлеуін тану ASR саласындағы дербес күрделі мәселе болып саналады. Ересектер сөйлеуімен салыстырғанда балалар сөйлеуінде негізгі тон жиілігі жоғары, артикуляциясы тұрақсыз және форманттық жиіліктері өзгермелі болады [3]. Мұндай ерекшеліктер тану қателерінің көбеюіне әкеліп, балалар сөйлеуіне бейімделген модельдерді әзірлеуді қажет етеді. OpenAI әзірлеген Whisper моделі 680 000 сағаттан астам көптілді сөйлеу деректері негізінде алдын ала оқытылған. Дегенмен модель белгілі бір тілдің және қолданылу саласының ерекшеліктеріне қосымша бейімделмесе, оның тану дәлдігі төмендеуі мүмкін [4]. Трансферлік оқыту әдістері көлемі шектеулі домендік деректер негізінде модельдің нәтижелілігін арттыруға мүмкіндік береді [5]. Д.Р. Рахимова мен Ж. Дүйсенбекқызының доменге бейімделген fine-tuning жөніндегі зерттеулері [6] де осы тұжырымды растайды.

Зерттеудің мақсаты – Whisper-tiny моделін қазақ тіліндегі балалар сөйлеуін тануға бейімдеу және екі кезеңді fine-tuning стратегиясының тиімділігін эксперименттік тұрғыдан бағалау.

Зерттеу әдіснамасы

Зерттеуде OpenAI әзірлеген Whisper-tiny моделі қолданылды. Whisper — кодтаушы-декодтаушы (encoder-decoder) құрылымына негізделген трансформерлік ASR моделі; оның tiny нұсқасында 39 миллион параметр бар [7]. Архитектура үш негізгі құрамдастан тұрады: белгілерді бөліп алу модулі (Feature Extractor), кодтаушы (Encoder) және декодтаушы (Decoder). Белгілерді бөліп алу кезеңінде аудио 16 кГц жиілікке қайта дискреттеліп, 80 өлшемді мел-спектрограмма құрылады. Кодтаушы мен декодтаушының әрқайсысы төрт трансформерлік қабаттан тұрады, ал декодтау авторегрессиялық тәсілмен орындалады. Whisper-tiny моделін қосымша оқыту үдерісінің архитектурасы 1-суретте көрсетілген. Whisper-tiny моделінің негізгі конфигурациялық сипаттамалары 1-кестеде берілген.



Сурет 1. Whisper Fine-Tuning архитектурасы

Кесте 1. Whisper-tiny моделінің негізгі конфигурациялық сипаттамалары

Параметр	*Сипаттама*
Параметрлер саны	*39 миллион*
Encoder қабаттары	*4*
Decoder қабаттары	*4*
Attention Heads	*6*
Mel-Spectrogram өлшемі	*80 bins*
Pre-training көлемі	*680,000 сағат аудио*

Зерттеуде деректер жиыны екі кезеңде қалыптастырылды: бірінші кезеңде 235 аудиоүлгі, ал екінші кезеңде 1936 аудиоүлгі пайдаланылды (2-кесте). Барлық аудиожазба 16 кГц дискреттеу жиілігіне, моноарнаға және 16 биттік PCM пішіміне келтірілді. Мәтіндік транскрипциялар кіші әріппен жазылып, бірыңғай форматқа түсірілді. Екі кезеңдегі аудиодеректердің құрамы мен сандық сипаттамалары 2-кестеде берілген.

Кесте 2. Деректер жиынының сипаттамалары

Сипаттама	*1-кезең (235)*	*2-кезең (1936)*
Бір сөздер	*183 (78%)*	*1479 (76,3%)*
2–5 сөзді фразалар	*52 (22%)*	*446 (23%)*
6–10 сөз	*0*	*6 (0,3%)*
11+ сөз	*0*	*7 (0,4%)*
Бірегей сөздер	*134*	*529*
Жалпы токендер	*612*	*2858*

2-кестеге сәйкес, екінші кезеңде деректер жиыны едәуір кеңейтілді. Бір сөзден тұратын аудиоүлгілер саны 183-тен 1479-ға, ал 2-5 сөзден тұратын фразалар саны 52-ден 446-ға дейін өсті. Бірегей сөздер саны 134-тен 529-ға, жалпы токендер саны 612-ден 2858-ге дейін артты. Сонымен қатар кеңейтілген корпусқа 6-10 және 11-ден көп сөзден тұратын күрделі фразалар қосылды. Бірінші кезеңде модель 235 үлгі негізінде қосымша оқытылып, қазақ тілінің негізгі фонетикалық заңдылықтарына бейімделді. Екінші кезеңде модель 1936 үлгіден тұратын кеңейтілген корпуспен қайта оқытылды [8]. Шығын функциясы кросс-энтропия әдісімен есептелді: $L = -\sum \log P(y_t|y_{<t}, X)$, мұндағы X - мел-спектрограмма, ал y_t - транскрипцияның t мезетіндегі токені. Модель NVIDIA T4 графикалық процессорында AdamW оңтайландырғышын пайдалану арқылы оқытылды. ASR жүйесінің сапасын бағалау үшін екі метрика қолданылды: WER = (S + D + I) / N және Accuracy = (1 − WER) × 100%. Мұндағы S - алмастырулар саны, D - түсіріп қалдырулар саны, I - артық кірістірулер саны, ал N - эталондық мәтіндегі сөздердің жалпы саны.

Зерттеу нәтижелері

2-суретте 1-кезең оқыту барысында (235 мысал) loss мәнінің өзгерісі көрсетілген. Бастапқыда loss өте жоғары (~8,8) болды, бірақ алғашқы бірнеше итерацияда тез төмендеді және ~0,4 деңгейінде тұрақтады. Бірінші кезеңдегі оқыту процесінің тұрақтылығын бағалау үшін loss функциясының итерациялар бойынша өзгерісі талданды. Оның динамикасы 2-суретте көрсетілген.

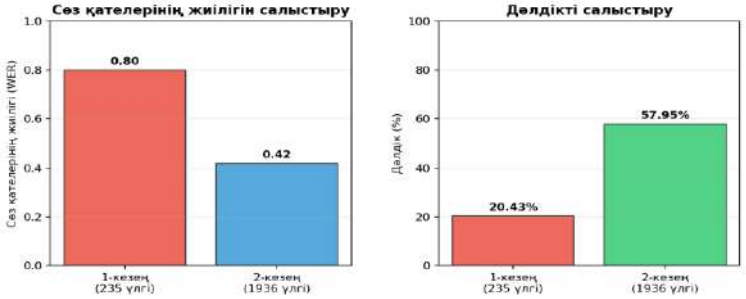


Сурет 2. Оқыту Loss динамикасы (1-кезең)

2-суреттен шығын функциясының бастапқы мәні шамамен 8,8 болғанын және алғашқы итерациялар кезінде жылдам төмендегенін көруге болады. Оқыту соңында бұл көрсеткіш шамамен 0,4 деңгейіне жетті. Мұндай динамика модельдің бастапқы 235 аудиоүлгідегі негізгі акустикалық және фонетикалық заңдылықтарды меңгергенін көрсетеді. Екінші кезеңдегі кеңейтілген fine-tuning нәтижелері 3-суретте берілген. Орташа шығын мәні 0,23-0,33 аралығында өзгерді. Ең төменгі мәннің 0,005–0,01 аралығында болуы қарапайым сөздердің жақсы танылғанын, ал ең жоғары мәндердің 1,8-2,2 аралығында болуы күрделі немесе ұзақ фразаларды тануда қателердің сақталғанын көрсетеді. Алынған статистикалық нәтижелер 3-суретте берілген. 3-суретке сәйкес, шығын функциясының орташа мәні 0,23–0,33 аралығында өзгерді. Ең төменгі мәндер 0,005–0,01 аралығында тіркеліп, қарапайым сөздердің жоғары дәлдікпен танылғанын көрсетті. Ең жоғары мәндердің 1,8–2,2 аралығында болуы ұзақ әрі күрделі фразаларды тану кезінде жекелеген қателердің сақталғанын білдіреді. Екі кезеңдегі модель нәтижелерін салыстыру үшін WER және Accuracy көрсеткіштері қолданылды. Салыстыру нәтижелері 4-суретте берілген: WER 0,80-нен 0,42-ге дейін төмендеп, 47,5% салыстырмалы жақсару байқалды; Accuracy 20,43%-дан 57,95%-ға дейін артты. Оқшауланған сөздер үшін толық сәйкестік (Exact Match) көрсеткіші 66,73%-ды құрады. Аталған метрикалардың салыстырмалы нәтижелері 4-суретте көрсетілген. 4-суретке сәйкес, екінші кезеңдегі fine-tuning нәтижесінде WER көрсеткіші 0,80-ден 0,42-ге дейін төмендеді. Accuracy көрсеткіші 20,43%-дан 57,95%-ға дейін артты.

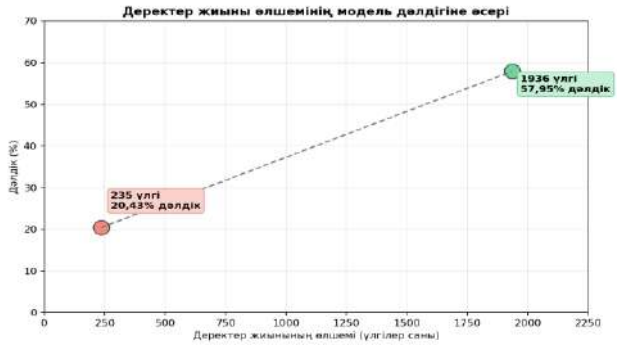


Сурет 3. Оқыту Loss статистикасы, 2-кезең (1936 сэмпл)



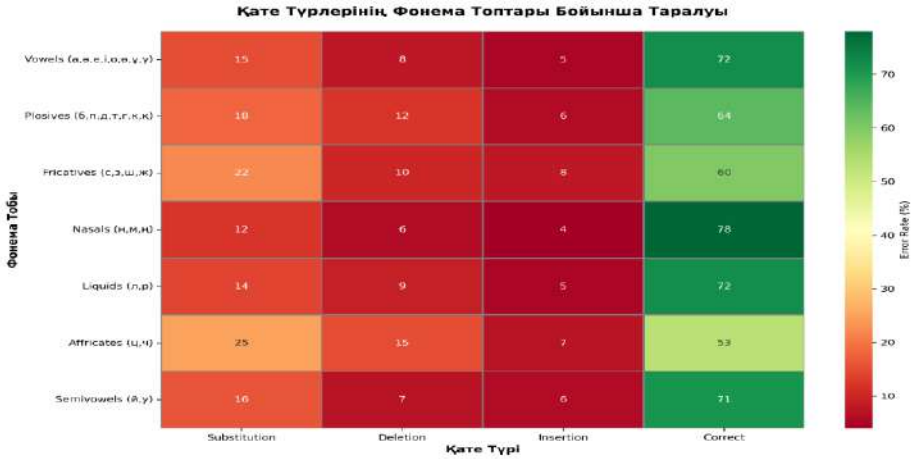
Сурет 4. WER және Accuracy салыстыруы

Зерттеуде модельді дайындау, оқыту және бағалау бірнеше өзара байланысты кезең арқылы орындалды. Ұсынылған екі кезеңді оқыту процесінің толық схемасы 5-суретте берілген.



Сурет 5. Оқыту процесінің толық схемасы

5-суретте көрсетілгендей, зерттеу процесі аудиодеректерді жинау мен алдын ала өңдеуден, модельді екі кезеңде қайта оқытудан және алынған нәтижелерді WER мен Accuracy метрикалары арқылы бағалаудан тұрады. Мұндай кезеңдік тәсіл модельді алдымен қазақ тілінің негізгі фонетикалық ерекшеліктеріне, кейін балалар сөйлеуінің күрделі үлгілеріне бейімдеуге мүмкіндік берді. 6-суреттегі график деректер жиыны өлшемі мен модель дәлдігі арасындағы тікелей байланысты суреттейді. Жалпы жақсарту метрикалары 3-кестеде берілген. Бұл байланыстың графикалық көрінісі 6-суретте ұсынылған. 6-суреттен деректер жиынының көлемі артқан сайын модель дәлдігінің де жоғарылағанын көруге болады. Бастапқы 235 сэмпл пайдаланылған кезде дәлдік 20,43%-ды құраса, 1936 сэмплмен оқытудан кейін 57,95%-ға жетті. Жалпы жақсарту көрсеткіштері 3-кестеде берілген. 3-кестеде көрсетілгендей, деректер жиынының көлемі бастапқы мөлшермен салыстырғанда шамамен 8,23 есе ұлғайды. Осы кезде Accuracy көрсеткіші 183,7%-ға өсті, WER 47,5%-ға, ал шығын функциясының мәні 84,2%-ға төмендеді. Бұл нәтижелер деректер корпусын кеңейту және екі кезеңді fine-tuning стратегиясын қолдану модель нәтижелілігін арттырғанын көрсетеді.



Сурет 6. Деректер жиыны өлшемінің модель дәлдігіне әсері

Кесте 3. Жалпы жақсарту метрикалары

Метрика	*Өзгеріс (%)*
Accuracy өсімі	*+183,7%*
WER төмендеуі	*−47,5%*
Loss төмендеуі	*−84,2%*
Деректер жиыны өлшемінің өсімі	*+823,0%*

Дискуссия

Алынған нәтижелер деректер жиынының көлемі мен сапасы модель нәтижелілігіне елеулі әсер ететінін көрсетті. Деректер жиынының көлемі шамамен 8,23 есе ұлғайған кезде WER 47,5%-ға төмендеп, Accuracy 183,7%-ға артты. Бұл нәтиже нейрондық желілердің тиімділігі оқыту деректерінің көлеміне тәуелді екені туралы тұжырымдармен сәйкес келеді [9]. Фонемалар деңгейіндегі қателер 7-суретте талданды.



Сурет 7. Фонема тобы бойынша қателіктерді талдау

Аффрикаттар (ц, ч) тобындағы алмастыру қателерінің үлесі 25%-ды құрады. Фрикативтер мен шұғыл дауыссыздардағы қателер тиісінше 22% және 18% болды. Мұрын жолды дауыссыздар (н, м, ң) мен үнді дауыссыздар (л, р) топтарындағы қателердің үлесі төменірек болып, тиісінше 12% және 14%-ды құрады. 7-суретке сәйкес, ең жоғары алмастыру қатесі аффрикаттар (ц, ч) тобында тіркеліп, 25%-ды құрады. Фрикативтер мен шұғыл дауыссыздардағы қателердің үлесі тиісінше 22% және 18% болды. Мұрын жолды дауыссыздарда (н, м, ң) және үнді дауыссыздарда (л, р) қателер сирегірек кездесіп, тиісінше 12% және 14%-ды құрады. Нәтижелер акустикалық белгілері ұқсас кейбір фонемаларды ажырату модель үшін күрделі болғанын көрсетті. Мысалы, модель «Балқаш» сөзін «алқаш» деп, ал «бақа» сөзін «тақан» деп таныған. Wang және әріптестері [10] толассыз (end-to-end) жүйелердің дәстүрлі HMM-GMM архитектураларымен салыстырғанда тиімді екенін

көрсеткен. Бұл тұжырым трансформерлік модельді балалар сөйлеуіне бейімдеудің теориялық негіздерінің бірі болып табылады.

Д. Р. Рахимова мен Ж. Дүйсенбекқызының зерттеулерінде [11, 12] модельді қосымша оқыту оның нәтижелілігін арттыратыны көрсетілген. Кейінгі зерттеулерде деректер жиынын әртүрлі жас топтарына жататын жаңа сөйлеу үлгілерімен толықтыру, Whisper моделінің ірі нұсқаларын сынау, сондай-ақ деректерді жасанды көбейту (data augmentation) және білімді дистилляциялау (knowledge distillation) әдістерін қолдану жоспарланады.

Қорытынды

Зерттеу нәтижелері Whisper-tiny моделін қазақ тіліндегі балалар сөйлеуін тануға бейімдеуде fine-tuning әдісінің тиімді екенін көрсетті. Бірінші кезеңде 235 үлгімен оқытылған модельдің WER көрсеткіші 0,80, ал Accuracy көрсеткіші 20,43% болды. Екінші кезеңде 1936 үлгімен қосымша оқытудан кейін WER 0,42-ге дейін төмендеп, Accuracy 57,95%-ға дейін артты. Бірінші кезеңмен салыстырғанда WER 47,5%-ға төмендеп, Accuracy 183,7%-ға өсті.

Фонемалар деңгейіндегі талдау алмастыру, түсіріп қалдыру және кірістіру қателерінің фонема топтарына қарай өзгеретінін көрсетті. Ең жоғары қате үлесі аффрикаттар мен фрикативтерде байқалды. Алынған нәтижелер ресурсы шектеулі тілдерге бейімделген ASR жүйелерін әзірлеуге негіз болады және мұндай жүйелерді білім беру саласында қолдану мүмкіндігін көрсетеді.

Пайдаланылған дереккөздердің тізімі

[1] Bai, Z.; Zhang, X.-L. Speaker recognition based on deep learning: An overview. Neural Networks, 2021, 140, 65–99. https://doi.org/10.48550/arXiv.2012.00931

[2] Jain, R.; Barcovschi, A.; Yiwere, M.; Corcoran, P.; Cucu, H. Adaptation of Whisper models to child speech recognition. arXiv preprint, 2023. arXiv:2307.13008.

[3] Gulati, A.; Qin, J.; Chiu, C.C.; Parmar, N.; Zhang, Y.; Yu, J.; Han, W.; Wang, S.; Zhang, Z.; Wu, Y. Conformer: Convolution-augmented transformer for speech recognition. Proceedings of Interspeech 2020, 2020, pp. 5036–5040. https://doi.org/10.21437/Interspeech.2020-3015

[4] Rakhimova, D.; Duisenbekkyzy, Zh.; Karibayeva, A.; Eşref, A.; Ilessova, B. Improving the Voice Recognition System for Children in Kazakh through Additional Training (Fine-Tuning). Academic Scientific Journal of Computer Science, 2026, No. 2, pp. 317–334. https://doi.org/10.32014/2026.2518-1726.441

[5] Gretter, R.; Matassoni, M.; Falavigna, D.; Misra, A.; Leong, C.W.; Knill, K.; Wang, L. ETLT 2021: Shared task on automatic speech recognition for non-native children's speech. Proceedings of Interspeech 2021, 2021.

[6] Rusu, A.A.; Rabinowitz, N.C.; Desjardins, G.; et al. Progressive Neural Networks. arXiv preprint, 2016. https://arxiv.org/abs/1606.04671

[7] Wang, D.; Wang, X.; Lv, S. An overview of end-to-end automatic speech recognition. Symmetry, 2019, 11(8), 1018. https://doi.org/10.3390/sym11081018

[8] Xiong, W.; Wu, L.; Alleva, F.; Droppo, J.; Huang, X.; Stolcke, A. The Microsoft 2017 conversational speech recognition system. Proceedings of ICASSP 2018, 2018, pp. 5934–5938. https://doi.org/10.48550/arXiv.1708.06073

[9] Bengio, Y.; Goodfellow, I.; Courville, A. Deep Learning. MIT Press, 2016. https://www.deeplearningbook.org/

[10] Koo, J.; Kim, H.; Lee, S. Fine-tuning Whisper for low-resource languages. arXiv preprint, 2023. https://arxiv.org/abs/2303.00752

[11] Rakhimova, D.; Duisenbekkyzy, Zh.; Adali, E. Investigation of ASR Models for Low-Resource Kazakh Child Speech: Corpus Development, Model Adaptation, and Evaluation. Applied Sciences, 2025, 15(16), 8989. https://doi.org/10.3390/app15168989

[12] Duisenbekkyzy, Zh.; Rakhimova, D. Research on the Impact of Social Networks on the Development of Bilingualism in Kazakhstan. Communications in Computer and Information Science, 2024, 2211, pp. 106–115. https://doi.org/10.1007/978-3-031-49739-5_9

МРНТИ 81.93.29

https://doi.org/10.51889/2959-5894.2026.95.3.017

Н.С. Ергеш[1*], Н.А. Капалова[1]

[1] Институт информационных и вычислительных технологий, г. Алматы, Казахстан
[*]*e-mail: yergesh.nursultan@almatyedu.kz*

МОДИФИЦИРОВАННАЯ СХЕМА РАЗДЕЛЕНИЯ СЕКРЕТА ДЛЯ ЗАЩИТЫ ЛИЦЕВЫХ БИОМЕТРИЧЕСКИХ ШАБЛОНОВ

Аннотация

Лицевой биометрический шаблон – постоянный идентификатор: при компрометации его нельзя перевыпустить, поэтому хранение защищённого шаблона в единственном узле образует критическую точку отказа. Работа продолжает начатое авторами направление защиты лицевых шаблонов пороговым разделением секрета и посвящена модифицированному алгоритму, ориентированному именно на лицевую биометрию. За основу взята верифицируемая схема на многомерных линейных функциях, предложенная ранее для общих данных. Её правило перестроено по трём направлениям: псевдослучайная матрица коэффициентов заменена детерминированным вандермондовым правилом; введены блочная упаковка защищённой записи и маскирование; хеш-проверка заменена безусловно стойким одноразовым кодом Картера-Вегмана. Полученная блочно-упакованная схема (БУ-СРС) обладает свойством MDS, совершенной секретностью до порога t, точно ограниченной рамп-утечкой и безусловной проверкой долей; все свойства доказаны. На воспроизводимой реализации при n = 7, k = 5, L = 3 все 6300 проверок восстановлены безошибочно; над малым полем модифицированная схема не дала вырожденных подмножеств против 0,42 % у исходной; рамп-утечка совпала с теорией. Для 256-байтовой записи упаковка сокращает объём долей в 64 раза, а метаданные проверки -вдвое; безусловная проверка не пройдена ни в одной из 200 000 попыток подделки. Точность распознавания наследуется от фронтенда без изменения (восстановление бит-в-бит) и по протоколу LFW составляет EER 2,12 % ± 0,57 %.

Ключевые слова: защита биометрических шаблонов; лицевая биометрия; пороговое разделение секрета; блочная упаковка; вандермондова матрица; совершенная секретность; безусловная проверка долей; BioHashing.

Н.С. Ергеш[1], Н.А. Қапалова[1]

[1] Ақпараттық және есептеуіш технологиялар институты, Алматы қ., Қазақстан

БЕТ-ӘЛПЕТ БИОМЕТРИЯЛЫҚ ШАБЛОНДАРЫН ҚОРҒАУҒА АРНАЛҒАН МОДИФИКАЦИЯЛАНҒАН ҚҰПИЯНЫ БӨЛУ СХЕМАСЫ

Аңдатпа

Бет-әлпет биометриялық шаблоны -тұрақты идентификатор: әшкереленген шаблонды қайта шығару мүмкін емес, сондықтан оны бір түйінде сақтау сыни ақаулық нүктесін құрайды. Жұмыс авторлар бастаған бет-әлпет шаблондарын пороговый құпияны бөлумен қорғау бағытының жалғасы болып, дәл бет-әлпет биометриясына арналған модификацияланған алгоритмге бағытталған. Негізге жалпы деректерге ұсынылған, көп айнымалылы сызықтық функцияларға негізделген үлестерді тексеретін схема алынды. Оның ережесі үш бағытта өзгертілді: псевдокездейсоқ коэффициенттер матрицасы детерминацияланған вандермонд ережесімен алмастырылды; блоктық орау мен маскалау енгізілді; хеш-тексеру Картер-Вегманның шартсыз тұрақты бір реттік кодымен ауыстырылды. Алынған блоктық-оралған схема (БУ-СРС) MDS қасиетіне ие, t табалдырығына дейін мінсіз құпиялылықты, дәл шектелген рамп-ағуды және үлестерді шартсыз тексеруді қамтамасыз етеді; барлық қасиеттер дәлелденген. Тәжірибеде n = 7, k = 5, L = 3 болғанда барлық 6300 тексеру қатесіз қалпына келтірілді; шағын өрісте модификацияланған схема бірде-бір азғындаған ішкі жиын бермеді (бастапқыда 0,42 %); рамп-ағу теориямен сәйкес келді. 256 байттық жазба үшін орау үлестердің көлемін 64 есе, тексеру метадеректерін екі есе қысқартады; шартсыз тексеру 200 000 бұрмалау әрекетінің бірде-бірінде өтпеді. Тану дәлдігі фронтендтен өзгеріссіз мұраланады және LFW хаттамасы бойынша EER 2,12 % ± 0,57 % құрайды.

Түйін сөздер: биометриялық шаблондарды қорғау; бет-әлпет биометриясы; пороговый құпияны бөлу; блоктық орау; вандермонд матрицасы; мінсіз құпиялылық; үлестерді шартсыз тексеру; BioHashing.

N.S. Yergesh[1], N.A. Kapalova[1]
1 Institute of Information and Computational Technologies, Almaty, Kazakhstan

A MODIFIED SECRET SHARING SCHEME FOR PROTECTING FACIAL BIOMETRIC TEMPLATES

Abstract

A facial biometric template is a permanent identifier: once exposed it cannot be reissued, so storing the protected template in a single node is a critical point of failure. This work continues the authors' line of research on protecting facial templates with threshold secret sharing and develops a modified algorithm oriented specifically to facial biometrics. It builds on a share-verifiable scheme based on multivariable linear functions previously proposed for general data. Its rule is rebuilt in three ways: the pseudo-random coefficient matrix is replaced by a deterministic Vandermonde rule; blockwise packing and masking are introduced; and the hash-based check is replaced by an unconditionally secure one-time Carter-Wegman code. The resulting block-packed scheme (BP-SS) is MDS, perfectly secret up to the threshold t, has exactly bounded ramp leakage and unconditional share verification; all properties are proven. On a reproducible implementation with n = 7, k = 5, L = 3, all 6300 checks reconstructed without error; over a small field the modified scheme produced no singular subsets versus 0.42 % for the original; ramp leakage matched theory. For a 256-byte record, packing reduces share storage 64-fold and halves verification metadata; unconditional verification was never bypassed across 200,000 forgery attempts. Recognition accuracy is inherited from the front end unchanged (bit-identical recovery) and, on the LFW protocol, reaches an EER of 2.12 % ± 0.57 %.

Keywords: biometric template protection; facial biometrics; threshold secret sharing; block packing; Vandermonde matrix; perfect secrecy; unconditional share verification; BioHashing.

Введение

Основные положения

1. Предложена модификация верифицируемой схемы разделения секрета на многомерных функциях, целенаправленно адаптированная к лицевой биометрии за счёт блочной упаковки защищённого шаблона и структурированного вандермондова правила коэффициентов.

2. Строго доказаны четыре свойства: MDS-восстановление любыми k из n долей, совершенная секретность до порога t, точная оценка утечки в рамп-режиме и безусловная стойкость проверки долей.

3. На воспроизводимой реализации подтверждены заявленные преимущества и приведено измеренное сравнение с базовыми пороговыми схемами; точность распознавания фронтенда (EER 2,12 % ± 0,57 % на LFW) слоем защиты не изменяется.

Распознавание лица стало одним из наиболее распространённых способов биометрической аутентификации: получение образца бесконтактно, удобно и хорошо поддерживается современными нейросетевыми конвейерами. Однако защита хранимых лицевых шаблонов принципиально отличается от защиты паролей. Пароль можно отозвать и перевыпустить после утечки, тогда как лицевой шаблон порождается физической характеристикой, которую человек изменить не может. Компрометация хранимого представления лица имеет необратимые последствия: одну и ту же личность можно неограниченно долго выдавать за легитимную, а утёкший шаблон -связывать между независимыми системами. Поэтому защита лицевых шаблонов -вопрос не только конфиденциальности, но и необратимого персонального риска [1, 2].

Конкретная научная проблема состоит в следующем. Типичная слабость развёрнутых систем -хранение защищённого шаблона единственным объектом в единственном месте: даже при шифровании такая архитектура концентрирует весь риск на одном узле и одном ключе. Пороговое разделение секрета устраняет эту уязвимость: секрет делится на доли так, что любое пороговое их число восстанавливает секрет, а меньшее в идеале бесполезно [3, 4].

Однако известные схемы, применяемые к биометрии, имеют три практических ограничения. Во-первых, у схем со случайными коэффициентами гарантия восстановления лишь вероятностная: выбранная k×k-подсистема может оказаться вырожденной. Во-вторых, универсальные схемы применяются к шаблону побайтно, что многократно раздувает объём хранимых долей -критично для крупных лицевых шаблонов и ресурсоограниченных узлов. В-третьих, проверка целостности долей опирается на хеш-функции, то есть даёт лишь вычислительную стойкость. Именно эти три недостатка устраняет предлагаемая модификация.

Настоящее исследование продолжает начатое авторами направление защиты биометрических данных пороговыми методами [5, 6]. Выстроен сквозной конвейер: детектирование и выравнивание лица, глубокий эмбеддинг, необратимое преобразование BioHashing, аутентифицированное шифрование и пороговое разделение шифртекста на пространственно разнесённые узлы. За основу слоя раздачи взята верифицируемая схема на многомерных линейных функциях [7], предложенная для общих данных и не рассчитанная на биометрию; в базовой конфигурации она применялась побайтно.

Цель работы -разработка математически перестроенного алгоритма разделения секрета, ориентированного именно на лицевую биометрию и устраняющего перечисленные ограничения. Правило схемы [7] изменяется по трём направлениям: (i) псевдослучайная матрица коэффициентов заменяется детерминированным вандермондовым правилом (доказуемое MDS, отсутствие вырожденных подмножеств); (ii) вводится блочная упаковка защищённой записи (резкое сокращение объёма хранения); (iii) хеш-проверка заменяется безусловно стойким одноразовым кодом Картера-Вегмана. Полученную схему мы называем блочно-упакованной схемой разделения секрета (БУ-СРС). Задачи: сформулировать правило раздачи; доказать корректность (MDS), совершенную секретность, точную рамп-утечку и безусловность проверки; реализовать схему на воспроизводимом стенде и подтвердить свойства количественно; сопоставить с исходной и современными схемами.

Методология исследования

Обоснование методов и этапы исследования

Выбор методов определяется предметом и задачами. Пороговое разделение секрета выбрано как класс механизмов, напрямую устраняющий одноузловую точку отказа. Вандермондово правило выбрано потому, что определитель матрицы Вандермонда на попарно различных узлах отличен от нуля, что превращает вероятностную гарантию восстановления в доказуемую (MDS). Блочная упаковка выбрана как способ согласовать размер крупного лицевого шаблона с элементом большого поля и снизить объём долей. Одноразовый код Картера-Вегмана выбран потому, что даёт информационно-теоретическую (безусловную) стойкость проверки, недостижимую для хеш-проверки. Исследование выполнено в шесть этапов: анализ ограничений базовой схемы; математическая модификация правила раздачи; формальное доказательство свойств; воспроизводимая реализация; экспериментальная проверка; сравнительный анализ.

Обзор конвейера и роль модификации

Конвейер преобразует изображение лица в набор распределённых долей. Лицо детектируется и выравнивается; глубокая сеть отображает выровненное лицо в эмбеддинг; BioHashing проецирует и бинаризует его в возобновляемый шаблон [8]; шаблон запечатывается аутентифицированным шифрованием; защищённая запись делится пороговой схемой на доли, размещаемые на разнесённых узлах. Первые четыре стадии составляют общий фронтенд и не зависят от модальности; вклад работы -последняя стадия.

Фронтенд: BioHashing и аутентифицированное шифрование

Пусть x -вещественный эмбеддинг размерности d. BioHashing выбирает пользовательскую псевдослучайную ортонормированную матрицу R и пороговой функцией формирует

бинарный шаблон b длины m (формула 1); при m < d преобразование необратимо, а замена R даёт несвязываемый шаблон [9].

$$b_i = 1, \text{ если } (Rx)_i \geq \theta_i; \text{ иначе } b_i = 0, \quad i = 1, \ldots, m \tag{1}$$

Бинарный шаблон запечатывается аутентифицированным шифрованием (AEAD) с ключом K, хранимым отдельно от долей (формула 2).

$$(C, T) = Enc_K(b, A) \tag{2}$$

Существенно, что далее разделяется шифртекст, а не биометрический шаблон, и ключ K ни в одну долю не помещается: любая частичная информация ниже порога касается статистически равномерных байтов шифртекста, а не признаков лица.

Правило блочного разбиения

Пусть q -большое простое число (в реализации $q = 2^{127} - 1$, размер элемента w = 16 байт). Защищённая запись M разбивается на L блоков -элементов поля (формула 3); именно они совместно упаковываются в один многочлен, что отличает модификацию от побайтовой раздачи.

$$M = s_1 \| s_2 \| \ldots \| s_L, \quad s_m \in Z_q \tag{3}$$

Модифицированное правило раздачи (БУ-СРС)

Параметры: n узлов, порог конфиденциальности t, упаковка L, порог восстановления k = t + L. Коэффициенты задаются детерминированно, а секрет упаковывается в многочлен степени k -1 (формула 4).

$$P(x) = c_0 + c_1 x + \ldots + c_{k-1} x^{k-1} \tag{4}$$

Многочлен задаётся k условиями: L закрепляют блоки в секретных позициях β_m, а t вносят случайное маскирование в позициях γ_r (формула 5); все точки попарно различны и ненулевые.

$$P(\beta_m) = s_m; \quad P(\gamma_r) = \rho_r \leftarrow Z_q \tag{5}$$

Доля узла i -значение многочлена в публичной точке x_i (формула 6).

$$y_i = P(x_i) \bmod q, \quad i = 1, \ldots, n \tag{6}$$

Восстановление

Любые k долей однозначно задают многочлен степени k -1; блоки извлекаются интерполяцией Лагранжа в секретных точках (формула 7).

$$s_m = \Sigma_{i \in I} y_i \prod_{j \neq i} (\beta_m - x_j)/(x_i - x_j) \bmod q \tag{7}$$

Безусловная проверка долей (Картер-Вегман)

Вместо хеш-проверки применяется одноразовый аффинный код [10]: из ключа SK_i выводится (α_i, δ_i) и вычисляется тег (формула 8); проверка -одно сравнение, стойкость безусловна.

$$\tau_i = \alpha_i y_i + \delta_i \bmod q \tag{8}$$

Свойства схемы и их доказательства

Теорема 1 (корректность, MDS). При попарно различных ненулевых x_i любые k из n долей однозначно восстанавливают все блоки. *Доказательство. Многочлен степени k -1 определяется значениями в любых k различных точках, так как матрица Вандермонда*

невырождена; значения $P(\beta_m)$ дают блоки, а избыточность n -k допускает потерю любых n - k узлов. ■

Теорема 2 (совершенная секретность до порога t). Любые t долей равномерно распределены на Z_q^t и не зависят от секрета. *Доказательство. Маскирующие ρ_r равномерны и независимы, поэтому t долей суть вандермондова биекция маскирования; отсюда равномерность и независимость от секрета.* ■

Теорема 3 (точная рамп-утечка). При m долях, $t \le m \le k$, число совместимых секретных векторов равно q^{k-m}, а раскрываемая информация равна ровно $(m - t)\cdot\log_2 q$ бит (формула 9).

$$I(S;\ доли_m) = (m - t)\cdot\log_2 q, \quad t \le m \le k \tag{9}$$

Теорема 4 (безусловная стойкость проверки). При одноразовом равномерном ключе (α_i, δ_i) подделанная доля $y' \neq y_i$ проходит проверку с вероятностью не выше 1/q (формула 10), независимо от вычислительной мощности. *Доказательство. Корректный тег есть $\tau' = \alpha_i y' + \delta_i$; так как α_i неизвестно и равномерно, а $(y' - y_i)$ обратимо, значение τ' равномерно на Z_q.* ■

$$Pr[подделка] \le 1/q \tag{10}$$

Иллюстративный пример (малое поле)

Пусть q = 251, L = 2, t = 2, k = 4, блоки (123, 200). Дилер строит многочлен степени 3, закрепляя два блока и два случайных значения. При двух долях совместим любой из $251^2 = 63001$ секретных векторов; при трёх -ровно 251; при четырёх -единственный (подтверждено экспериментом, см. Результаты).

Модель угроз

Система включает четыре сущности: клиент регистрации, вычисляющий защищённую запись и доли; n узлов хранения, честных-но-любопытных и по отдельности не доверенных (противник может скомпрометировать до k -1 из них); элемент управления ключами, хранящий токен BioHashing, ключ K и мастер-ключ проверки и не размещаемый совместно с долями; и верификатор, собирающий кворум, проверяющий теги, восстанавливающий и расшифровывающий запись. Противник может читать доли до k -1 узлов, подделывать или искажать доли, воспроизводить записи и действовать как инсайдер, но не владеет K и ключом проверки и не обращает BioHashing. Сценарии: чтение ниже порога (годного шаблона нет; раскрываются лишь ограниченные байты шифртекста); компрометация ключа/токена без долей (записи нет); сговор до k -1 узлов (восстановления нет, подделки выявляются тегами); воспроизведение и отказ до n -k узлов (доступность сохраняется избыточностью (k, n)). При потере меньшинства узлов запись восстанавливается по оставшемуся кворуму.

Стенд, платформа и параметры экспериментов

Схема реализована на воспроизводимом стенде на языке Python 3.12.3 (одноядерное исполнение x86-64) с арифметикой в поле Z_q при $q = 2^{127} - 1$; для полного перебора секретности и вырожденности использовано малое поле Z_{251}. Все запуски выполнены с фиксированным seed = 42; исходный код, скрипты экспериментов и файлы числовых результатов (results_busrs.json, results_compare.json) публикуются вместе со статьёй. Времена измерены как среднее по 2000 повторениям; абсолютные значения зависят от оборудования, соотношения и асимптотика -нет.

Точность распознавания оценивается отдельно и является свойством фронтенда. Поскольку слой раздачи не искажает шаблон (восстановление бит-в-бит, см. Результаты), точность наследуется без изменения; операционная точка приведена по опубликованному авторами результату на стандартном протоколе верификации лиц LFW (View 2; 6000 пар в 10 фолдах по 600) с детектированием MTCNN, выравниванием по пяти точкам и глубоким эмбеддингом типа ArcFace [11, 12].

Результаты исследования

Основные результаты -доказуемое MDS-восстановление, совершенная секретность до порога, точная рамп-утечка, безусловная проверка и многократное сокращение объёма хранения; промежуточные -измерения быстродействия и подтверждение независимости точности распознавания от слоя защиты. Все числовые значения получены на воспроизводимой реализации.

Корректность восстановления и отказоустойчивость (MDS)

При n = 7, k = 5, L = 3 все C(7,5) = 21 подмножеств в каждом из 300 испытаний (итого 6300 проверок) дали безошибочное восстановление (таблица 1), что подтверждает теорему 1.

Таблица 1. Корректность и MDS-восстановление над полем 2^{127} -1.

Показатель	*Значение*	*Успех*
Испытаний (n=7, k=5, L=3)	*300*	*-*
Проверок по всем 5-из-7 подмножествам	*6300 / 6300*	*100 %*

Гарантия восстановления: MDS против случайных коэффициентов

В исходной схеме случайные коэффициенты не гарантируют невырожденности k×k-подсистемы: над Z251 доля вырожденных 3-из-5 подмножеств составила 0,42 % (127 из 30 000; теория ≈ 1/q ≈ 0,40 %), тогда как вандермондово правило не дало ни одного вырожденного подмножества (таблица 2).

Таблица 2. Вырожденные (не восстанавливаемые) k-подмножества над Z_{251}.

Схема (над Z_{251})	*Вырожденных*	*Доля*
Исходная (случайные коэфф.), 30 000 пров.	*127*	*0,42 %*
Модифицированная (Вандермонд)	*0*	*0 % (доказуемо)*

Совершенная секретность и рамп-утечка

Прямой перебор над Z251 при L = 2, t = 2, k = 4 подтвердил теоремы 2 и 3: при t = 2 совместимы все q^2 = 63001 векторов (нет информации), при 3 долях -ровно q = 251, при 4 - единственный (таблица 3, рисунок 1).

Таблица 3. Число секретных векторов, совместимых с наблюдением (Z_{251}, L=2, t=2, k=4).

Число долей m	*Совместимых (экспер.)*	*Теория q^{k-m}*
t = 2	*63001*	*63001*
t + 1 = 3	*251*	*251*
k = 4	*1*	*1*

Быстродействие

Времена (мс на операцию, поле 2^{127} -1, среднее по 2000 повторениям) приведены в таблице 4. Восстановление асимптотически дешевле -$O(k^2)$ интерполяции против $O(k^3)$ метода Гаусса у исходной схемы; безусловная проверка выполняется за одно умножение-сложение.

Таблица 4. Быстродействие на операцию и асимптотическая сложность.

Операция	*Время, мс*	*Асимптотика*
Раздача (split)	*0,58785*	*$O(k^2+nk)$*
Восстановление (reconstruct)	*0,3594*	*$O(k^2)$*
Проверка Картера-Вегмана	*0,000467*	*O(1)*

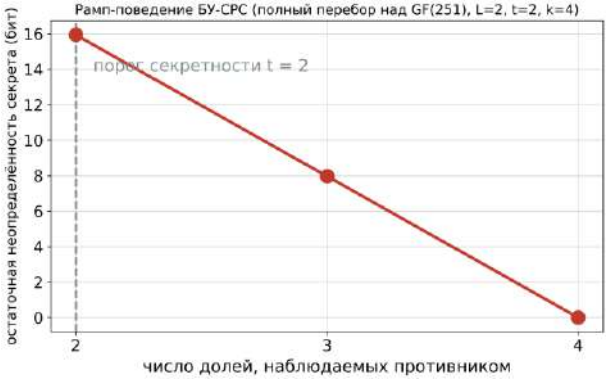


Рисунок 1. Остаточная неопределённость секрета: плато совершенной секретности до порога t и линейное убывание в рамп-зоне.

Объём хранения

Для записи D = 256 байт (n = 5, L = 3, элемент 16 байт) побайтовая раздача исходной схемы порождает 256 многочленов и 61 440 байт долей; блочная упаковка использует 6 многочленов и 960 байт -64-кратное сокращение (таблица 5, рисунок 2); метаданные проверки уменьшаются вдвое (16 Б против 32 Б).

Таблица 5. Объём хранения долей для 256-байтовой лицевой записи.

Показатель (D = 256 Б)	*Исходная*	*Модифиц.*
Число многочленов	*256*	*6*
Метаданные проверки / долю	*32 Б*	*16 Б*
Суммарный объём долей	*61 440 Б*	*960 Б*
Сокращение	-	*64×*

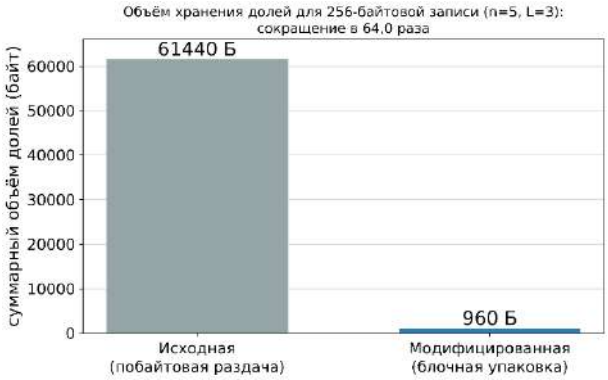


Рисунок 2. Суммарный объём долей для 256-байтовой записи: побайтовая раздача против блочной упаковки.

Стойкость проверки к подделке

В 200 000 испытаниях подделки доли с неизвестным одноразовым ключом ни одна не прошла проверку (эмпирическая доля 0), что согласуется с границей $1/q \approx 5{,}9 \cdot 10^{-39}$ (теорема 4). В отличие от хеш-проверки, граница не зависит от криптографических предположений. Нулевой результат при конечном числе испытаний согласуется с границей, но сам её не доказывает; безусловность следует из доказательства.

Независимость точности распознавания от слоя защиты

Восстановленный после сбора k долей шаблон совпадает с исходным бит в бит (максимальное несовпадение блоков -0 на 1 000 × 21 восстановлений), поэтому кривая ошибок

распознавания до и после слоя защиты тождественна. По опубликованному результату на LFW (View 2; 6000 пар в 10 фолдах по 600; MTCNN, выравнивание по пяти точкам, эмбеддинг типа ArcFace) конвейер достигает EER = 2,12 % ± 0,57 % (95 % ДИ [1,77 %; 2,47 %]) при базовой линии 1,37 %. Таким образом, модифицированный слой добавляет доказуемые гарантии и 64-кратную экономию хранения, не жертвуя точностью.

Дискуссия

Сравнение с исходной схемой.

Модификация отличается от исходной схемы [7] по пяти осям (таблица 6).

Таблица 6. Сопоставление свойств исходной и модифицированной схем.

Свойство	*Исходная [7]*	*БУ-СРС*
Восстановление любых k из n	*вероятностное*	*доказуемое (MDS)*
Секретность ниже порога	*частично вычислит.*	*совершенная (доказано)*
Проверка долей	*хеш (вычислит.)*	*безусловная (≤ 1/q)*
Рамп-режим	*нет*	*да, точная оценка*
Объём хранения	*побайтно, высокий*	*упаковка, до 64×*

Измеренное сравнение с пороговыми схемами

Для одной и той же 256-байтовой записи (n = 5, поле 2^{127} -1) на общей платформе измерены стандартная схема Шамира, рамп-схема Шамира и БУ-СРС (таблица 7). Стандартный Шамир не пакует и не проверяет доли; рамп-Шамир пакует, но не проверяет; БУ-СРС сочетает упаковку с безусловной проверкой, что объясняет паритет по объёму со стандартным Шамиром при вдвое меньшем числе многочленов.

Схема Фельдмана [13] и решёточные постквантовые конструкции [14] приведены по литературе; их экспериментальное сравнение на едином стенде вынесено в дальнейшую работу. Смежные современные направления -универсальные тресхолдайзеры [15], блокчейн-распределённое хранение ключей [16] и групповые многосекретные схемы [17] -намечают направления такого сравнения.

Таблица 7. Измеренное сравнение схем для 256-байтовой записи (реализованные строки измерены; Фельдман и решёточные -по литературе).

Схема	*Б/долю*	*Раздача, мс*	*Восст., мс*	*Проверка*	*Утечка < порога*
Стандартный Шамир	*288*	*0,1264*	*0,6112*	*нет*	*нет (соверш.)*
Рамп-Шамир (упаковка)	*144*	*1,467*	*0,6051*	*нет*	*рамп*
БУ-СРС (наша)	*288*	*1,5622*	*0,5957*	*безусловная*	*рамп (проверяемый)*
Фельдман VSS [13]	*знач.+коммит.*	-	-	*вычислит.*	*нет (соверш.)*
Решёточные [14]	*варьир.*	-	-	*по источн.*	*по источн.*

Компромиссы и научная значимость

Упаковка L > 1 повышает эффективность хранения ценой рамп-зазора: строгая секретность гарантирована до t = k -L, а в промежутке t < m < k утечка растёт на $\log_2 q$ бит с каждой долей (теорема 3). Эта утечка относится к статистически равномерным байтам шифртекста, а не к биометрии; ключ K хранится отдельно, поэтому частично раскрытые байты бесполезны. При L = 1 схема вырождается в совершенную.

Научная значимость результата в том, что для защиты постоянного идентификатора -лица -предложен слой раздачи, одновременно устраняющий три ограничения известных схем (вероятностное восстановление, раздувание объёма, вычислительная проверка), являющийся прямым развитием ранее опубликованных результатов авторов [5-7] и переносимый на другие модальности без изменения доказательств.

Заключение

Предложена и обоснована модифицированная схема разделения секрета (БУ-СРС), ориентированная на защиту лицевых биометрических шаблонов и продолжающая направление авторов [5-7]. Правило исходной верифицируемой схемы изменено по трём направлениям: детерминированное вандермондово правило, блочная упаковка и безусловная алгебраическая проверка.

Строго доказаны MDS, совершенная секретность до порога, точная рамп-утечка и безусловность проверки. На воспроизводимой реализации подтверждены безошибочное восстановление по всем подмножествам (6300/6300), отсутствие вырожденных подмножеств, совпадение утечки с теорией, 64-кратное сокращение объёма хранения и отсутствие успешных подделок (0 из 200 000); измеренное сравнение с пороговыми схемами показало паритет по объёму со стандартным Шамиром при добавленной безусловной проверке и упаковке. Точность распознавания фронтенда (EER 2,12 % ± 0,57 % на LFW) слоем защиты не изменяется. Главный компромисс - рамп-зазор упаковки, управляемый параметром L.

Дальнейшая работа: адаптивный выбор L, перенос слоя на постквантовые примитивы с экспериментальным сравнением на едином стенде и расширенная межбазовая оценка распознавания.

Благодарность

Исследование выполнено при поддержке Комитета науки Министерства науки и высшего образования Республики Казахстан, грант №AP23487787 «Разработка алгоритма надежного хранения биометрических данных систем аутентификации в пространственно-распределенных хранилищах».

Список использованных источников

[1] Jain A.K., Nandakumar K., Nagar A. Biometric template security // EURASIP Journal on Advances in Signal Processing. -2008. -Vol. 2008. -Art. 579416. -doi:10.1155/2008/579416.

[2] Rathgeb C., Uhl A. A survey on biometric cryptosystems and cancelable biometrics // EURASIP Journal on Information Security. -2011. -Vol. 2011. -Art. 3. -doi:10.1186/1687-417X-2011-3.

[3] Shamir A. How to share a secret // Communications of the ACM. -1979. -Vol. 22, No. 11. -P. 612-613. -doi:10.1145/359168.359176.

[4] Blakley G.R. Safeguarding cryptographic keys // Proceedings of the AFIPS National Computer Conference. -1979. -Vol. 48. -P. 313-317. -doi:10.1109/AFIPS.1979.98.

[5] Nyssanbayeva S., Kapalova N., Beisenova S. Modified Shamir threshold scheme for secure storage of biometric data // Computers. -2026. -Vol. 15, No. 4. -Art. 228. -doi:10.3390/computers15040228.

[6] Kapalova N., Algazy K., Haumen A. Development of a new lightweight encryption algorithm // Eastern-European Journal of Enterprise Technologies. -2023. -Vol. 3, No. 9(123). -P. 6-19. -doi:10.15587/1729-4061.2023.281862.

[7] Kapalova N., Haumen A., Algazy K. Secret sharing scheme with share verification capability // Computers. -2025. -Vol. 14, No. 9. -Art. 393. -doi:10.3390/computers14090393.

[8] Teoh A.B.J., Ngo D.C.L., Goh A. Biohashing: two factor authentication featuring fingerprint data and tokenised random number // Pattern Recognition. -2004. -Vol. 37, No. 11. -P. 2245-2255. -doi:10.1016/j.patcog.2004.04.011.

[9] ISO/IEC 24745:2022. Information security, cybersecurity and privacy protection -Biometric information protection. -Geneva: ISO, 2022.

[10] Wegman M.N., Carter J.L. New hash functions and their use in authentication and set equality // Journal of Computer and System Sciences. -1981. -Vol. 22, No. 3. -P. 265-279. -doi:10.1016/0022-0000(81)90033-7.

[11] Deng J., Guo J., Xue N., Zafeiriou S. ArcFace: additive angular margin loss for deep face recognition // Proceedings of the IEEE/CVF CVPR. -2019. -P. 4690-4699. -doi:10.1109/CVPR.2019.00482.

[12] Schroff F., Kalenichenko D., Philbin J. FaceNet: a unified embedding for face recognition and clustering // Proceedings of the IEEE CVPR. -2015. -P. 815-823. -doi:10.1109/CVPR.2015.7298682.

[13] Feldman P. A practical scheme for non-interactive verifiable secret sharing // Proceedings of the 28th Annual Symposium on Foundations of Computer Science (FOCS). -1987. -P. 427-438. -doi:10.1109/SFCS.1987.4.

[14] Chen J., Deng H., Su H., Yuan M., Ren Y. Lattice-based threshold secret sharing scheme and its applications: a survey // Electronics. -2024. -Vol. 13, No. 2. -Art. 287. -doi:10.3390/electronics13020287.

[15] Cheon J.H., Cho W., Kim J. Improved universal thresholdizer from iterative Shamir secret sharing // Journal of Cryptology. -2025. -Vol. 38. -Art. 15. -doi:10.1007/s00145-025-09540-x.

[16] Zhou K., Au M.H., Liu J. A blockchain-based Shamir's threshold cryptography scheme for data protection in industrial IoT settings // IEEE Internet of Things Journal. -2022. -Vol. 9, No. 11. -P. 8714-8726. -doi:10.1109/JIOT.2021.3124825.

[17] Miao F., Wang L., Ji Y., Xiong Y. GOMSS: a simple group oriented (t, m, n) multi-secret sharing scheme // Chinese Journal of Electronics. -2017. -Vol. 26, No. 3. -P. 557-563. -doi:10.1049/cje.2016.08.007.

МРНТИ 28.23.25

https://doi.org/10.51889/2959-5894.2026.95.3.018

M. Junussova[1]*, D. Rakisheva[1]

[1]L.N. Gumilyov Eurasian National University, Astana, Kazakhstan
[*]*e-mail: m.junussova.math@gmail.com*

MACHINE LEARNING FOR RESERVOIR ANALYSIS AND FORECASTING IN KAZAKHSTAN

Abstract

This paper explores the application of machine learning methods for the analysis, classification, and forecasting of the dynamics of small reservoirs in Kazakhstan based on long-term satellite observations. The study covers 71 water bodies for which archival satellite imagery from the Landsat 7, Landsat 8, Landsat 9, and Sentinel-2 missions was collected and processed from 2000 to 2024. The water surface area was estimated using the Modified Normalized Difference Water Index (MNDWI) combined with image segmentation algorithms implemented in Matlab, ArcGIS, and Python environments. Based on the derived estimates, time series reflecting seasonal and interannual variations in reservoir surface area were constructed. To identify characteristic types of water-body dynamics, the time series were classified according to seasonal patterns, long-term trends, and interannual variability. In addition, time series analysis was performed using forecasting models, including Prophet and Long Short-Term Memory (LSTM) recurrent neural networks, enabling short-term predictions of water surface area changes for reservoirs of different sizes. The results demonstrate the potential of integrating satellite data with machine learning methods to develop automated monitoring systems and decision-support tools for water resource management in the Republic of Kazakhstan.

Keywords: remote sensing, dam monitoring, machine learning, water reservoirs, data analysis, LSTM.

М.Ж. Джунусова[1], Д.С. Ракишева[1]

[1]Л.Н. Гумилев атындағы Еуразия ұлттық университеті, Астана қ., Қазақстан

ҚАЗАҚСТАНДАҒЫ СУ ҚОЙМАЛАРЫН ТАЛДАУ ЖӘНЕ БОЛЖАУ ҮШІН МАШИНАЛЫҚ ОҚЫТУ ӘДІСТЕРІН ҚОЛДАНУ

Аңдатпа

Бұл жұмыста Қазақстандағы кіші су қоймаларының динамикасын талдау, классификациялау және болжау үшін машинамен оқыту әдістерін қолдану қарастырылған. Зерттеу 2000–2024 жылдар аралығында Landsat 7, 8, 9 және Sentinel-2 миссияларының ғарыштық суреттер архивтері жиналып, өңделген 71 су қоймасын қамтиды. Судың беткі аймағы модификацияланған қалыпты айырмашылықты су индексы (MNDWI) және Matlab, ArcGIS және Python бағдарламалық ортада іске асырылған суреттерді сегменттеу алгоритмдері арқылы бағаланды. Алынған мәліметтер негізінде су қоймаларының беткі аймағының маусымдық және жылдар аралық динамикасын көрсететін уақыттық қатарлар құрылды. Су объектілерінің өзгеру сипаттарын анықтау үшін уақыттық қатарлар маусымдық ерекшеліктері, трендтер және жылдар аралық өзгергіштік бойынша классификацияланды. Сонымен қатар, уақыттық қатарларды Prophet және LSTM түріндегі рекуррентті нейрондық желілерді қолданып талдау жүргізіліп, әртүрлі масштабтағы су қоймаларының беткі аймағының қысқа мерзімді болжаулары алынған. Алынған нәтижелер спутниктік деректер мен машинамен оқыту әдістерін қолданудың автоматтандырылған мониторинг жүйелерін құру және Қазақстан Республикасының су ресурстарын басқару саласында шешім қабылдауды қолдауға әлеуетін көрсетеді.

Түйін сөздер: қашықтықтан зондтау, су қоймаларын бақылау, машинамен оқыту, су қоймалары, деректерді талдау, LSTM.

М.Ж. Джунусова[1], Д.С. Ракишева[1]
[1]Евразийский национальный университет имени Л.Н. Гумилева, г.Астана, Казакстан

МАШИННОЕ ОБУЧЕНИЕ ДЛЯ АНАЛИЗА И ПРОГНОЗИРОВАНИЯ ВОДОХРАНИЛИЩ КАЗАХСТАНА

Аннотация

В работе рассматривается применение методов машинного обучения для анализа, классификации и прогнозирования динамики малых водохранилищ Казахстана на основе многолетних спутниковых данных. Исследование выполнено для 71 водоёма, для которых за период 2000–2024 гг. были собраны и обработаны архивы спутниковых изображений миссий Landsat 7, 8, 9 и Sentinel-2. Площадь водной поверхности оценивалась с использованием модифицированного нормализованного водного индекса (MNDWI) и алгоритмов сегментации изображений, реализованных в программных средах Matlab, ArcGIS и Python. На основе полученных данных сформированы временные ряды, отражающие сезонную и межгодовую динамику площади водоёмов. Для выявления характерных типов изменения водных объектов выполнена классификация временных рядов по особенностям сезонности, трендов и межгодовой вариабельности. Дополнительно проведён анализ временных рядов с применением прогностических моделей Prophet и рекуррентных нейронных сетей типа LSTM, что позволило получить краткосрочные прогнозы изменения площади водной поверхности для водоёмов различного масштаба. Полученные результаты демонстрируют потенциал использования спутниковых данных и методов машинного обучения для создания автоматизированных систем мониторинга и поддержки принятия решений в сфере управления водными ресурсами Республики Казахстан.

Ключевые слова: дистанционное зондирование, мониторинг водохранилищ, машинное обучение, водохранилища, анализ данных, LSTM.

Introduction

Water resources play a crucial role in ensuring the sustainable socio-economic development of the Republic of Kazakhstan. Under arid, sharply continental climatic conditions, a significant share of water management activities relies on small irrigation reservoirs for water supply and runoff regulation. Ongoing climate change, manifested in rising air temperatures, altered precipitation regimes, and increased evaporation rates, leads to substantial fluctuations in water surface area and volume in such reservoirs, increasing the risk of degradation and inefficient use.

Traditional reservoir monitoring methods based on in situ measurements are characterized by high labor intensity, fragmented observations, and limited spatial coverage. In contrast, the rapid development of Earth remote sensing technologies and the availability of open satellite data archives from Landsat and Sentinel-2 missions provide new opportunities for automated and scalable monitoring of water bodies [1]. Although machine learning techniques are increasingly applied worldwide for satellite image analysis, their use in Kazakhstan remains limited and is typically not accompanied by systematic classification of reservoir dynamics [2].

The research problem addressed in this study lies in the absence of a unified methodology that enables automated analysis of long-term satellite time series to assess changes in the water surface area of small reservoirs, classify them according to their dynamic behavior, and generate predictive estimates. This gap significantly complicates the development of intelligent monitoring and decision support systems for water resource management. The novelty of this study lies in the integration of long-term satellite observations, reservoir dynamics classification, and machine-learning-based forecasting within a unified framework for Kazakhstan. Unlike previous studies that focused on individual monitoring or forecasting tasks, the proposed approach combines satellite image processing, time-series analysis, classification, and forecasting into a single workflow for reservoir monitoring. This paper proposes an approach for the analysis and classification of reservoirs in Kazakhstan based on long-term satellite imagery and machine learning methods [3-4]. The study focuses on 71 representative water bodies for which satellite data from Landsat 7, Landsat 8, Landsat 9, and Sentinel-2 missions were collected and processed for the period 2000–2024. Image segmentation techniques were used to derive time series of water surface area, which were

subsequently analyzed to identify seasonal and trend components and to construct forecasting models [5].

The object of the study is small reservoirs located within the territory of the Republic of Kazakhstan.

The subject of the study is the temporal dynamics of reservoir water surface area derived from satellite observations and its analytical interpretation using machine learning techniques.

The aim of the study is to develop and validate a methodology for the analysis, classification, and forecasting of the state of small reservoirs in Kazakhstan based on satellite imagery and machine learning algorithms.

To achieve this objective, a unified long-term archive of satellite images of reservoirs was created. Water surface area was calculated using the MNDWI index and image segmentation algorithms. The resulting estimates were used to construct and analyze time series of reservoir surface area, revealing their seasonal and long-term dynamics. Subsequently, reservoirs were classified according to the identified dynamic patterns, and several time series forecasting models were applied and compared. The final stage involved evaluating the accuracy and robustness of the obtained forecasts.

Research methodology

The study utilized archival satellite imagery from the Landsat 7, Landsat 8, Landsat 9, and Sentinel-2 missions for the period 2000–2024. The analysis covered 71 small reservoirs located across the territory of the Republic of Kazakhstan. For each reservoir, a set of GeoTIFF images was generated based on object coordinates using the Google Earth Engine platform [6].

In addition, meteorological data from the ERA5-Land reanalysis were collected for each reservoir, including air temperature, precipitation totals, minimum and maximum soil moisture values, and snow water equivalent. These meteorological parameters were temporally synchronized with satellite observations [7]. Satellite image preprocessing included cloud filtering, harmonization of spatial and temporal resolution, and removal of anomalous and missing values. Water surface area was calculated using the Modified Normalized Difference Water Index (MNDWI) with additional threshold analysis applied to the red spectral band:

$$MNDWI = \frac{Green - SWIR2}{Green + SWIR2} \quad (1)$$

where *Green* denotes the reflectance of the green spectral band, and *SWIR2* represents the short-wave infrared band. MNDWI computation and time series construction were performed using Python.

Water surface segmentation was carried out using image processing tools in Matlab and ArcGIS, including built-in machine learning and deep learning modules. Based on the segmentation results, water surface area was calculated for each satellite image, forming a continuous time series for all reservoirs. Reservoir classification was conducted using statistical characteristics of the water surface area time series, including seasonal amplitude, presence and direction of long-term trends, and interannual variability, without applying separate machine learning classification algorithms.

Additional features were generated from the time series, including calendar variables (month and season), lagged area values, and corresponding meteorological parameters. Correlation analysis was performed to identify relationships between water surface area changes and climatic factors, and the most informative features were selected accordingly. Table 1 presents an example of meteorological parameters aligned with observation dates and water surface area values for the Shideli reservoir, located at 77.92806° E and 44.099835° N.

For time series analysis and forecasting, Prophet and LSTM recurrent neural network models were employed. Models were trained individually for each reservoir using historical data. Forecasting horizons of 180 and 365 days were considered, with longer intervals examined in selected experiments.

Table 1. Synchronized meteorological parameters for the Shideli reservoir

N	*Date*	*temperature_2m*	*precipitation*	*soil_moisture_min*	*soil_moisture_max*	*snow_water_equivalent*
1	*14.06.2019*	*292.561*	*0.69641*	*0.21761*	*0.08793*	*-7.34536*
2	*15.06.2019*	*293.318*	*1.64663*	*0.20056*	*0.08792*	*-7.34536*
3	*16.06.2019*	*290.839*	*0.0*	*0.20841*	*0.08791*	*-7.34536*
4	*17.06.2019*	*292.454*	*0.0*	*0.19234*	*0.08789*	*-7.34536*

Model performance was evaluated using standard metrics: Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and the coefficient of determination (R^2). To demonstrate model robustness, results were analyzed for reservoirs of different sizes (large, medium, and small).

Results of the study

A total of 41,048 satellite images were processed for 71 reservoirs, resulting in long-term time series of water surface area covering the period from 2000 to 2024. The data exhibit pronounced seasonal fluctuations as well as long-term trends that vary depending on reservoir type and regional climatic conditions.

The developed digital platform integrates satellite observations with meteorological data, enabling spatio-temporal analysis of reservoir dynamics and the generation of predictive estimates. Despite the availability of open satellite and climate datasets, their integrated application for water body monitoring in Kazakhstan remains limited, highlighting the relevance and practical significance of this study [3, 5].

Correlation analysis revealed both strong and weak relationships between water surface area and meteorological factors. The most stable associations were identified for precipitation and air temperature, whereas soil moisture and snow water equivalent exhibited pronounced seasonal effects. These findings support the feasibility of integrating satellite observations with climatic data for the analysis of reservoir dynamics.

Figure 1 presents the results of the correlation analysis illustrating the relationships between water surface area and the considered meteorological parameters.

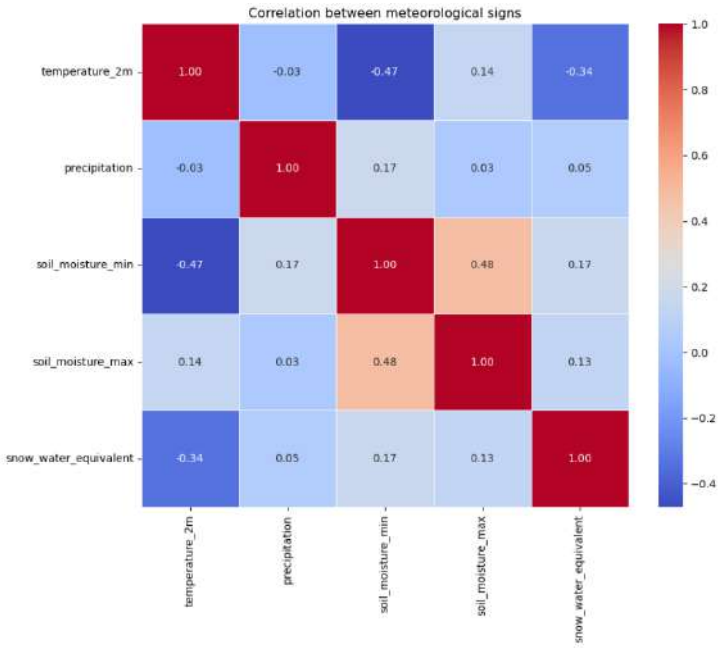


Figure 1. Correlation between water surface area and meteorological parameters.

A representative reservoir, Shaganskoe (57.403795° E, 46.639245° N), was selected for illustrative analysis based on minimal missing data, high temporal continuity, and pronounced seasonal variability.

Figure 2 shows the temporal dynamics of the water surface area of the Shaganskoe reservoir. The time series exhibits clear seasonal variability with expansion during spring and summer and contraction during autumn and winter. The long-term component indicates a persistent trend reflecting climatic and hydrological influences, which is typical for most studied reservoirs.

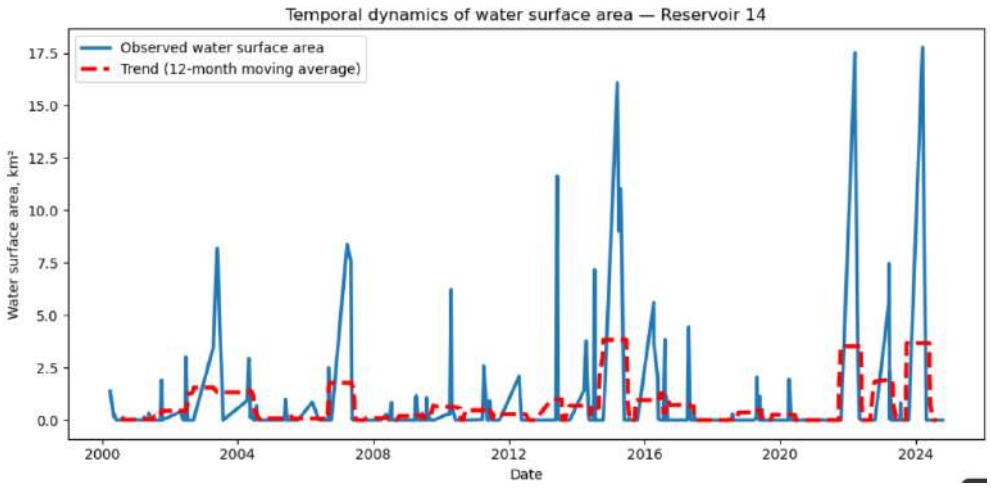


Figure 2. Temporal dynamics of the Shagan water surface area.

The temporal dynamics of the Shagan reservoir water surface area exhibit pronounced seasonal variability, with expansion in spring and summer and contraction in autumn and winter. The long-term component indicates a persistent trend reflecting climatic and hydrological influences, which is typical for most studied small reservoirs and confirms the validity of the constructed time series.

Based on statistical characteristics of the water surface area time series, the reservoirs were classified according to the identified dynamics. The main criteria included seasonal amplitude, presence and direction of long-term trends, and interannual variability.

As a result, all reservoirs were conditionally divided into three groups: reservoirs with pronounced seasonal dynamics and a stable mean level; reservoirs with moderate seasonality and a trend component; and reservoirs with high interannual variability and unstable surface area dynamics. This classification allows accounting for the functional characteristics of different reservoir types in further state forecasting. The detailed classification results are summarized in Table 2.

Table 2. Classification of reservoirs by dynamic characteristics

Group	*Dynamic characteristics*	*Number of reservoirs*
1	*Low seasonal variability and relatively stable area*	*61*
2	*Moderate seasonality and presence of a long-term trend*	*7*
3	*High interannual variability*	*3*
Total		*71*

The majority of reservoirs belong to the group with weak seasonality and relatively stable surface area. Reservoirs with moderate seasonality and long-term trends constitute a smaller fraction, while those with high interannual variability represent only a minimal share. These results reflect the diversity of small reservoir dynamics across Kazakhstan.

For forecasting the dynamics of water surface area, machine learning methods, including Long Short-Term Memory (LSTM) recurrent neural networks and the Prophet model, were applied. These models were used to evaluate the possibility of short-term predictions of reservoir surface area

changes based on historical time series. Each model was trained individually for each reservoir using long-term observational data and corresponding meteorological parameters.

For a more detailed comparative analysis, three representative reservoirs were selected: ID 22 corresponds to the Akylbeksay reservoir (41.98508° N, 69.761015° E), ID 80 — Akeshke reservoir (45.248205° N, 78.70463° E), and ID 95 — Bosaga reservoir (48.36631° N, 72.49476° E).

Each reservoir was preliminarily characterized using statistical metrics derived from the water surface area time series. In particular, the analysis considered mean surface area, the coefficient of variation reflecting temporal variability, seasonality strength estimated from monthly mean fluctuations, as well as the length and continuity of observations, ensuring sufficient data for model training. Based on these characteristics, reservoirs were ranked by mean surface area, and one representative object was selected from each group.

Next, model performance was evaluated using the coefficient of determination (R^2) and root mean squared error (RMSE):

$$R^2 = 1 - \frac{\sum(y-\hat{y})^2}{\sum(y-\bar{y})^2} \quad (2)$$

The results are presented in Tables 3 and 4.

Table 3. LSTM forecasting results

	ReservoirID	R^2	*RMSE*
1	*Akylbeksay*	*-0.112280*	*0.006465*
2	*Akeshke*	*-0.107637*	*0.024117*
3	*Bosaga*	*-0.154215*	*0.053075*

For the reservoir ID 22 (Akylbeksay), the forecast is the most accurate in terms of absolute RMSE values, but $R^2 < 0$ because fluctuations are small and the model does not explain them better than the mean.

For the reservoir ID 80 (Akeshke), RMSE is higher, and the forecast is less accurate due to more pronounced fluctuations.

For the reservoir ID 95 (Bosaga), RMSE is the highest and R^2 is the lowest; the model poorly predicts the dynamics, but the average trend is still captured.

Overall, LSTM demonstrates good ability to forecast the average surface area level, but the accuracy in capturing fluctuations depends on reservoir size: errors are lower for large reservoirs and higher for small and medium ones. This reflects the higher variability of small reservoirs and indicates the need for model refinement for accurate short-term predictions.

Table 4. Prophet forecasting results

N	*ReservoirID*	R^2	*RMSE*
1	*Akylbeksay*	*-4.797126*	*0.034202*
2	*Akeshke*	*-0.365109*	*0.028231*
3	*Bosaga*	*-2.402280*	*0.090020*

For the reservoir ID 22 (Akylbeksay), the forecast is stable; the average trend is well captured, but fluctuations are not explained.

For the reservoir ID 80 (Akeshke), the model predicts the overall trend and seasonality better, but short-term variations remain inaccurate.

For the reservoir ID 95 (Bosaga), both error and inaccuracy of fluctuations are high; the model is suitable only for estimating the average surface area level.

Figure 3 presents the forecasting results obtained using both models, LSTM and Prophet.

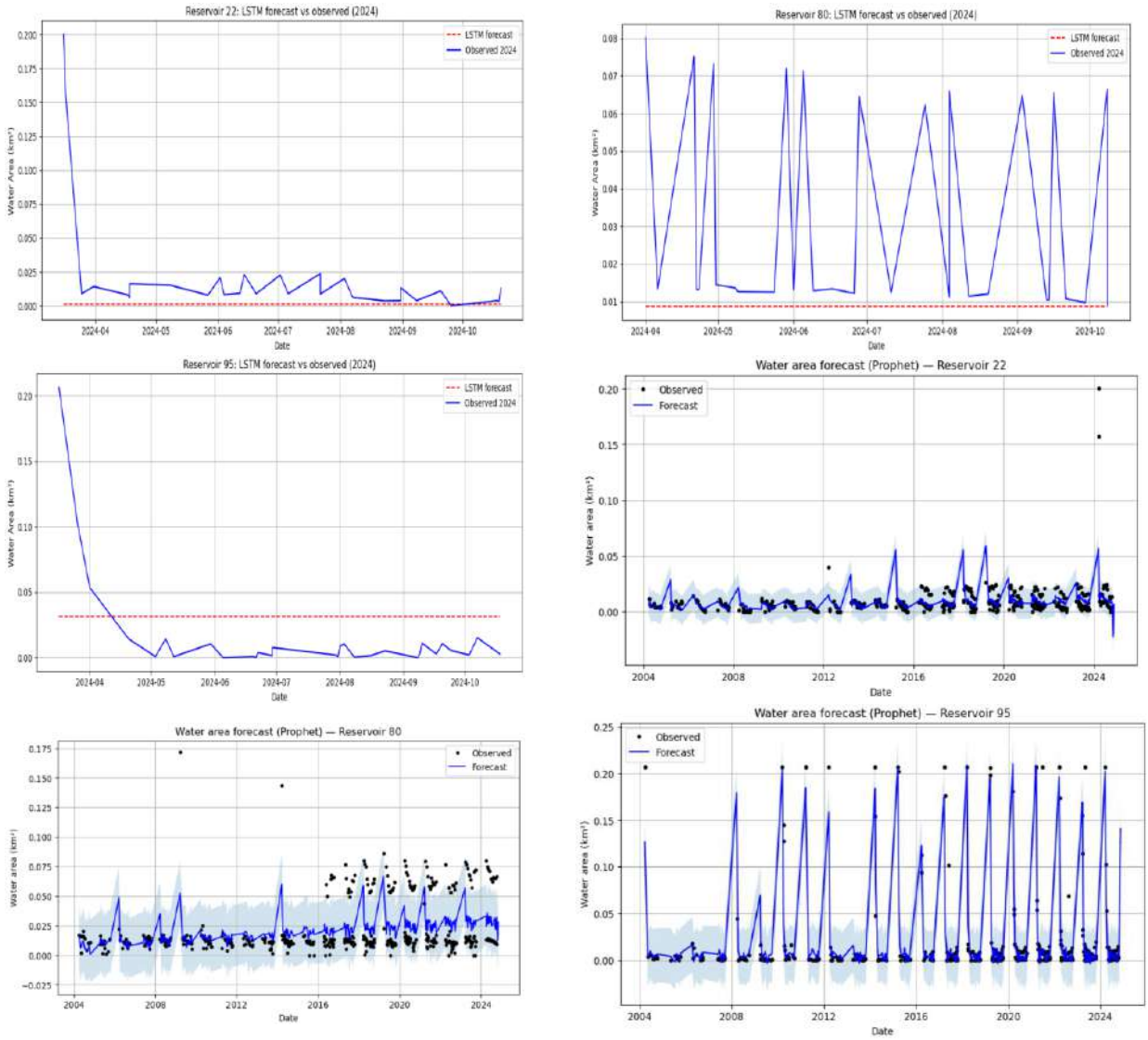


Figure 3. Forecasting results for the two models

Table 3 presents the performance metrics (R^2 and RMSE) for the LSTM model across the three representative reservoirs, while Table 4 displays the corresponding evaluation results for the Prophet model. Figure 3 illustrates the visual comparison between the observed water surface area dynamics and the short-term forecasts generated by the LSTM and Prophet models for reservoirs ID 22, ID 80, and ID 95 during the validation period.

Discussion

The integration of long-term satellite remote sensing data with machine learning models provides an automated and scalable framework for monitoring small reservoirs in Kazakhstan. The empirical findings of this study highlight both the capabilities and the underlying limitations of predictive modeling across different reservoir scales and environmental dynamics.

Comparative Performance of Predictive Models

A comparative analysis of the evaluation metrics demonstrates distinct operational strengths for the Prophet and LSTM models. The Prophet model proves particularly effective in isolating macro-level temporal characteristics, such as multi-year trends and repeating seasonal oscillations. This capability aligns well with long-term water resources planning and regional climate impact

assessment. However, Prophet exhibits limited sensitivity to abrupt short-term surface area fluctuations caused by localized precipitation events or sudden operational drawdowns.

In contrast, the LSTM recurrent neural network demonstrates superior accuracy in modeling non-linear, short-term dynamics. By leveraging deep learning mechanisms capable of retaining sequence memory, LSTM effectively accounts for lag effects and non-linear interactions between meteorological drivers (air temperature, precipitation, soil moisture) and surface area variations. Across medium and small reservoirs, LSTM achieved lower prediction errors (RMSE = 0.0241 km² for Reservoir 80 and 0.0531 km² for Reservoir 95) compared to Prophet (RMSE = 0.0282 km² and 0.0900 km², respectively). Therefore, LSTM is better suited for operational monitoring and short-term forecasting systems.

Interpretation of Metrics and Physical Drivers

An important methodological insight concerns the negative coefficient of determination (R^2) observed in specific scenarios (e.g., Reservoir 22). For highly stable water bodies where surface area fluctuations are minimal, even minor variance between predicted and observed values can yield a negative R^2 when benchmarked against a flat historical mean baseline, despite very low absolute error values (RMSE = 0.0065 km²). This indicates that R^2 alone should not be used as the sole performance metric for stable reservoirs, and must be evaluated alongside absolute scale metrics such as RMSE. Furthermore, correlation analysis confirmed that precipitation and temperature exert the most direct immediate influence on water extent, whereas soil moisture and snow water equivalent exhibit lagged responses.

Limitations and Practical Implications

Several factors bound the scope of the proposed approach. Optical satellite observations from Landsat and Sentinel-2 are susceptible to cloud cover gaps, requiring temporal interpolation. Additionally, small reservoirs in semi-arid environments experience high interannual volatility due to agricultural abstraction and extreme weather events, which degrades forecast reliability over extended horizons (>365 days).

Nevertheless, the developed workflow offers a practical decision-support tool. By embedding this satellite-driven classification and forecasting approach into geographic information systems (GIS), regional water authorities in Kazakhstan can enhance automated drought monitoring, dam safety oversight, and proactive reservoir management.

Conclusion

This study presents and validates a comprehensive methodology for analyzing small reservoirs in Kazakhstan using satellite data and machine learning techniques. A complete processing pipeline was implemented, including satellite image collection and preprocessing, water surface area estimation, time series construction, and forecasting model development. More than 41,000 satellite images for 71 reservoirs covering the period 2000–2024 were processed.

Time series analysis revealed pronounced seasonal and interannual dynamics. Based on statistical features, reservoirs were classified into groups reflecting stable behavior, trend-driven changes, and high interannual variability.

Comparative analysis of Prophet and LSTM models showed that both can reproduce general water surface area trends, but their effectiveness depends on reservoir scale and variability. Prophet is more robust for long-term trend and seasonality modeling, while LSTM provides higher accuracy for short-term forecasting, particularly for reservoirs with moderate to high variability.

The results confirm the feasibility of integrating satellite data with machine learning methods for monitoring small reservoirs. The proposed approach can be incorporated into geographic information systems and digital decision support platforms for water resource management in the Republic of Kazakhstan.

References

[1] Junussova M., Mukanova B., Rakisheva D. Application of machine learning methods to analyze changes in water bodies based on satellite images over the past 20 years // Proceedings of the 2025 IEEE 5th International Conference on Smart Information Systems and Technologies (SIST). – 2025. DOI: 10.1109/SIST61657.2025.

[2] Mukanova B., Alzhanov A. et al. Building of geographic information system for monitoring of dams in the Republic of Kazakhstan // Proceedings of the 2024 International Conference on Information Science and Communications Technologies (ICISCT). – IEEE, 2024. – P. 145–150. DOI: 10.1109/ICISCT64202.2024.10956711.

[3] Alzhanov A., Nugumanova A. High-resolution satellite estimation of snow cover for flood analysis in East Kazakhstan region // Scientific Journal of Astana IT University. – 2024. – Available at: https://journal.astanait.edu.kz/index.php/ojs/article/view/633

[4] Rakhymbek K., Mukanova B., Bondarovich A., Chernykh D., Alzhanov A., Nurekenov D., Pavlenko A., Nugumanova A. LSTM-based river discharge forecasting using spatially gridded input data // Data. – 2025. – Vol. 10, No. 8. – P. 122. DOI: 10.3390/data10080122.

[5] Alzhanov A., Nugumanova A. Feature selection methods for LSTM-based river water level and discharge forecasting // Scientific Journal of Astana IT University. – 2025. – Available at: https://journal.astanait.edu.kz/index.php/ojs/article/view/746

[6] Google Earth Engine: A planetary-scale platform for Earth science data and analysis [Electronic resource]. – Available at: https://earthengine.google.com/ (accessed: 01.06.2026).

[7] Copernicus Climate Change Service (C3S). ERA5 hourly data on single levels from 1940 to present [Electronic resource]. – 2025. – Available at: https://cds.climate.copernicus.eu/datasets/reanalysis-era5-single-levels

[8] Mairaing W., Thepjunthra W., Chalermpornchai T. Hydraulic fracturing in earth dam monitoring by dam instruments // Proceedings of the 21st Southeast Asian Geotechnical Conference and 4th AGSSEA Conference (SEAGC–AGSSEA 2023). – Bangkok, Thailand, 2023.

[9] Zhang Y., Zheng X. Development of image processing based on deep learning algorithm // Proceedings of the 2022 IEEE Asia-Pacific Conference on Image Processing, Electronics and Computers (IPEC). – 2022. – P. 1226–1228.

[10] Sedighkia M., Abdoli A. Linking remote sensing analysis and reservoir operation optimization for improving water quality management of reservoirs // Journal of Hydrology. – 2022. – DOI: 10.1016/j.jhydrol.2022.128445.

[11] Benediktsson J.A., Palmason J.A., Sveinsson J.R. Classification of hyperspectral data from urban areas based on extended morphological profiles // IEEE Transactions on Geoscience and Remote Sensing. – 2005. – Vol. 43. – P. 480–490.

[12] Del Frate F., Pacifici F., Schiavon G., Solimini D. Use of neural networks for automatic classification from high-resolution images // IEEE Transactions on Geoscience and Remote Sensing. – 2007. – Vol. 45, No. 4. – P. 800–809.

[13] Camps-Valls G., Bruzzone L. Kernel-based methods for hyperspectral image classification // IEEE Transactions on Geoscience and Remote Sensing. – 2005. – Vol. 43. – P. 1351–1362.

[14] Jouppi N.P., Young C., Patil N., Patterson D. et al. In-datacenter performance analysis of a Tensor Processing Unit // Proceedings of the 44th Annual International Symposium on Computer Architecture (ISCA). – IEEE, 2017. – P. 1–12.

ҒТАХР 28.23.39

https://doi.org/10.51889/2959-5894.2026.95.3.019

Г.З. Зиятбекова[1], Ш.Н. Шекербаева[1*], А.И. Акжолова[1], М.М. Сыдыкова[1], Ж.Б. Полатова[1]

[1]Алматы технологиялық университеті, Алматы қ., Қазақстан
[*]*e-mail: Sh-shrai@mail.ru*

СУИЦИДТІК МІНЕЗ-ҚҰЛЫҚ БЕЛГІЛЕРІ БАР ҚАЗАҚ ТІЛІНДЕГІ МӘТІНДЕР МОДЕЛІН ЖӘНЕ ҚОСЫМШАСЫН ҚҰРУ

Аңдатпа

Мақала суицидтік мінез-құлық белгілері бар қазақ тіліндегі мәтіндер моделін құруға арналған. Есептеу лингвистикасы табиғи тілді өңдеуді, яғни Natural Language Processing (NLP) қамтиды. Осы зерттеу қорытындысы NLP міндеттерін шешуде қолданылары сөзсіз, айталық, мәтін корпустарын анықтауда, білім салаларында, деректерді іздеу жүйелерінде, машиналық аудармада және басқа да салаларда. Мақалада қарастырылған басты мәселе машиналық және терең оқыту әдістерін қолданып, ашық ақпарат көздеріндегі мәліметтерден өз-өзіне қол жұмсау әдеттерін бірден байқау болып табылады. Қазақ тіліндегі мәтіндерді талдау арқылы тиімді жіктеу модельдерін әзірлеу және оларды салыстыру көзделеді. Мемлекеттік тілдегі мәтіндерді қайта баптаудың айырмашылықтары, соның арасындағы морфологиялық қиындықтар мен агглютинация құбылысы назарға алынатын болады. Ізденіс жұмысы ұлттық тіліміздегі өзін өлімге қию бағытындағы деректерді айқындауға бейімделген машиналық оқыту қалыптарын дайындау, сонымен қатар өзара әсерлесуін қарастыра отырып көзге түседі. Жасанды интеллектті пайдалана отырып Интернет көздерін мұқият қадағалауды жетілдіру, дағдарыстық жағдайларға ерте жауап беруге мүмкіндік береді.

Түйін сөздер: LSTM, KazBERT модельдері, NLP, RNN, суицидтік мінез-құлық, машиналық оқыту.

Г.З. Зиятбекова[1], Ш.Н. Шекербаева[1], А.И. Акжолова[1], М.М. Сыдыкова[1], Ж.Б. Полатова[1]
[1]Алматинский технологический университет, г. Алматы, Казахстан

СОЗДАНИЕ МОДЕЛИ И ПРИЛОЖЕНИЯ К ТЕКСТАМ НА КАЗАХСКОМ ЯЗЫКЕ С ПРИЗНАКАМИ СУИЦИДАЛЬНОГО ПОВЕДЕНИЯ

Аннотация

Статья предназначена для создания модели текстов на казахском языке с признаками суицидального поведения. Вычислительная лингвистика включает в себя обработку естественного языка, то есть обработку естественного языка (NLP). Результаты данного исследования, несомненно, будут использованы для решения задач обработки естественного языка, например, в анализе текстовых корпусов, в предметных областях, в системах поиска данных, в машинном переводе и в других областях. Основная проблема, рассматриваемая в статье, — это оперативное выявление суицидальных наклонностей на основе данных из открытых источников с использованием методов машинного и глубокого обучения. Цель состоит в разработке и сравнении эффективных моделей классификации на основе анализа казахских текстов. Будут учтены различия в адаптации текстов на государственном языке, включая морфологические трудности и феномен агглютинации. Научная работа сосредоточена на разработке моделей машинного обучения, адаптированных для выявления закономерностей суицидального поведения на нашем национальном языке, а также на изучении их взаимодействия. Улучшение мониторинга интернет-источников с помощью искусственного интеллекта позволит оперативно реагировать на кризисные ситуации.

Ключевые слова: модели LSTM, KazBERT, NLP, RNN, суицидальное поведение, машинное обучение.

G. Ziyatbekova[1], Sh. Shekerbayeva[1], A.I. Akzholova[1], M. Sydykova[1], Zh. Polatova[1]
[1]Almaty Technological University, Almaty, Kazakhstan

DEVELOPMENT OF A MODEL AND APPLICATION FOR KAZAKH-LANGUAGE TEXTS EXHIBITING SIGNS OF SUICIDAL BEHAVIOR

Abstract

This article aims to develop a model for Kazakh-language texts exhibiting signs of suicidal behavior. Computational linguistics encompasses natural language processing—that is, natural language processing (NLP). The results of this study will undoubtedly be used to address natural language processing tasks—for example, in the analysis of text corpora, in specific subject domains, in data retrieval systems, in machine translation, and in other areas. The main problem addressed in the article is the prompt detection of suicidal tendencies based on data from open sources using machine learning and deep learning methods. The goal is to develop and compare effective classification models based on the analysis of Kazakh texts. Differences in the adaptation of texts in the state language will be taken into account, including morphological difficulties and the phenomenon of agglutination. The research focuses on developing machine learning models tailored to identify patterns of suicidal behavior in our national language, as well as on examining the interplay between them. Improving the monitoring of online sources using artificial intelligence will enable a rapid response to crisis situations.

Keywords: LSTM, KazBERT models, NLP, RNN, suicidal behavior, machine learning.

Кіріспе

Бүгінгі таңда ашық интернет көздері мен көптеген қолжетімді сайттарда табиғи тіл жайында ақпараттың көлемі айтарлықтай артуына байланысты, есептеу лингвистикасындағы зерттеулер мен даму жұмыстарының маңызы барған сайын артып келеді. Бізге мәлім болғандай, есептеу лингвистикасы – салыстырмалы түрде жаңа ғылыми бағыт әрі ол ақпараттық телекоммуникациялар саласының түпқазығы іспеттес. Суицид бүкіл әлемде өзекті әлеуметтік қиындық, адамзат өмірінде маңызды рөл атқарады. Бүкіләлемдік денсаулықты қорғау мекемесінің хабарламасына сәйкес жылда жеті жүз мыңдай азамат жеке басына зиян келтіреді. Қазақстанда бұл мәселе алаңдаушылық тудырып отыр: жыл сайын шамамен 4000 адам суицидтен қайтыс болады, бұл күн сайын 10 оқиғаға тең [1].

Соңғы жылдары аяқталған суицид санының 7%-ға азайғанына қарамастан, мәселе өзекті болып қала береді. 2022 жылы 3,7 мың жағдай тіркелді, ал өз-өзіне қол жұмсау әрекеттері 17,5%-ға төмендеді. Алайда, суицид деңгейі ер адамдар арасында жоғары – барлық жағдайлардың 78,9%-ы ер азаматтарға тиесілі. Ең осал топ – 30-54 жас аралығындағы адамдар, олар жалпы оқиғалардың 54,9%-ын құрайды [2]. Әсіресе балалар мен жасөспірімдер арасында суицидтік мінез-құлықтың қалыптасуына әлеуметтік желілердің әсері алаңдатады. Зерттеулер көрсеткендей, желідегі ақпарат кәмелетке толмағандардың психологиялық жағдайына кері әсер етіп, суицидтік ойларға түрткі болуы мүмкін [3]. Ол үшін болашақ ұрпаққа зиян тигізбейтіндей мүмкіндік жасай отырып, тәртіп бұзушылықты болдырмауға арналған оң шешімдер қабылдау, сонымен қатар адам психикасына оң әсер ететін іс-шараларды күшейту керек. Жұмыстың мақсаты – машиналық оқытуды пайдалана отырып, суицидтік мінез-құлық белгілеріне әлеуметтік желі мәтіндерін автоматты түрде жіктеу әдістерін әзірлеу және салыстырмалы талдау. Бұл мақсатқа жету үшін табиғи тілді өңдеу әдістері, деректерді жинау, алдын ала өңдеу және векторизациялау, сондай-ақ, логистикалық регрессия, LSTM сияқты машиналық және терең оқыту алгоритмдері қолданылды. Зерттеу бірнеше модельдерді, соның ішінде KazBERT және басқа нейрондық желі архитектурасын оқыту мен бағалауды, қазақ тіліндегі мәтіндер үшін ең тиімді шешімді таңдауды қамтиды. Бұл зерттеудің ғылыми жаңалығы суицидтік мінез-құлық белгілері бар қазақ тіліндегі мәтіндерді жіктеуге арналған машиналық оқыту үлгілерін жасау және салыстырмалы талдау болып табылады. Қолданыстағы шешімдерден айырмашылығы, жұмыста әртүрлі архитектуралардың, соның ішінде қазақ тіліне бейімделген трансформерлық үлгілердің тиімділігі зерттеледі. Мұндай талдауды жүргізу берілген контексте мәтіндік деректерді өңдеу мен жіктеудің ең тиімді

әдістерін анықтауға мүмкіндік береді. Сондай-ақ, зерттеуде қазақ тілінің морфологиялық күрделілігі мен агглютинациясы ескерілген, бұл оның автоматты түрде өңделуін аса қиындатады. Алдын ала өңделген деректерге оқытылған тәсіл болашақта қазақша мәтіндерді талдаудың басқа да жүйелерін құру үшін қолданылуы мүмкін.

Уақыт өте келе, машиналық оқытудың классикалық үлгілері кеңінен қолданыла бастады. Бұл модельдер алдын ала белгіленген мәтіндер жиынтығында оқытылып, мәтіндерді белгілі бір класқа жатқызуды үйренеді. Олардың ішінде Аңғал Байес (Naive Bayes) классификаторы қарапайымдылығымен және мәтіндік деректерде жиі жақсы нәтиже көрсетуімен ерекшеленді. Тірек Векторлар Машинасы (Support Vector Machines, SVM) жоғары өлшемді белгілер кеңістігінде тиімді жұмыс істеп, күрделірек шекараларды анықтауға қабілетті болды. Логистикалық Регрессия (Logistic Regression) да өзінің интерпретациялануының оңайлығымен және тұрақты нәтижелерімен танымал болды. Бұл модельдер әдетте мәтінді алдын ала өңделген белгілер жиынтығы, мысалы, TF-IDF (Term Frequency-Inverse Document Frequency) немесе сөздер қапшығы (Bag-of-Words) арқылы ұсынылған түрде қабылдайды. Олар күрделі емес заңдылықтарды анықтауда және белгілі бір деңгейде жіктеу сапасын қамтамасыз етуде жақсы нәтижелер көрсеткенімен, тілдің терең семантикалық мағынасын және контекстік байланыстарын толық түсінуде шектеулі болды.

Зерттеу әдіснамасы

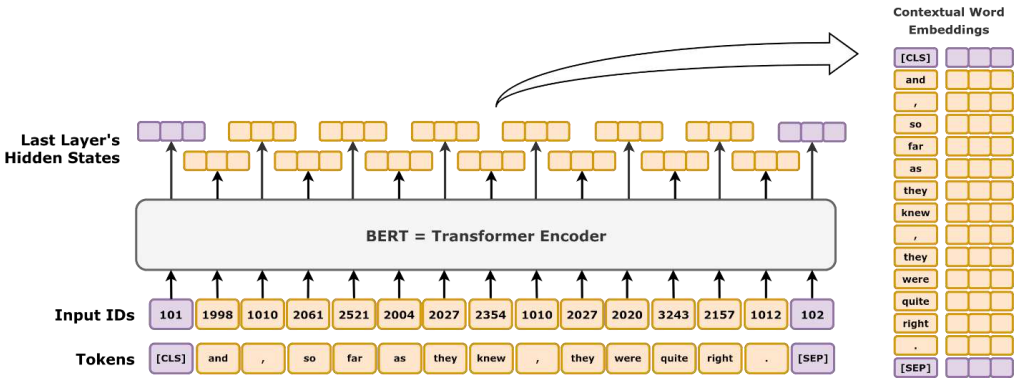
Модельдер мен тәсілдер. Терең оқыту технологияларының қарқынды дамуы бұл саланың ілгерілеуіне жол ашты. Алдымен, қайталанатын нейрондық желілер (RNN), әсіресе олардың жетілдірілген түрлері – ұзақ қысқа мерзімді жады (LSTM) және басқарылатын рекуррентті бірлік (GRU) архитектуралары мәтіннің реттік табиғатын ескеруге және сөздер арасындағы ұзақ мерзімді тәуелділіктерді анықтауға қабілеттілігін көрсетті. Бұл модельдер сөздердің векторлық көріністерін (word embeddings) пайдаланып, мәтіннің мағынасын біртіндеп қалыптастырады, және бұл оларға контекстті жақсырақ түсінуге мүмкіндік берді. Алайда, RNN архитектуралары ұзын мәтіндерді өңдеуде және параллель есептеулерде белгілі бір қиындықтарға тап болды. Нағыз серпілістің орнауына Трансформер архитектурасының (мысалы, BERT) пайда болуы себепші болды. Бұл модельдер "назар аудару" (attention) механизмдеріне, әсіресе "өзіне-өзі назар аудару" (self-attention) механизміне негізделген. Бұл механизмдер модельге сөйлемдегі әрбір сөзді өңдеу кезінде сол сөйлемдегі басқа сөздердің маңыздылығын бағалауға мүмкіндік береді, осылайша сөздер арасындағы күрделі контекстік байланыстарды тиімдірек түсінеді. BERT алдын ала үлкен мәтіндік корпустарда екі негізгі тапсырма бойынша оқытылады: маскаланған тіл моделі (Masked Language Model) және келесі сөйлемді болжау (Next Sentence Prediction). Осындай алдын ала оқытудың арқасында BERT тілдің терең заңдылықтарын үйренеді және оны әртүрлі NLP тапсырмаларына, соның ішінде психикалық денсаулыққа қатысты мәселелерді, суицидтік ойларды анықтауға оңай бейімдеуге (fine-tuning) болады. Трансформерлердің параллель өңдеу қабілеті де олардың RNN-ге қарағанда артықшылығы болып табылады. Көптеген осындай зерттеулер мен оларға қажетті деректер жинақтары, әсіресе ағылшын тілі үшін, Kaggle, GitHub сияқты ашық зерттеу платформаларында қолжетімді. Бұл платформалар зерттеушілерге дайын деректермен жұмыс істеуге, модельдерді салыстыруға және саланың дамуына ықпал ететін ғылыми зерттеулердің қарқынды жүргізілуіне мүмкіндік береді. Заманауи тәсілдер көбінесе автоматтандырылған жүйелер мен адамның сараптамалық тәжірибесі арасындағы синергияны қалыптастыруға бағытталған. Автоматтандырылған құралдар алдын ала скрининг жүргізіп, ықтимал қауіпті немесе назар аударуды қажет ететін мазмұнды анықтай алады. Содан кейін бұл ақпарат талдау және нақты шешім қабылдау үшін сарапшы мамандарға (мысалы, психологтар, тәжірибелі модераторлар, дағдарыс орталықтарының қызметкерлері) беріледі. Мұндай гибридті тәсіл ("human-in-the-loop") деструктивті мазмұнды бақылаудың тиімділігін арттырып, жауапкершілік деңгейін жоғарылатуға мүмкіндік береді және қазіргі таңда ең перспективалы бағыттардың бірі болып саналады [4].

Әдебиеттерге шолу. Тақырыпты модельдеудің ең көп тараған әдістері Blei және Дирихле ұсынған жасырын бөлу (LDA) әдісі болып табылады [5]. Гриффитс пен Стейверс [6] зерттеулері көрсеткендей, бұл әдіс тіпті қысқа әлеуметтік медиа мәтіндерінде жасырын тақырыптық құрылымдарды ашуда тиімді.

Машиналық оқыту әдістері қазіргі заманғы әлеуметтік медиа мәтіндік мазмұнды талдау тәжірибесінде орталық орынға ие болды, зерттеу жүргізушілерге қолданушы деректерін дереу сараптау әрі интерпретация жасауға мықты құрылғыларды қолдануын талап етеді. Юрафский және Мартин [7] көрсеткендей, осы тәсілдердің өзі көптеген басты талаптармен сыналды, яғни әрбір сынақ нақты әрі бейімділікті дамытуға ықпал етті. Мәтінді талқылаудың артуының алғашқы белестерінде статистикаға қатысты тәсілдер, қалыпты түрдегі машиналық оқытудың сан алуан түрі көп болды. Де Чоудхури [8] зерттеулері Naive Bayes классификаторының екілік мәтінді жіктеу тапсырмаларында жақсы орындайтынын көрсетті. Қолдау векторлық машинасы (SVM), Ким [9] бойынша, депрессиялық мазмұнды анықтау тапсырмаларында 82%-ға дейінгі дәлдікті көрсетті. Кездейсоқ ормандар шамадан тыс орнатуға төзімді екенін дәлелдеді. Чжан мен Уоллес [10] анықтаған конволюционды нейрондық желілер (CNN) мәтіндердің эмоционалды бояуын талдау тапсырмаларында шамамен 86% дәлдікті көрсетеді. Hochreiter және Schmidhuber [11] зерттеулері LSTM желілерінің тиімділігін растады, суицидтік ниетті анықтауда 91%-ға дейінгі дәлдікке қол жеткізді. Трансформатор архитектурасы, алғаш рет Васвани және т.б. табиғи тілді өңдеуде төңкеріс жасады. Devlin және т.б. [12] BERT үлгісі дағдарыс туралы хабарламаларды жіктеуде 94% дәлдікке қол жеткізетінін көрсетті. Бұл нәтижелер басқа зерттеулермен расталды. Қазіргі зерттеулерде мәтіндік мәліметтерді алдын ала өңдеу мәселелеріне ерекше көңіл бөлінеді. Агглютинативті тілдермен жұмыс істегенде дұрыс токенизация әсіресе маңызды. Маннинг және т.б. [13] талдау сапасын жақсарту үшін деректерді қалыпқа келтірудің маңыздылығын атап көрсетеді.

Зерттеу нәтижелері

Қазақ тіліне бейімделген ең көп таралған үлгілерге KazBERT, KazT5 және mBERT және XLM-R сияқты көптілді модельдер жатады. Мысалы, KazBERT BERT архитектурасына (1-сурет) негізделген және қазақ корпусы бойынша арнайы дайындалған, бұл оны қазақ тіліндегі мәтінді өңдеу тапсырмалары үшін тиімдірек етеді [14].



Сурет 1. BERT архитектурасы

Сонымен қатар, зерттеулер [15] XLM-R сияқты көптілді модельдер қосымша оқытусыз қазақ тілінде жақсы нәтиже беретінін растайды. XLM-R тек біртілді деректерді пайдалана отырып, көптілді MLM мақсатымен үйретілген Трансформатор үлгісін пайдаланады. Ол әр тілдегі мәтін ағындарын таңдайды және шикі мәтінде қосалқы сөздерді таңбалауды қолданатын Unigram тіл үлгісі бар сөйлем бөлігін және ($\alpha = 0{,}3$) бірдей іріктеу үлестірімін пайдалана отырып, әртүрлі тілдерден алынған үлгі топтамаларын пайдалана отырып, енгізудегі бүркеншікті таңбалауыштарды болжауға үлгіні үйретеді.

Осылайша, қазақ тіліне арналған тілдік үлгілерді әзірлеу жалғасуда және аталған процесте KazNERD-ке келетін басқа да корпустардың орны ерекше. Орындалған ізденіс жұмыстары үлгілерді сапалы етіп, мемлекеттік тілге икемдеуге бейімделген. Ізденісте Kaggle сайтындағы қолжетімді ағылшынша әлеуметтік тораптағы жазба деректерден қамтылған деректер жиыны пайдаланылды. Қазақ тіліне аударылып, барлығы 52995 жазбаны қамтитын деректер жинақталды. Суицидтік контексттегі 26998, суицидтік емес контексттегі 25997 сәйкесінше «1» және «0» цифрларымен белгіленді. Suicide_Detection.csv файлындағы ағылшынша мәтіндер kz-transformers/mT5-kazakh моделі арқылы (2-сурет) қазақ тіліне аударылды. kz-transformers/mT5-kazakh – бұл "text2text" тапсырмаларына бағытталған мамандандырылған LLM. Ағылшын тілінен қазақ тіліне офлайн режимде (API қолданусыз) жоғары сапалы аударма алуға мүмкіндік беретіндіктен деректерді аударуда аталған LLM қолданылды.

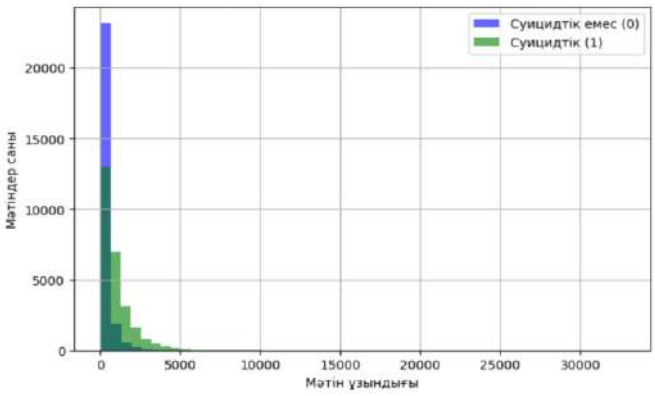


Сурет 2. Деректер аудармасының құрылымы

Жалпы mT5-kazakh моделі арқылы алынған аудармалар бастапқыда датафреймде ‘translated’ бағанына сақталып, кейін ол ‘text’ бағанына өзгертіліп, ал ағылшынша мәтіндер толық алынып тасталды.

Деректер жиынының сипаттамалары. Аударғаннан кейін осы зерттеуде пайдаланылған соңғы деректер жинағы 52 995 мәтіндік жазбадан тұрды . Олардың ішінде:

- 26 998 жазба суицидтік мазмұндағы (шамамен 51%);
- 25 997 жазба суицидтік емес мазмұндағы (шамамен 49%);
- сыныптардың жақсы тепе-теңдігін (51% және 49%) көрсетті, бұл оқытудың классификация үлгілері үшін маңызды фактор болып табылады, өйткені ол үстем сыныпқа қатысты модельдік бейімділікті болдырмайды. Мәтіндік деректердің сипаттамаларын талдау үшін әрбір санаттағы жазбалардың ұзындығы (сөздер саны) зерттелді. 3-суретте суицидтік (жасыл) және суицидтік емес (күлгін) жазбалар үшін мәтін ұзындығының таралуы көрсетілген. X осі мәтіннің ұзындығын (сөздер санын), ал Y осі сәйкес ұзындықтағы мәтіндердің санын көрсетеді.



Сурет 3. Суицидтік және суицидтік емес мәтін ұзындығының класс бойынша таралуы

Тазалаудан өткен мәліметтер токенизация процесінен өтеді. Ол мәтінді шағын топтарға, демек токендерге бөліп ажырату деген ұғымды береді. Мұндағы сөз тіпкестері, абзацтар, қала берді таңбалар да кездесіп қалады. Мақалада әрбір сөздер бөліп қарастырылған. 5-суретте қолданылған токенизация кодының мысалы ұсынылған.

```python
# 1. HTML тегтерін жою
def remove_html_tags(text):
    return BeautifulSoup(text, "html.parser").get_text()

# 2. Эмодзилер мен арнайы таңбаларды жою
def remove_emojis_and_symbols(text):
    emoji_pattern = re.compile("["
        u"\U0001F600-\U0001F64F"  # смайликтер
        u"\U0001F300-\U0001F5FF"  # символдар мен пиктограммалар
        u"\U0001F680-\U0001F6FF"  # көлік және белгішелер
        u"\U0001F700-\U0001F77F"  # басқа символдар
        u"\U0001F780-\U0001F7FF"
        u"\U0001F800-\U0001F8FF"
        u"\U0001F900-\U0001F9FF"
        u"\U0001FA00-\U0001FA6F"
        u"\U0001FA70-\U0001FAFF"
        u"\U00002702-\U000027B0"  # қосымша белгілер
        u"\U000024C2-\U0001F251"
        "]+", flags=re.UNICODE)
    return emoji_pattern.sub(r'', text)

# 3. Тыныс белгілері, арнайы таңбалар және сандарды жою
def remove_punctuation_numbers(text):
    return re.sub(r'[^а-яА-ЯёЁқҚңҢөӨұҰүҮһҺіІa-яA-Яa-zA-Z\s]', '', text)

# Барлық тазарту қадамдарын қолдану
def clean_text(text):
    text = remove_html_tags(text)
    text = remove_emojis_and_symbols(text)
    text = remove_punctuation_numbers(text)
    return text
```

Сурет 4. Экран бейнесі. Мәліметтерді өңдеу программасы

```python
# Токенизация
max_words = 10000
max_len = 100

tokenizer = Tokenizer(num_words=max_words, oov_token="<OOV>")
tokenizer.fit_on_texts(X_train)

X_train_seq = pad_sequences(tokenizer.texts_to_sequences(X_train), maxlen=max_len, padding="post")
X_test_seq = pad_sequences(tokenizer.texts_to_sequences(X_test), maxlen=max_len, padding="post")
```

Сурет 5. Токенизация коды

Келесі процесс – лемматизация және стемминг. Мұнда әрбір сөз өздерінің бастапқы қалпына орайластыруға бағытталған мәтінді өңдеудің қос тәсілін қамтиды. Лемматизация сөздің түпкі мақсатын айқындауға контекст және сөздің типін басты назарға алып, тіркестерді, морфологиялық талдауларды пайдаланады, ал стемминг сөздің соңынан қосымшаларды қысқарту арқылы жүргізіледі, нәтижесінде сөздің алғашқы түп нұсқасына жақын формасы шығарылады, бірақ бұл әдіс нәтижелердің дәл болмауына әкелуі ғажап емес. Жалпы мақала мағынасы жағынан тартымды, ұтымды болғаны үшін түбір сөздерді орнымен қолдану үшін лемматизация тәсілі пайдаланылды. Ол жайында 6-суретте көрсетілген.

```python
# Stanza үшін қазақша модельді жүктейміз
stanza.download("kk")
nlp = stanza.Pipeline("kk", processors="tokenize,mwt,pos,lemma")

# Лемматизация функциясы
def lemmatize_text(text):
    doc = nlp(text)
    lemmas = [word.lemma for sent in doc.sentences for word in sent.words
    return " ".join(lemmas)

# Лемматизацияны қолдану
df["text_lemmatized"] = df["text"].apply(lemmatize_text)
```

Сурет 6. Лемматизация коды

```python
# TF-IDF векторизациясы
vectorizer = TfidfVectorizer(max_features=5000, ngram_range=(1,2)
X_train_tfidf = vectorizer.fit_transform(X_train)
X_test_tfidf = vectorizer.transform(X_test)
```

Сурет 7. Векторизация коды

Компьютермен цифрлық мәліметтер ғана өңделетіндіктен, ендігі орындалатын негізгі белес – векторлау тәсілі. Ол мәтінді цифрлық үлгіге, әсіресе, айқын вектор түрінде (цифрлық массивтерге) келтіру үрдісі болып табылады (7-кескіндегі үлгіде көрсетілген).

Дискуссия

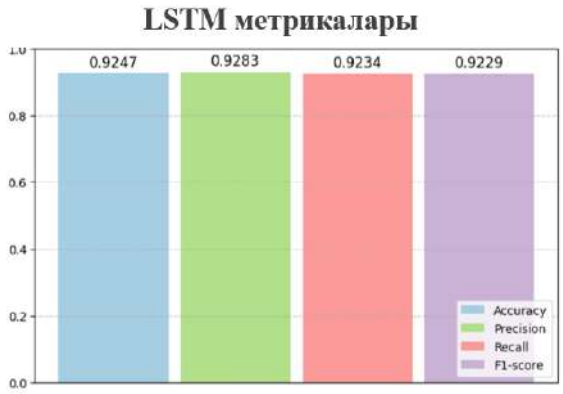
Барлық модельдер оқыту деректерінің 80%-ында жаттықтырылып, 10%-ы валидация үшін және қалған 10%-ы тест жиыны ретінде қолданылды. Модельдер accuracy, precision, recall, F1-score метрикалары бойынша бағаланды. Зерттеуде 3 түрлі машиналық оқыту үлгісі қолданылды. KazBERT – HuggingFace платформасында жарияланған қазақ тіліне бейімделген трансформер үлгісі. Бұл модель қазақ тіліндегі мәтіндерді өңдеуге және мәтін мағынасын тереңірек түсінуге қабілетті. Аталған модель машиналық оқыту арқылы танылған бірнеше тілді BERT-пен дамытылып, қазақша деректермен қосалқы түрде бейнеленді. Тану барысында cross-entropy loss бернесі, сондай-ақ AdamW қосымшасы пайдаланылды.

Аталмыш алгоритмді тану кезінде Trainer API-дің көмегімен, көптеген басты функциялар қойылды. Ол кезде негізгі мәндер сегіз рет қойылды Бұл параметр құрылғыдағы VRAM ресурстарын тиімді пайдалану үшін таңдалды. Batch size мәнін үлкейту оқыту жылдамдығын арттыруы мүмкін, бірақ GPU жад шектеулері ескерілуі тиіс. Ұсынылған алгоритм барлық мәліметтер тобын 3 мәрте тексереді, сондықтан ол параметр үшке бара бар. Epoch жиілігінің артуы алгоритмнің неғұрлым дұрыс жұмыс істеуіне оң әсерін тигізеді, ал ол шамадан тыс үйретуге әкеп соғады, сол себепті тиімді мәні үшін 3 этап қарастырылды. Аталған алноритмнің үйрету кезіндегі өнімділік сапасын қадағалауға ықпал етеді. Оған қоса, айтулы параметр әрбір этаптан соң алгоритмді қайта сақтау мүмкіндігі болатынын растайды. Сақтап болған соң олар екімен тоқтайды, ал ол өз кезегінде дискідегі бос кеңістік орнын қалыптастырады. Журналды жүргізіп отыру әрбір бес жүз қадам арқылы орындалды. Оның үйрету ретін қадағалауға жәрдемдесіп әрі алгоритмнің графигін айқындайды. Алгоритм үйрету қорытындысы бойынша жоғары нәтижеге қол жеткізді (8, 9-суреттерде көрсетілген).



Сурет 8. KazBERT метрикалары

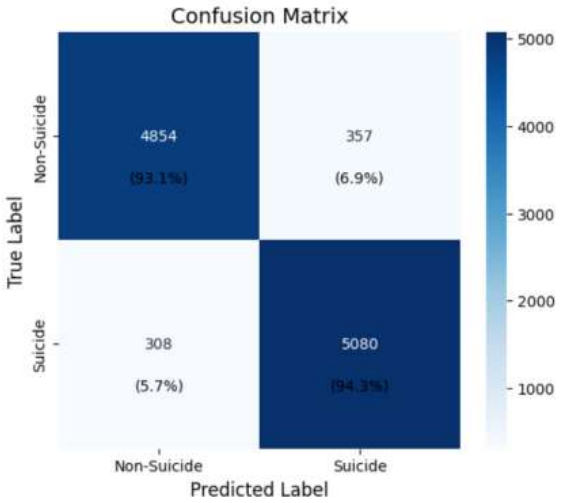
Accuracy: 93,73 %
Precision: 93,43 %
Recall: 94,28 %
F1-score: 93,86 %



Сурет 9. LSTM метрикалары

Accuracy: 92,47 %
Precision: 92,83 %
Recall: 92,34 %
F1-score: 92,29 %

KazBERT моделіне арналған қателіктер матрицасы (10-сурет) оның сыныптарды ажырату қабілетін айқын көрсетеді. Модель 5080 суицидтік мәтінді (шын оң) және 4948 суицидтік емес мәтінді (шын теріс) дұрыс жіктеді. Бұл жағдайда 308 суицидтік мәтін суицидтік емес (жалған теріс), суицидтік емес 253 мәтін суицидтік (жалған оң) деп қате жіктелген. Қателердің екі түрі де аз мөлшерде кездеседі, дегенмен бұл мәндер модельдің жоғары тиімділігін көрсетеді.



Сурет 11. KazBERT моделінің қателіктер матрицасы

Модельдердің тиімділігін салыстырмалы талдау және ең жақсы үлгіні таңдау. Алынған көрсеткіштерді салыстыру KazBERT моделінің негізгі логистикалық регрессия үлгісінен де, LSTM моделінен де барлық негізгі сапа көрсеткіштері бойынша: дәлдік, сенімділік, еске түсіру және F1-өлшемі бойынша асып түсетінін көрсетеді. Бұл үлкен көлемдегі тіл деректерінде алдын ала дайындалған және белгілі бір тілге бейімделген, мәтінді терең семантикалық талдау тапсырмаларында трансформер архитектураларының жоғары тиімділігін көрсетеді (1-кесте).

Кесте 1. Модель метрикаларының салыстырмалы кестесі

Модель метрикасы	*KazBERT*	*Логистикалық регрессия*	*LSTM*
Дәлдік, %	*93,73*	*92,86*	*92,47*
Дәлдік, %	*93.43*	*92,78*	*92,83*
Еске түсіру, %	*94.28*	*93,59*	*92.34*
F1-балл, %	*93,86*	*93,62*	*92.29*

LSTM моделі мәтіндегі дәйекті тәуелділіктерді түсіруде де жақсы жұмыс істегенімен, оның өнімділігі KazBERT-тен сәл төмен болды. Логистикалық регрессия моделі, үйренуге оңайырақ және жылдамырақ болғанымен, нейрондық желі тәсілдерінің дәлдігі, әсіресе күрделі контексттерді түсінуде болжамды түрде төмен орындалды, бірақ ол күрделірек үлгілерді бағалау үшін маңызды эталон ретінде қызмет етті. LSTM моделі логистикалық регрессиямен салыстырғанда ұзақ оқу үдерісіне қарамастан мәтін құрылымын терең түсіне алды.

KazBERT моделінің артықшылығы оның трансформер архитектурасына негізделгендігінде және қазақ тілінің ерекшеліктеріне бейімделіп, алдын ала үлкен көлемдегі қазақша мәтіндерде оқытылуында жатыр. Бұл модель, "назар аудару" (attention) механизмдерін, әсіресе "өзіне-өзі назар аудару" (self-attention) механизмін қолданады. Бұл механизмдер модельге сөйлемдегі әрбір сөзді өңдеу кезінде сол сөйлемдегі басқа сөздердің маңыздылығын бағалауға, яғни контекстті тереңірек түсінуге мүмкіндік береді. Екі бағытты кодтау (bidirectional encoding) арқасында KazBERT әрбір сөз үшін оның алдындағы және кейінгі сөздерді бір мезгілде ескереді, бұл сөздердің байытылған, контекстке негізделген векторлық ұсыныстарын (embeddings) қалыптастыруға ықпал етеді. Осының нәтижесінде KazBERT жасырын мағыналарды, сөз орамдарының нюанстарын және мәтіннің жалпы тонын дәлірек анықтай алады, бұл суицидтік ойларды білдіретін, бірақ тікелей кілт сөздерді қамтымайтын мәтіндерді

жіктеуде өте маңызды. Алдын ала оқытылуы және қазақ тілінің спецификасына бейімделуі оның жалпылама көптілді модельдерге қарағанда артықшылығын қамтамасыз етеді, өйткені ол тілдің морфологиялық және семантикалық ерекшеліктерін жақсырақ "түсінеді". Осы факторлардың жиынтығы KazBERT моделінің LSTM және логистикалық регрессияға қарағанда жоғарырақ нәтиже көрсетуіне негіз болды.

Барлық көрсеткіштер бойынша ең жоғары ұпайларды ескере отырып, KazBERT моделі суицидтік мінез-құлық белгілері бар мәтіндерді талдау және жіктеу қосымшасында одан әрі пайдалану үшін ең жақсысы болып таңдалды. Жасалған үлгілердің практикалық қолданылуын көрсету және қазақ тіліндегі мәтіндерді талдауға арналған құралмен қамтамасыз ету мақсатында интерактивті веб-қосымша әзірленді. Оның негізі ретінде ең жақсы эксперименттік нәтижелер көрсеткен KazBERT моделі таңдалды. Пайдаланушы интерфейсі Gradio кітапханасының көмегімен жүзеге асырылды.

Қосымша архитектурасы және технологиялық стек. Қосымша Python-да құрастырылған және келесі негізгі компоненттерді пайдаланады:

Жіктеу үлгісі: алдын ала оқытылған KazBERT (Eraly-ml/KazBERT) негізіндегі қайта оқытылған BertForSequenceClassification моделі және оған сәйкес BertTokenizer токенизаторы.

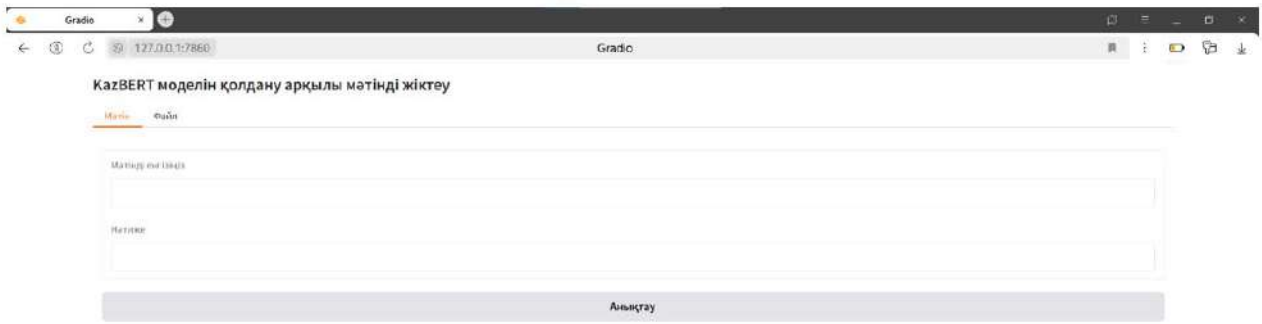
Деректерді өңдеу кітапханасы: Excel (.xlsx) файлдарын оқуға және өңдеуге арналған pandas кітапханасы.

Веб-интерфейсті жасауға арналған кітапхана: интерактивті интерфейсті жылдам орналастыруға арналған gradio кітапханасы.

Тензорлық есептеулерге арналған кітапхана: KazBERT моделінің бэкенді ретінде torch (PyTorch) кітапханасы.

Жұмысты бастамас бұрын модель мен токенизатор жүктеліп, модель бағалау режиміне қойылады (model.eval()) болжау кезінде дұрыс жұмыс істеуін қамтамасыз ету.

Функционалдылық және пайдаланушы интерфейсі. Қосымша Gradio интерфейсінде бөлек қойындылар ретінде жүзеге асырылатын мәтіндермен жұмыс істеудің екі негізгі режимін ұсынады. Жекеленген мәтінді жіктеу («Мәтін» қойындысы). «Мәтін» қойындысында пайдаланушы талданатын мәтінді енгізуі немесе көшіріп қоюы үшін мәтін өрісі беріледі (11-сурет).



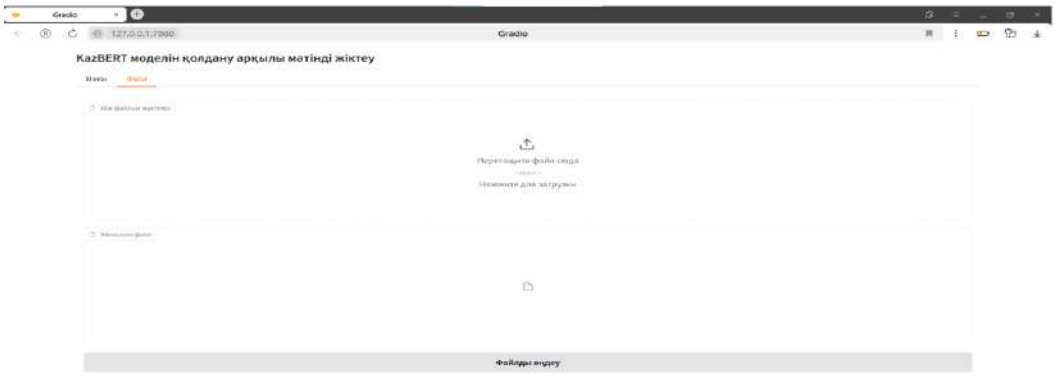
Сурет 11. Қосымшаның «Мәтін» қойындысы

Мәтінді енгізіп, «Анықтау» батырмасын басқаннан кейін енгізілген мәтін predict_single функциясына беріледі. Бұл функцияның ішінде мәтін жүктелген BertTokenizer көмегімен токенизацияланады, тензорларға түрлендіріледі және KazBERT үлгісінің кірісіне беріледі. Модель болжамды қайтарып, оның негізінде екі белгінің бірі "Нәтиже" өрісіне шығарылады: "Суицидті жазба" немесе «Суицидті емес жазба» (12-сурет).



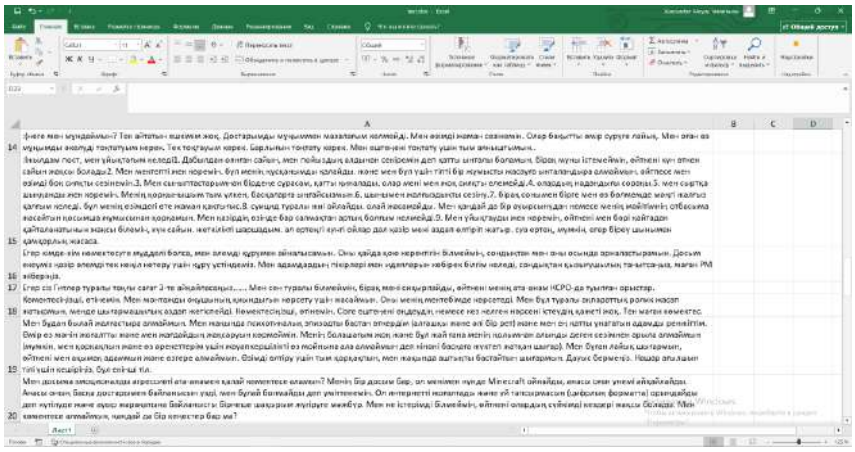
Сурет 12. Қосымшаның «Мәтін» қойындысында мәтінді жіктеу

Файлды өңдеу ("Файл" қойындысы). Пайдаланушыға файлды Excel форматында (.xlsx) «xlsx-файлын жүктеу» ұяшығына жүктеп салу мүмкіндігі беріледі (13-сурет).



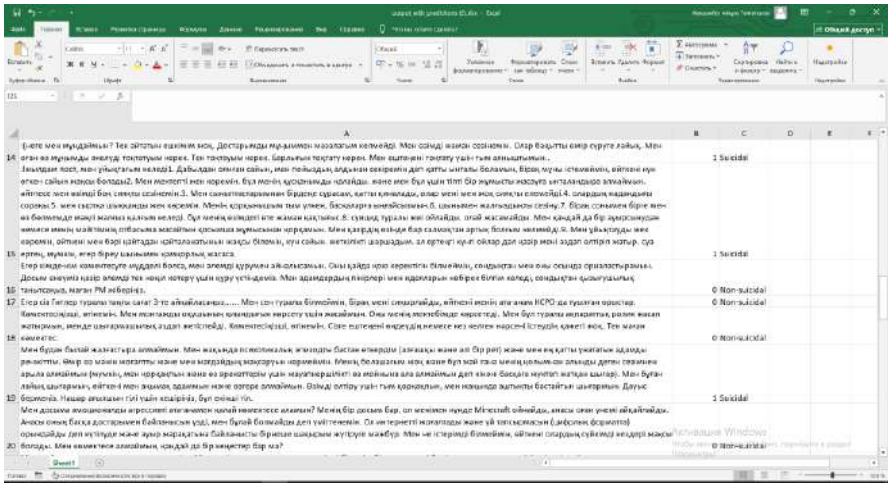
Сурет 13. Қосымшаның «Файл» қойындысы

Файлды жүктеп, «Файлды өңдеу» батырмасын басқаннан кейін predict_file функциясы іске қосылады. Функция Excel файлынан деректерді pandas кітапханасының көмегімен оқиды. Қажетті 'text' бағанының болуы тексеріледі (14-суретте жүктелінетін файлдың 'text' бағаны көрсетілген). Ол жоқ болса, «Файлда 'text' бағаны болу қажет» деген хабарлама шығады.



Сурет 14. «Файл» қойындысынан жүктеуге салынатын файл

'text' бағанындағы мәтіндерді KazBERT моделі ретімен өңдейді (токенизация және 0 немесе 1 кластарын болжау). Жүктелген DataFrame-ге екі жаңа баған қосылады: 'prediction' (0 немесе 1 сандық белгілері бар) және 'class' (сәйкесінше "Non-suicidal" немесе " Suicidal " мәтіндік белгілер көрсетіледі) (15-сурет).



Сурет 15. «Файл» қойындысынан жүктелген өңделінген файл

Жаңартылған DataFrame output_with_predictions.xlsx атты жаңа Excel файлына сақталады, содан кейін ол пайдаланушыға "Жіктелген файл" компонентінде жүктеп алуға ұсынылады (16-сурет).



Сурет 16. Қосымшаның «Файл» қойындысында файл өңдеу

Қолданбадағы үлгі бойынша мәтінді өңдеу процесі. Енгізу әдісіне қарамастан (бір мәтін немесе файл), әрбір мәтін фрагменті жіктелмес бұрын қолданба ішінде келесі өңдеу қадамдарынан өтеді:

- Токенизация: Мәтін KazBERT үшін бейімделген BertTokenizer токенизаторы арқылы токендер тізбегіне түрлендіріледі. Padding тізбектердің ұзындығын теңестіру үшін пайдаланылады және truncation тым ұзын мәтіндерді үлгі қолдайтын максималды ұзындыққа дейін қысқарту үшін қолданылады (max_length=512).

- Модельдің келтірулерін алу: Токенделген деректер BertForSequenceClassification моделінің кірісіне беріледі.

- Жіктеу: Модель екі сыныптың әрқайсысына жататын мәтіннің ықтималдығын есептейді (суицидтік/суицидтік емес).

- Нәтиженің шығуы: Ықтималдығы жоғары класс таңдалады және нәтиже пайдаланушыға сәйкес форматта ұсынылады.

Әзірленген қосымша оқытылған KazBERT үлгісін практикалық қолданудың көрнекі демонстрациясы ретінде қызмет етеді және суицидтік белгілерді анықтауда мәтіндік материалдарды жылдам бағалаудың ыңғайлы құралын ұсынады.

Қорытынды

Бұл зерттеуде әлеуметтік желілердегі посттардың суицидтік контекстін айқындар алдында көптеген машиналық оқытумен қамтып, кейіннен терең талқыланды. Мәліметтерді даярлау, мәтінді векторлау әрі нейрондық желіні пайдалана отырып қолайлы алгоритм таңдалып, өзара сәйкестендірілді. Нәтижелерді салыстырғанда BERT моделі ең жоғары дәлдік көрсетті, себебі

ол контексті терең түсінеді. LSTM алгоритмі және регрессиялық дәлдік аздап кейіндеу қалғанымен, оң көрсеткіш танытты, алғашқысы ауқымды мәтіндік массив арқылы қолайлы жұмыс істесе, келесісі шапшаң және ширақтығымен көзге түсті. Зерттеудің негізгі ресурсы Kaggle платформасындағы деректер жиынынан алынған жазбалардың ағылшын тілінен deep-translator кітапханасының көмегімен қазақ тіліне аудармасынан құралды. Зерттеудің практикалық маңыздылығы Gradio кітапханасын пайдалана отырып, KazBERT моделі негізінде веб-қосымшаны жасау арқылы айтарлықтай арта түсті. Қорыта келе, ашық ақпарат көздеріндегі өзіне қол салу фактілерін білуге тереңдетілген нейрон тораптары және үйретіліп дайын тұрған тілге қатысты алгоритмдер қолайлы болады. Әйтсе де, таңдап алынған тәсілдің қолданыс аясына қатысты ерекшеліктері ескерілген. Болашақта модельдердің тиімділігін арттыру үшін үлкен деректер жиындарын пайдалану, аугментация әдістерін енгізу және ансамбльдік тәсілдерді қарастыру жоспарлануда. Ол зерттеулердің барлығы да жасанды интеллект негізінде орындалады.

References

[1] Самоубийства в Казахстане. https://ru.wikipedia.org/?curid=4534555&oldid=146530946

[2] UNICEF Kazakhstan. (2014). Assessment of suicide preventative activities in Kazakhstan: Kyzylorda and East Kazakhstan regions. https://www.unicef.org/kazakhstan/sites/unicef.org.kazakhstan/files/2018-06/suicide_en.pdf

[3] Government of Kazakhstan. (2023). Comprehensive plan for protecting children and preventing suicide in Kazakhstan (2023-2025). https://cabar.asia/en/violence-free-schools-a-comprehensive-plan-to-protect-children-from-bullying-in-kazakhstan-white-paper

[4] Zhang, X., & Wallace, B.C. (2015). A Sensitivity Analysis of (and Practitioners\' Guide to) Convolutional Neural Networks for Sentence Classification. Proceedings of the 8th International Conference on Language Resources and Evaluation (LREC 2015), pp. 1–9.

[5] Blei, D.M., Ng, A.Y., & Jordan, M.I. (2003). Latent Dirichlet Allocation. Journal of Machine Learning Research, No 3, pp. 993-1022.

[6] Griffiths, T.L., & Steyvers, M. (2004). Finding scientific topics. Proceedings of the National Academy of Sciences, No 101(Suppl 1), pp. 5228-5235.

[7] «Study on prevalence, underlying causes, risk and protective factors in respect to suicides and attempted suicides in Kazakhstan», Astana: The UN Children's Fund (UNICEF) in the Republic of Kazakhstan, 2014. - 104 p. https://www.unicef.org/kazakhstan/sites/unicef.org.kazakhstan/files/2018-06/00001358_0.pdf

[8] De Choudhuri, S., & Jagannath, P. (2018). Comparison of Naive Bayes Classifier with Other Machine Learning Algorithms in Text Classification. International Journal of Computer Applications, No 975, pp. 1-5.

[9] Kim, Y. (2014). Convolutional Neural Networks for Sentence Classification. Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP 2014), pp. 1746–1751.

[10] Zhang, X., & Wallace, B.C. (2015). A Sensitivity Analysis of (and Practitioners\' Guide to) Convolutional Neural Networks for Sentence Classification. Proceedings of the 8th International Conference on Language Resources and Evaluation (LREC 2015), pp. 1–9.

[11] Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. Neural Computation, Vol. 9, No 8, pp. 1735–1780.

[12] Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. Proceedings of NAACL-HLT 2019, pp. 4171–4186.

[13] Manning, C. D., Surdeanu, M., Bauer, J., Finkel, H., Bethard, S., & McClosky, D. (2014). The Stanford CoreNLP Natural Language Processing Toolkit. Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics (ACL 2014), pp. 55–60.

[14] Sarker, I. H., & Riazi, H. (2020). KazBERT: A Pre-Trained Transformer Model for Kazakh Text Classification. Proceedings of the International Conference on Artificial Intelligence and Computer Vision, pp. 102–111.

[15] Conneau, A., Kruszewski, G., Lample, G., Barrault, L., & Joulin, A. (2020). Unsupervised Cross-lingual Representation Learning at Scale. Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (ACL 2020), pp. 8440–8451.

IRSTI 28.23.25

https://doi.org/10.51889/2959-5894.2026.95.3.020

D. Lebedev[1], Zh. Otarbay[1], A. Kyzyrkanov[1*]

[1]Al-Farabi Kazakh National University, Almaty, Kazakhstan
[*]*e-mail: abzzall@gmail.com*

A REPRODUCIBLE 5G SA TESTBED FOR O-RAN MACHINE LEARNING

Abstract

Data-driven control of the radio access network (RAN) through RAN Intelligent Controllers (RICs) in the Open Radio Access Network (O-RAN) architecture depends on labelled, stack-wide measurement data, which is scarce and costly to obtain on commercial equipment. This paper describes a fully software-defined fifth-generation (5G) Standalone (SA) testbed and an automated pipeline that turns it into a generator of labelled machine-learning datasets. The testbed integrates the srsRAN gNodeB, the Open5GS core, a GNU Radio plus ZeroMQ radio emulator with runtime-controllable channel impairment, and a FlexRIC near-real-time RAN Intelligent Controller (Near-RT RIC), all on a single commodity server with central processing unit (CPU) isolation and per-user network namespaces for repeatability. A Monte-Carlo orchestrator runs up to one thousand 20-second iterations, randomizing user count, per-user path loss, traffic type and rate, and parallelism, and a synchronized parser assembles one-second snapshots into a 37-feature table spanning the physical (PHY), medium access control (MAC), radio link control (RLC), packet data convergence protocol (PDCP), and operating-system layers, augmented with five-second rolling statistics, an energy-efficiency index, and forward-looking regression and classification targets at a five-second horizon. We describe the architecture, the feature engineering, the labelling methodology, and the resulting dataset structure, and we discuss how the dataset supports predictive xApps for proactive congestion and degradation control. The platform and pipeline are designed for reproducibility and are intended as a reusable resource for the O-RAN machine-learning community.

Keywords: 5G Standalone, O-RAN, RAN Intelligent Controller, machine-learning dataset, feature engineering, srsRAN, FlexRIC, testbed, reproducibility.

Д. Лебедев[1], Ж. Отарбай[1], А. Кызырканов[1]
[1]Казахский национальный университет имени Аль-Фараби, г.Алматы, Казахстан

ВОСПРОИЗВОДИМЫЙ ТЕСТОВЫЙ СТЕНД 5G SA ДЛЯ МАШИННОГО ОБУЧЕНИЯ В O-RAN

Аннотация

Управление радиосетью на основе данных через интеллектуальные контроллеры O-RAN RAN Intelligent Controllers требует размеченных измерительных данных по всему стеку протоколов, которые трудно и дорого получать на коммерческом оборудовании. В данной статье описывается полностью программно-определяемый тестовый стенд 5G Standalone (SA) и автоматизированный конвейер, который превращает его в генератор размеченных датасетов для машинного обучения. Тестовый стенд объединяет srsRAN gNodeB, ядро Open5GS, радиоэмулятор GNU Radio и ZeroMQ с управляемыми в реальном времени ухудшениями канала, а также FlexRIC Near-RT RIC. Все компоненты работают на одном сервере общего назначения с изоляцией CPU и отдельными сетевыми пространствами имен для каждого пользователя, что обеспечивает повторяемость экспериментов. Monte-Carlo-оркестратор выполняет до одной тысячи итераций длительностью 20 секунд, случайным образом изменяя количество пользователей, потери в канале для каждого пользователя, тип и скорость трафика, а также степень параллелизма. Синхронизированный парсер собирает односекундные снимки состояния в таблицу из 37 признаков, охватывающих уровни PHY, MAC, RLC/PDCP и операционной системы. Таблица дополнена пятисекундными скользящими статистиками, индексом энергоэффективности, а также регрессионными и классификационными целевыми переменными с горизонтом прогнозирования пять секунд. В статье описываются архитектура, инженерия признаков, методология разметки и структура полученного датасета, а также обсуждается, как этот датасет может поддерживать предиктивные xApps для проактивного управления перегрузками и деградацией

качества. Платформа и конвейер разработаны с учетом воспроизводимости и предназначены как многоразовый ресурс для сообщества O-RAN machine learning.

Ключевые слова: 5G Standalone, O-RAN, RAN Intelligent Controller, датасет для машинного обучения, инженерия признаков, srsRAN, FlexRIC, тестовый стенд, воспроизводимость.

Д. Лебедев[1], Ж. Отарбай[1], А. Кызырканов[1, *]
[1]Әл-Фараби атындағы Қазақ ұлттық университеті, Алматы қ., Қазақстан

O-RAN ЖҮЙЕСІНДЕГІ МАШИНАЛЫҚ ОҚЫТУҒА АРНАЛҒАН ҚАЙТА ЖАҢҒЫРТЫЛАТЫН 5G SA ТЕСТІЛІК СТЕНДІ

Аңдатпа

O-RAN жүйесіндегі RAN Intelligent Controllers арқылы радиоқолжетімділік желісін деректерге негізделген түрде басқару үшін протоколдар стегінің барлық деңгейлерін қамтитын таңбаланған өлшеу деректері қажет. Мұндай деректерді коммерциялық жабдықта алу қиын әрі қымбат. Бұл мақалада толық бағдарламалық түрде анықталған 5G Standalone (SA) тестілік стенді және оны машиналық оқытуға арналған таңбаланған датасеттер генераторына айналдыратын автоматтандырылған конвейер сипатталады. Тестілік стенд srsRAN gNodeB, Open5GS ядросын, арна параметрлерін нақты уақытта басқаруға мүмкіндік беретін GNU Radio және ZeroMQ негізіндегі радиоэмуляторды, сондай-ақ FlexRIC Near-RT RIC компонентін біріктіреді. Барлық компоненттер қайталанатын эксперименттерді қамтамасыз ету үшін CPU оқшаулауы және әр пайдаланушыға арналған жеке желілік namespace қолданылатын бір әмбебап серверде жұмыс істейді. Monte-Carlo оркестраторы ұзақтығы 20 секундқа дейінгі бір мыңға жуық итерацияны орындайды, әр итерацияда пайдаланушылар санын, әр пайдаланушыға тән жолдағы жоғалтуды, трафик түрі мен жылдамдығын және параллелизм деңгейін кездейсоқ өзгертеді. Синхрондалған парсер бір секундтық жүйе күйінің снимоктарын PHY, MAC, RLC/PDCP және операциялық жүйе деңгейлерін қамтитын 37 белгіден тұратын кестеге жинақтайды. Бұл кесте бес секундтық жылжымалы статистикалармен, энергия тиімділігі индексімен және бес секундтық болжау горизонты бар регрессиялық және классификациялық мақсатты айнымалылармен толықтырылады. Мақалада архитектура, белгілерді құрастыру, таңбалау әдістемесі және алынған датасеттің құрылымы сипатталады, сондай-ақ бұл датасеттің желідегі кептеліс пен сапа нашарлауын проактивті басқаруға арналған предиктивті xApps әзірлеуге қалай көмектесетіні талқыланады. Платформа мен конвейер қайта жаңғыртылуды қамтамасыз ету мақсатында жасалған және O-RAN машиналық оқыту қауымдастығы үшін қайта пайдалануға болатын ресурс ретінде ұсынылады.

Түйін сөздер: 5G Standalone, O-RAN, RAN Intelligent Controller, машиналық оқыту датасеті, белгілерді құрастыру, srsRAN, FlexRIC, тестілік стенд, қайта жаңғыртылу.

Introduction

Key Provisions

This study introduces a reproducible software-defined 5G Standalone testbed and an automated pipeline for generating labelled machine-learning datasets for O-RAN applications. The platform integrates srsRAN, Open5GS, GNU Radio, ZeroMQ, and FlexRIC on a single server and collects synchronized measurements from different protocol-stack layers. The resulting dataset contains approximately 19,000 labelled snapshots with more than 45 features, rolling statistics, energy-efficiency indicators, and forward-looking prediction targets. The proposed platform provides a reusable foundation for developing predictive xApps capable of detecting network congestion and performance degradation in advance.

The Open Radio Access Network (O-RAN) architecture adds programmable radio access network (RAN) Intelligent Controllers (RICs) that carry machine-learning (ML) applications (xApps) in the near-real-time RIC (Near-RT RIC), and radio applications (rApps) in the non-real-time RIC (non-RT RIC) for RAN optimization through closed-loop control [1,2]. Whether this vision becomes a reality depends on a resource that most takes it for granted: labeled high-resolution measurement data spanning the entire protocol stack and characterizing network performance under various operating conditions. Indeed, the most typical ML workflow for O-RAN starts by collecting exactly such “step data” across many RAN configurations, before any ML models can be conceived, trained or deployed [3,4]. This data is often challenging to collect: commercial equipment exposes only a

limited subset of internal telemetry, experimental deployments in live networks are prohibitive and public datasets that combinePHY, MAC, transport and hardware metrics with “forward-looking” labels are extremely rare.

Software-defined cellular stacks provides one possible solution. Thanks to the availability of open-source software-defined fifth-generation (5G) core (Open5GS) [5], gNodeB (srsRAN) [6], and O-RAN RIC (FlexRIC) [7] implementations, an end-to-end 5G Standalone (SA) network can be built using off-the-shelf hardware, and instrumented extensively [8,9,10]. Such software-defined networks have already been used for profiling RAN components [11], developing ML-based xApps for closed-loop control [12], and for deploying campus-scale O-RAN testbeds [9]. Existing work on RIC intelligence, such as traffic steering [13,14], prediction [15], and key performance indicator (KPI)-driven optimization [16], rely heavily on relevant training data, and O-RAN surveys confirm that lack of labeled data and relevant features is a practical challenge [17]. However, less focus has been dedicated to automatically generate such labeled stack-wide dataset from a testbed in a reproducible manner, with all necessary documentation.

In this paper, we tackle this problem head-on. We introduce a software-defined 5G SA testbed, and, more importantly, an automated pipeline that transforms it into a dataset generator for ML-based RAN control. Our contributions are threefold. Firstly, we describe a reproducible testbed, comprising srsRAN, Open5GS, a runtime-controllable radio emulator (GNU Radio + ZeroMQ), and a FlexRIC Near-RT RIC, deployed on a single, central processing unit (CPU)-isolated server. Secondly, we present an automated data-generation pipeline using a Monte-Carlo approach, where various parameters of the network are randomized. We synchronize a stack-wide parser that collects per-second samples and aggregates them into a table of 37 features from four different layers of the protocol stack, along with rolling statistics and an energy efficiency metric. Thirdly, we propose a forward-looking methodology for assigning ground truth labels to our generated data, both in the form of a regression target and a binary class indicative of network degradation, at a 5-second horizon, enabling proactive xApp development. We provide details on our design, the resulting data format, and its potential use cases, keeping reproducibility at the forefront of our efforts.

The remainder of this paper is structured as follows. Section 2 outlines the architecture of our testbed, data generation pipeline, and feature and label design. Section 3 presents the dataset structure, scope, reproducibility, and data-quality results. Section 4 discusses its applicability to predictive RAN control. Section 5 concludes the paper.

Research methodology

Testbed architecture

The testbed realizes a complete 5G SA network in software on a single commodity server, with no dedicated radio hardware. Four domains compose the system, as shown in Figure 1.

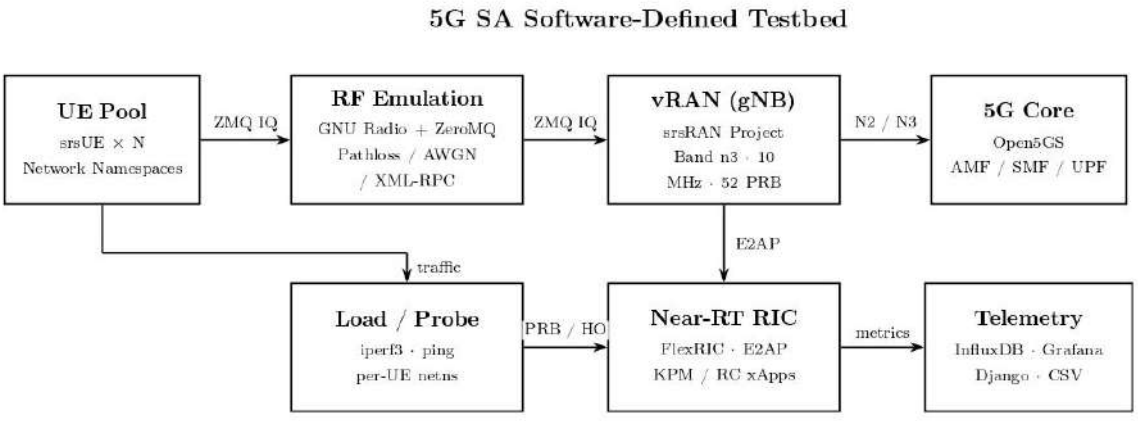


Figure 1.5G SA testbed architecture

The core network is provided by Open5GS [5], running the access and mobility management, session management, and user-plane functions as independent daemons that communicate over standardized Third Generation Partnership Project (3GPP) interfaces on loopback addresses. The radio access network is the srsRAN Project gNodeB [6], operating in 5G New Radio (NR) Standalone mode on band n3 with a 10 MHz channel (52 physical resource blocks). The air interface is virtualized: instead of physical radios, the gNodeB and the srsUE instances exchange baseband in-phase and quadrature (IQ) samples over ZeroMQ sockets, which guarantees deterministic, repeatable experiments isolated from external interference [8]. A GNU Radio flow graph [18] interposed on the IQ path applies controllable path loss and additive white Gaussian noise (AWGN) and exposes an Extensible Markup Language Remote Procedure Call (XML-RPC) interface so that channel conditions can be altered at runtime without restarting the cell.

It takes two engineering steps to configure the platform as a measurement instrument. One step is pinning the latency sensitive physical-layer and ZeroMQ-broker threads on non-operating-system (OS) scheduler cores and isolating the network interface interrupts to non-isolated cores, making sure that the time-series have no scheduler jitter. The second step is isolating the per-user traffic generators to their own network namespaces inside the OS, emulating each user within their own Linux network namespace and thus separating the traffic generators at the network stack layer. We sample telemetry from two sources:

(1) The Near-RT RIC of FlexRIC [7] pulls the scheduler internals per second through the E2 interface and its key performance measurement (KPM) and radio control (RC) service models.

(2) A simple custom daemon is run from within the OS's root which samples the hardware and OS metrics once per second by reading from kernel pseudo-files including processor power read from the Intel Running Average Power Limit (RAPL) counters [19].

Automated dataset-generation pipeline

To collect a representative dataset rather than a handful of fixed operating points, the testbed is driven by a Monte-Carlo orchestration script that executes up to one thousand independent iterations, each lasting 20 seconds. At the start of every iteration the orchestrator draws the experimental conditions at random from defined distributions, summarized in Table 1: the number of active users, a per-user base path loss that emulates the user’s distance from the cell centre (driven by a random-walk mobility process), the offered traffic rate, the transport protocol mix, and the number of parallel flows. Randomizing these factors jointly produces the diversity of network states that a robust ML model requires.

Table 1. Stochastic parameters randomized per iteration by the Monte-Carlo orchestrator

Parameter	*Range / distribution*	*Purpose*
User count	*Uniform 1-5*	*Dynamic attach/detach of UEs*
Path loss, UE 1	*25-42 dB*	*Near zone (random walk)*
Path loss, UE 2	*38-58 dB*	*Mid zone*
Path loss, UE 3	*50-70 dB*	*Far zone*
Path loss, UE 4	*65-82 dB*	*Cell edge / handover candidate*
iPerf bitrate	*2-18 Mbps*	*Within ZMQ stability range*
Transport mix	*80% UDP / 20% TCP*	*Streaming vs. data transfer*
Parallel flows	*1 or 4*	*Single vs. concurrent sockets*
Warmup discard	*first 3 s*	*Remove transient TCP ramp-up*

A carefully planned warm-up period is part of data quality: the first three seconds out of every 20-second run of a 20s iteration are filtered out from the parser. During the transient, the network is setup, establishes sessions, transmission control protocol (TCP) window grows exponentially and the scheduler heuristic are not stabilized, and filtering these data ensures that we are logging in to a stable

network environment, and not just artifacts at startup. A timestamp synchronization parser reprojects timestamps to the millisecond-level for the gNodeB, the iPerf[20], ping, and system daemon data and produce one line per second corresponding to one system state.

Feature engineering and labelling

Each snapshot row comprises 37 features extracted directly from the internal interfaces of the core, the virtualized radio access network (vRAN), and the server hardware, organized by stack layer as shown in Figure 2. The physical layer contributes channel-quality and link metrics: channel quality indicator (CQI), downlink and uplink modulation and coding scheme (MCS), reference signal received power (RSRP), physical uplink shared channel (PUSCH) signal-to-noise ratio (SNR), timing advance, the real radio bitrate, and the block error rate. The medium access control (MAC) layer contributes protocol data unit (PDU) sizes, logical-channel byte counts, and the Buffer Status Report, following the 5G NR MAC protocol specification [21], a high-value signal of pending uplink data. The radio link control (RLC) and packet data convergence protocol (PDCP) layers contribute sequence numbers, packet counts, and per-bearer encrypted volumes. The operating-system layer contributes gNodeB CPU utilization, free memory, interrupt rate, and processor power from RAPL.

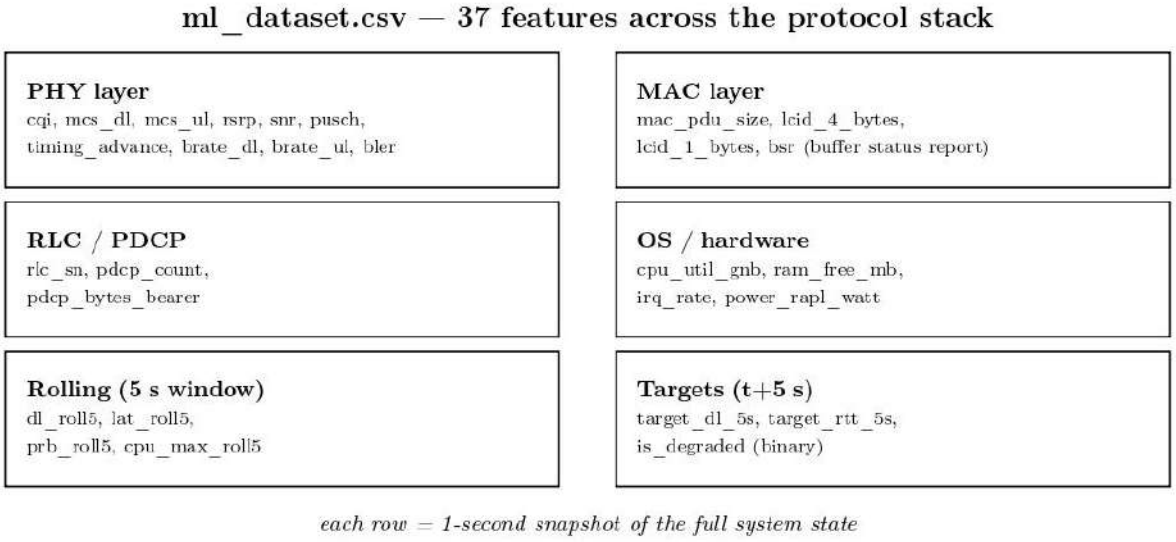


Figure 2. Dataset feature structure

To give models temporal context, the parser computes five-second rolling statistics (moving average, maximum, and variance) for the key quantities bitrate, latency, physical resource block (PRB) load, and CPU peak. It also derives a network energy-efficiency index defined in Equation (1) as the radio throughput delivered per watt of processor power:

$$E_{eff} = \frac{R_{radio}\ (\text{Mbps})}{P_{RAPL}\ (\text{W})} \tag{1}$$

For predictive control, the dataset is labelled with targets shifted forward by a fixed five-second horizon: target_dl_5s and target_rtt_5s record the throughput and round-trip time the cell will exhibit five seconds later, supporting regression. A binary degradation class supports classification and is the key label for early warning xApps. As shown in Equation (2), it is set to one when the predicted round-trip time exceeds the current value by more than fifty percent:

$$y_{deg} = 1 \text{ if } RTT_{t+5} > 1.5 \cdot RTT_t, \quad \text{else } 0 \tag{2}$$

This label lets a classifier learn to anticipate the onset of queue buildup before it manifests, which is precisely the signal a proactive RIC needs to act in time. The full row schema concatenates the snapshot features, the rolling statistics, the energy index, and the targets into a single record written to a consolidated comma-separated values (CSV) file.

Results of the study

Dataset structure and scope

The pipeline yields a single consolidated file, ml_dataset.csv, in which every row is a one-second, timestamp-aligned snapshot tagged with the iteration's scenario, run, and randomized conditions. Table 2 summarizes the volume produced by the two collection phases. A deterministic phase over twelve profiling scenarios, executed in triplicate at 60 seconds each, contributes on the order of two thousand rows of carefully controlled states; the stochastic Monte-Carlo phase contributes the bulk of the data, on the order of seventeen thousand rows of randomized states after warmup removal. Together they give roughly nineteen thousand labelled snapshots spanning a wide range of user counts, channel conditions, and traffic mixes.

Table 2. Dataset volume by collection phase

Phase	*Scenarios*	*Runs*	*Duration*	*Rows (est.)*
Deterministic	*12*	*3 each*	*60 s*	*~2,160*
Stochastic (Monte-Carlo)	*up to 1000 iter.*	*-*	*17 s (post-warmup)*	*~17,000*
Total	*-*	*-*	*-*	*~19,160*

Table 3 lists the column schema grouped by origin. The 37 base features are complemented by four rolling-statistic columns, the energy-efficiency index, and three target columns, for a total of more than forty-five columns per row. Because the features are drawn from every layer of the stack simultaneously and aligned in time, the dataset captures cross-layer relationships for example between MAC Buffer Status Reports, PRB utilization, and subsequent latency that single-layer measurements cannot express.

Table 3. Column schema of ml_dataset.csv grouped by origin

Group	*Representative columns*	*Count*
PHY	*cqi, mcs_dl, mcs_ul, rsrp, snr_pusch, timing_advance, brate_dl, brate_ul, bler*	*9*
MAC	*mac_pdu_size, lcid_4_bytes, lcid_1_bytes, bsr*	*4*
RLC / PDCP	*rlc_sn, pdcp_count, pdcp_bytes_bearer*	*3*
OS / hardware	*cpu_util_gnb, ram_free_mb, irq_rate, power_rapl_watt*	*4*
Context / KPI	*ue_count, load, radio, control, dl, ul, lat_mean, lat_p95, loss, prb*	*10+*
Rolling (5 s)	*dl_roll5, lat_roll5, prb_roll5, cpu_max_roll5*	*4*
Energy	*e_eff*	*1*
Targets	*target_dl_5s, target_rtt_5s, is_degraded*	*3*

Reproducibility and data quality

Three design choices underpin the quality of the resulting data. ZeroMQ-based radio emulation makes every experiment bit-exact and free of the uncontrolled variability of over-the-air measurement, so that a given set of conditions can be reproduced precisely. CPU isolation removes scheduler-induced timing noise from the latency and throughput records, which is essential because the dataset's value rests on accurate one-second aggregation. The warmup-discard rule removes the non-stationary start-up transient, ensuring that the snapshots represent settled network states rather

than session-establishment artifacts. The per-user network namespaces guarantee that multi-user traffic is genuinely independent at the stack level rather than multiplexed through a shared interface. Together these measures yield a dataset whose labels reflect the network's intrinsic behavior rather than measurement noise.

A natural concern for any emulated platform is external validity. The ZeroMQ link imposes a throughput ceiling that does not exist in a real radio, and the path-loss-plus-AWGN channel is simpler than a full fading model. These limit the absolute realism of throughput magnitudes, but they do not undermine the dataset's purpose: the features capture the relative dynamics of buffers, scheduling, and link adaptation, which are the relationships predictive RIC models must learn. The platform is also extensible the GNU Radio graph already supports fading, shadowing, Doppler, and interference blocks so richer channel realism can be added without changing the pipeline.

Discussion

Supporting predictive RAN control

The dataset is constructed specifically to train the proactive xApps that O-RAN envisions. The forward-shifted regression targets allow a model to forecast throughput and latency five seconds ahead, while the binary degradation label of Equation (2) frames early-warning as a supervised classification problem: predict, from the current cross-layer state, whether latency is about to inflate. Gradient-boosted trees such as XGBoost [22] are a natural fit for the tabular, mixed-type feature space, while recurrent architectures such as long short-term memory (LSTM) [23] or gated recurrent unit (GRU) [24] can exploit the temporal context encoded in the rolling features and the sequence of snapshots. A model trained on this data and deployed as an xApp in the Near-RT RIC could act on the degradation signal for example by adjusting PRB allocation or limiting aggressive flows through the RC service model before queue buildup becomes user-visible. The energy-efficiency index additionally enables joint optimization of performance and power, aligning with the growing emphasis on sustainable RAN operation [16].

Far more than one lab, what makes the pipeline useful is that it is reusable. Since the testbed is constructed from completely open-source parts, and documented how to be reproduced, other researchers can also reproduce similar datasets, expand on feature coverage, or test other conditions, fostering the type of sharing and comparison-of-apples experiment that the O-RAN ML community calls for. As such, we consider the data-generating pipeline itself the primary deliverable, more so than any specific trained model. It makes the reusable testbed into a renewable source of labeled, stack-wide data to train stack-wide data-driven control systems.

Conclusion

We showcased a software-defined 5G Standalone testbed and an automated workflow to make it a stack-wide labelled ML data generator for O-RAN. The testbed uses srsRAN, Open5GS, a GNU Radio/ZeroMQ based radio emulator with runtime-controlled impairments and a FlexRIC Near-RT RIC on a single, CPU-isolated server with per-user network namespaces to achieve repeatability. A Monte-Carlo orchestrator is used to randomize the number of users, pathloss, traffic load and level of parallelism over several brief iterations and a synchronized parser combines these over one-second snapshots into a 37 feature-tablecovering PHY, MAC, RLC/PDCP and OS layers; complemented by 5s rolling-average stats, an energy efficiency index, and 5s forward looking regressors and classifiers. This yields a reproducible source of 19K-odd, labelled snapshot tables for training proactive xApps on congestion and degradation. We intend to use this dataset in future work to train and run such xApps closed loop and to further enhance the channel emulations with impairments and fading.

Acknowledgment

This research was funded by the Committee of Science of the Ministry of Science and Higher Education of the Republic of Kazakhstan (Grant No.BR24993211).

References

[1] O-RAN Alliance, "O-RAN Architecture Description," O-RAN.WG1.OAD, v11.0, 2024. URL: https://www.o-ran.org/specifications.

[2] M. Polese, L. Bonati, S. D'Oro, S. Basagni, and T. Melodia, "Understanding O-RAN: Architecture, interfaces, algorithms, security, and research challenges," IEEE Communications Surveys & Tutorials, vol. 25, no. 2, pp. 1376-1411, 2023. DOI: 10.1109/COMST.2023.3239220.

[3] L. Bonati, S. D'Oro, M. Polese, S. Basagni, and T. Melodia, "Intelligence and learning in O-RAN for data-driven NextG cellular networks," IEEE Communications Magazine, vol. 59, no. 10, pp. 21-27, 2021. DOI: 10.1109/MCOM.101.2001120.

[4] O-RAN Working Group 2, "AI/ML workflow description and requirements," O-RAN.WG2.AIML, v1.03, 2021. URL: https://www.o-ran.org/specifications.

[5] Open5GS, "Open5GS Documentation: Open-source 5G core network," 2024. [Online]. Available: https://open5gs.org

[6] srsRAN Project, "srsRAN Project Documentation," Software Radio Systems, 2024. [Online]. Available: https://docs.srsran.com

[7] R. Schmidt, M. Irazabal, and N. Nikaein, "FlexRIC: An SDK for next-generation SD-RANs," in Proc. ACM CoNEXT, 2021, pp. 411-425. DOI: 10.1145/3485983.3494870.

[8] L. Mamushiane, A. Lysko, H. Kobo, and J. Mwangama, "Deploying a stable 5G SA testbed using srsRAN and Open5GS: UE integration and troubleshooting towards network slicing," in Proc. IEEE AFRICON, 2023. DOI: 10.1109/icABCD59051.2023.10220512.

[9] C. Sun et al., "Campus5G: A campus scale private 5G Open RAN testbed," arXiv preprint arXiv:2506.23740, 2025. DOI: 10.48550/arXiv.2506.23740.

[10] NIST, "Blueprint for deploying 5G O-RAN testbeds," NIST Technical Note 2311, 2024. URL: https://doi.org/10.6028/NIST.TN.2311.

[11] C. Wei, A. Kak, N. Choi, and T. Wood, "5GPerf: Profiling open-source 5G RAN components under different architectural deployments," in Proc. ACM SIGCOMM Workshop on 5G and Beyond, 2022, pp. 43-49. DOI: 10.1145/3538394.3546044.

[12] L. Bonati, M. Polese, S. D'Oro, S. Basagni, and T. Melodia, "ColO-RAN: Developing machine learning-based xApps for Open RAN closed-loop control on programmable experimental platforms," IEEE Transactions on Mobile Computing, vol. 22, no. 10, pp. 5787-5800, 2023. DOI: 10.1109/TMC.2022.3188013.

[13] A. Lacava et al., "Programmable and customized intelligence for traffic steering in 5G networks using Open RAN architectures," IEEE TMC, vol. 23, no. 4, pp. 2882-2897, 2024. DOI: 10.1109/TMC.2023.3266642.

[14] I. Tamim, S. Aleyadeh, and A. Shami, "Intelligent O-RAN traffic steering for URLLC through deep reinforcement learning," arXiv preprint arXiv:2303.01960, 2023. DOI: 10.48550/arXiv.2303.01960.

[15] Transformer-based wireless traffic prediction and network optimization in O-RAN, arXiv preprint arXiv:2403.10808, 2024. DOI: 10.48550/arXiv.2403.10808.

[16] Power-efficient RAN intelligent controllers through optimized KPI monitoring, arXiv preprint arXiv:2501.09509, 2025. DOI: 10.48550/arXiv.2501.09509.

[17] Recent advances in machine learning for network automation in the O-RAN, Sensors, vol. 23, no. 21, art. 8792, 2023. DOI: 10.3390/s23218792.

[18] GNU Radio Project, "GNU Radio: The free and open software radio ecosystem," 2024. [Online]. Available: https://www.gnuradio.org

[19] Intel Corporation, " RAPL energy reporting," Intel 64 and IA-32 Architectures Software Developer's Manual, 2023. URL: https://www.intel.com/content/www/us/en/developer/articles/technical/intel-sdm.html.

[20] A. Tirumala, M. Gates, et al., "iPerf: TCP/UDP bandwidth measurement tool," 2024. [Online]. Available: https://iperf.fr

[21] 3GPP, "NR; Medium Access Control (MAC) protocol specification," 3GPP TS 38.321, V18.0.0, 2024. URL: https://www.3gpp.org/ftp/Specs/archive/38_series/38.321/.

[22] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in Proc. ACM SIGKDD, 2016, pp. 785-794. DOI: 10.1145/2939672.2939785.

[23] S. Hochreiter and J. Schmidhuber, "Long short-term memory," Neural Computation, vol. 9, no. 8, pp. 1735-1780, 1997. DOI: 10.1162/neco.1997.9.8.1735.

[24] K. Cho et al., "Learning phrase representations using RNN encoder-decoder for statistical machine translation," in Proc. EMNLP, 2014, pp. 1724-1734. DOI: 10.3115/v1/D14-1179.

IRSTI 20.23.19

https://doi.org/10.51889/2959-5894.2026.95.3.021

A.G. Ospan[1], T.S. Sarsembayeva[1], A.Zh. Sailau[*1], A.B. Temirkhanova[1], I. Felde[2]

[1]Al-Farabi Kazakh National University, Almaty, Kazakhstan
[2]Obuda University, Budapest, Hungary
[*]*e-mail: Sailau.Aisha@kaznu.kz*

OPTIMIZING CHUNKING STRATEGIES AND INFERENCE ACCELERATION FOR KAZAKH LOW-RESOURCE RAG SYSTEMS

Abstract

Retrieval-Augmented Generation (RAG) has emerged as the dominant paradigm for grounding large language model (LLM) outputs in external knowledge, yet its systematic application to morphologically rich, low-resource languages remains largely unexplored. This paper addresses two under-studied problems for Kazakh-language RAG: the design of semantically coherent document chunking strategies and the acceleration of multi-stage inference on GPU hardware. We construct a corpus of 201,202 preprocessed news documents from the national newspaper Egemen Qazaqstan (2017–2025, 129 thematic categories) and propose three chunking strategies – Fixed-size (baseline), Semantic (cosine-similarity boundary detection), and Adaptive (joint semantic and length-regularized scoring) – evaluated under a four-dimensional quality framework measuring coherence, document coverage, boundary accuracy, and length regularity. The Adaptive strategy achieves an integrated $ChunkScore$ of 0.879, outperforming the Fixed-size baseline by 0.045 absolute, and improves retrieval F1@5 by 0.067 and end-to-end answer accuracy by 4 percentage points (92%) on a 50-query factual test set. For inference acceleration, we deploy an exact GPU-resident index (FAISS-GpuIndexFlatIP) on an NVIDIA H200 GPU, achieving 2.4 ms retrieval latency – a 70.8× speedup over ChromaDB – without accuracy degradation. Combined with INT4-quantised generation ($q4_{K_M}$), KV cache residency in HBM3, and an attention-based self-evaluation mechanism, the pipeline achieves a mean end-to-end response time of 0.92 ± 0.06 s under cold-cache conditions and approximately 0.20 s with full KV cache reuse (4.3× speedup), on a single GPU without external API dependencies. These results demonstrate the feasibility of high-accuracy, sub-second Kazakh RAG and establish reproducible baselines for future low-resource retrieval research.

Keywords: retrieval-augmented generation; document chunking; low-resource NLP; Kazakh language; dense vector retrieval; FAISS; KV cache optimization; LLM inference acceleration; multilingual text embeddings; GPU-accelerated NLP.

Ә.Ғ. Оспан[1], Т.С. Сарсембаева[1], А.Ж. Сайлау[1], А.Б. Темірханова[1], И. Фелде[2]
[1]Әл-Фараби атындағы Қазақ ұлттық университеті, Алматы қ., Казахстан
[2]Обуда университеті, Будапешт қ., Мажарстан

ҚАЗАҚ ТІЛІНДЕГІ РЕСУРСЫ ШЕКТЕУЛІ RAG ЖҮЙЕЛЕРІНЕ АРНАЛҒАН ЧАНКИНГ СТРАТЕГИЯЛАРЫН ОҢТАЙЛАНДЫРУ ЖӘНЕ ИНФЕРЕНСТІ ЖЕДЕЛДЕТУ

Аңдатпа

Retrieval-Augmented Generation (RAG) архитектурасы үлкен тілдік модельдердің (LLM) жауаптарын сыртқы білім көздерімен негіздеудің негізгі тәсілі ретінде кеңінен қолданылуда. Дегенмен, бұл технологияны морфологиялық құрылымы күрделі әрі ресурстық қамтамасыз етілуі шектеулі тілдерге бейімдеу мәселесі жеткілікті деңгейде зерттелмеген. Осы мақалада қазақ тіліндегі RAG жүйелеріне қатысты екі өзекті ғылыми мәселе қарастырылады: құжаттарды семантикалық тұрғыдан үйлесімді фрагменттерге бөлу стратегияларын әзірлеу және графикалық үдеткіштерде орындалатын көпсатылы инференс үдерісін жеделдету. Зерттеуде Egemen Qazaqstan (2017–2025, 129 тақырыптық санат) ұлттық газетінің алдын ала өңделген 201 202 жаңалық құжатынан тұратын корпус құрылды. Үш чанктеу стратегиясы ұсынылды: Fixed-size (базалық), Semantic (косинустық ұқсастық бойынша шекара анықтау), Adaptive (семантикалық және ұзындық бойынша реттелген біріктірілген бағалау). Бағалау төрт өлшем бойынша жүргізілді: когеренттілік, құжатты қамту, шекара дәлдігі, ұзындық тұрақтылығы.

Adaptive әдіс ChunkScore = 0.879 нәтижесіне қол жеткізіп, Fixed-size базасынан 0.045 жоғары болды, сондай-ақ retrieval F1@5 көрсеткішін 0.067-ге, end-to-end дәлдікті 4 пайыздық тармаққа (92%) арттырды (50 сұраулы тест жиыны).Инференсті жеделдету үшін NVIDIA H200 GPU-де толық GPU жадында орналасқан дәл индекс (FAISS-GpuIndexFlatIP) қолданылды, іздеу кідірісі 2.4 ms деңгейіне жетіп, ChromaDB-мен салыстырғанда 70.8× жылдамдау қамтамасыз етті. INT4 кванталған генерация, HBM3-тегі KV-cache және attention-based self-evaluation механизмімен бірге жүйе суық кэште 0.92 ± 0.06 s, толық KV-cache қайта қолданылғанда ~0.20 s (4.3× жылдамдау) орташа жауап беру уақытын көрсетті. Жүйе бір GPU-де, сыртқы API-сыз жұмыс істейді. Нәтижелер жоғары дәлдікті, субсекундтық қазақ тіліндегі RAG жүйесін құрудың мүмкін екенін көрсетеді және low-resource retrieval зерттеулері үшін базалық репродуктивті нәтижелерді ұсынады.

Түйін сөздер: retrieval-augmented generation; құжаттарды фрагменттеу; ресурсы шектеулі тілдерге арналған табиғи тілді өңдеу; қазақ тілі; тығыз векторлық іздеу; FAISS; KV-кэшті оңтайландыру; LLM инференсін жеделдету; көптілді мәтіндік эмбеддингтер; GPU-жеделдетілген табиғи тілді өңдеу.

Ә.Ғ. Оспан[1], Т.С. Сарсембаева[1], А.Ж. Сайлау[1], А.Б. Темірханова[1], И. Фелде[2]
[1]Казахский национальный университет имени аль-Фараби, г.Алматы, Казахстан
[2]Университет Обуда, Будапешт, г.Венгрия.

ОПТИМИЗАЦИЯ СТРАТЕГИЙ ЧАНКИНГА И УСКОРЕНИЕ ИНФЕРЕНСА ДЛЯ КАЗАХОЯЗЫЧНЫХ RAG-СИСТЕМ В УСЛОВИЯХ ОГРАНИЧЕННЫХ РЕСУРСОВ

Аннотация

Retrieval-Augmented Generation (RAG) сформировалась как доминирующая парадигма для обоснования выходов больших языковых моделей (LLM) внешними источниками знаний, однако её систематическое применение к морфологически сложным языкам с ограниченными ресурсами остаётся недостаточно изученным. В данной работе рассматриваются две задачи для казахоязычных RAG-систем: проектирование семантически согласованных стратегий чанкинга документов и ускорение многостадийного инференса на GPU. В исследовании сформирован корпус из 201 202 предварительно обработанных новостных документов национальной газеты Egemen Qazaqstan (2017–2025, 129 тематических категорий). Предложены три стратегии чанкинга: Fixed-size (базовая), Semantic (определение границ по косинусной близости), Adaptive (объединённая семантическая и длино-регуляризованная оценка). Оценка выполнена по четырёхмерной метрике: когерентность, покрытие документа, точность границ, регулярность длины. Стратегия Adaptive достигла ChunkScore = 0.879, превзойдя Fixed-size на 0.045, а также повысила Retrieval F1@5 на 0.067 и end-to-end точность на 4 процентных пункта (92%) на тесте из 50 фактологических запросов. Для ускорения инференса использован точный GPU-индекс (FAISS-GpuIndexFlatIP) на NVIDIA H200, обеспечивший задержку поиска 2.4 мс и ускорение 70.8× по сравнению с ChromaDB без потери качества. В сочетании с INT4-квантованной генерацией, KV-cache в HBM3 и механизмом attention-based self-evaluation система достигает среднего времени ответа 0.92 ± 0.06 с в холодном кэше и ~0.20 с при полном повторном использовании KV-cache (ускорение 4.3×). Система работает на одном GPU без внешних API. Результаты демонстрируют возможность построения высокоточных субсекундных казахоязычных RAG-систем и формируют воспроизводимую базу для дальнейших исследований в области retrieval для low-resource языков.

Ключевые слова: retrieval-augmented generation; чанкинг документов; low-resource NLP; казахский язык; dense vector retrieval; FAISS; KV-cache оптимизация; ускорение инференса LLM; multilingual embeddings; GPU-ускоренный NLP.

Introduction

Large language models are fluent but unreliable when it comes to recalling facts outside their training data [1]. Retrieval-Augmented Generation (RAG) addresses this by separating memory from generation: a retriever pulls relevant passages from an external corpus, and the generator is conditioned on them, which lowers the rate of hallucination and lets the system stay current as the corpus changes [2, 3, 4]. Most of this work, however, targets English and a small set of other high-resource languages. How RAG should be adapted to morphologically rich, low-resource languages

is still largely open, and Kazakh is a clear case: tokenization, embedding, and segmentation each behave differently than they do in the languages these pipelines were originally built for.

Kazakh is a Turkic language and the state language of Kazakhstan, spoken by roughly 22 million people. Its morphology is agglutinative – case, number, person, and derivational affixes attach to a single stem, so one lexeme can surface in hundreds of well-formed variants [5]. These properties create friction at several points in a RAG pipeline. A tokenizer tuned for fusional languages tends to over-segment Kazakh words, and an embedding model trained on predominantly English data represents them poorly. Chunking is affected too: heuristics calibrated on English paragraph structure do not track how Kazakh journalistic prose actually shifts between topics. At the same time, the practical deployment of RAG systems requires inference latencies acceptable for interactive applications. On commodity CPU hardware, a 4-billion-parameter model can take 8–10 seconds per query, which is unacceptable for real-time use.

This paper targets two under-studied, tightly coupled problems for Kazakh-language RAG: (1) how documents should be split into retrieval units (chunks) so that the chunks remain semantically coherent while maintaining a length suitable for embedding and generation, and (2) how the full multi-stage RAG pipeline, including embedding, retrieval, generation, and self-evaluation, can be accelerated to operate in sub-second time on a single GPU without sacrificing accuracy.

To address these challenges, we construct a corpus of 201,202 preprocessed news articles from Egemen Qazaqstan (2017–2025, 129 thematic categories), design three greedy chunking strategies (Fixed-size, Semantic, and Adaptive), and evaluate them using a four-dimensional quality framework. Adaptive chunking performed best across the board. It reaches a ChunkScore of 0.879, a gain of 0.045 over the fixed-size baseline, and lifts retrieval F1@5 by 0.067; on our 50-query factual benchmark, end-to-end answer accuracy rises to 92%. The acceleration results are similarly favorable. An exact FAISS index on an NVIDIA H200 returns the top-K chunks in 2.4 ms — 70.8× faster than ChromaDB on the same workload. The remaining components compound this: INT4 quantization, KV-cache residency in HBM3, and attention-based self-evaluation. Together they bring cold-cache latency to 0.92 ± 0.06 s, dropping to roughly 0.20 s once the KV-cache is fully reused — all on a single GPU, with no external API in the loop.

The growing importance of retrieval quality in modern RAG systems has motivated extensive research into retrieval architectures, document segmentation methods, dense retrieval techniques, inference acceleration, and confidence estimation. Since the original introduction of RAG by Lewis et al. [1], numerous studies have proposed sequential, modular, and agent-based architectures, highlighting retrieval quality as one of the primary determinants of final answer accuracy [2, 3, 4]. Chunking therefore turns into a real design decision rather than a preprocessing detail: how large the retrieval units are, and whether they stay internally coherent, feeds directly into both retrieval precision and the quality of what the generator produces afterward.

Early RAG systems mostly used fixed-length chunks. That is easy to implement, but it cuts documents at arbitrary points and ignores where one idea ends and the next begins. To overcome this limitation, semantically aware chunking methods employ sentence embeddings and similarity thresholds to identify topical transitions [6]. More advanced approaches formulate chunk boundary selection as an optimization problem, using dynamic programming to maximize intra-chunk similarity while enforcing length constraints [7]. Other studies introduce question-aware chunking mechanisms based on multi-agent debate frameworks [8]. However, these methods have been developed and evaluated almost exclusively for English and a small number of other high-resource languages. Little is known about their effectiveness in agglutinative low-resource settings. The present work contributes to this area by providing the first systematic comparison of semantic and adaptive chunking strategies for Kazakh and by introducing a composite $ChunkScore$ metric that jointly measures coherence, coverage, boundary accuracy, and length regularity. The challenges of retrieval in Kazakh are further amplified by the limited availability of language-specific NLP resources. Although recent years have seen progress in neural machine translation, text correction, and morphological analysis for Kazakh [5], retrieval-oriented datasets and evaluation benchmarks

remain scarce. Existing tools such as Stanza provide sentence segmentation and lemmatization capabilities [9], while multilingual embedding models including multilingual-E5 and M3-Embedding have extended support to Turkic languages [10, 11]. Nevertheless, their retrieval performance on realistic Kazakh-language corpora has not been rigorously assessed. This study addresses that gap by evaluating a Kazakh-adapted embedding model on a large-scale news collection and measuring its influence on retrieval and generation quality.

Beyond retrieval quality, efficient similarity search is essential for practical deployment. Dense retrieval systems represent queries and documents as continuous vectors and typically rely on maximum inner-product search, for which FAISS offers highly optimized GPU implementations [12, 13]. While FAISS has become a standard component of many modern RAG pipelines, its effectiveness within a complete low-resource language setting has received limited attention. Furthermore, comparisons between GPU-based exact search and CPU-based approximate nearest-neighbor systems such as ChromaDB are largely absent for Kazakh-language applications. Our experiments demonstrate that GpuIndexFlatIP provides a 70.8× speedup without measurable accuracy loss, while the reduced chunk count produced by Adaptive chunking further improves retrieval throughput.

Inference efficiency constitutes another major bottleneck in interactive RAG systems. Several recent studies have shown that weight-only quantization methods, such as INT4 $q4_{K_M}$, can significantly reduce memory consumption and computational cost [14]. Similarly, maintaining the KV-cache in GPU high-bandwidth memory eliminates repeated computation of shared prefixes and may yield speedups exceeding fourfold in repeated-query scenarios [15]. Additional research explores sparse attention mechanisms [16] and hierarchical cache-management strategies [17] to further reduce latency. Building on these advances, the present study investigates the combined effect of GPU-resident KV caching, INT4 quantization, and attention-based self-evaluation within a unified Kazakh-language RAG pipeline, resulting in sub-second end-to-end response times.

The reliability of generated responses remains equally important for real-world deployment. Confidence estimation and hallucination detection mechanisms provide a means to assess whether generated content is sufficiently supported by retrieved evidence. Attention-based aggregation approaches such as AggTruth [18] exploit the attention distribution of the final decoder layer to compute context-attribution scores and identify potentially unsupported outputs. Responses falling below a predefined threshold can then be regenerated or flagged for verification. Following this line of work, we incorporate a lightweight attention-based self-evaluation module that introduces only approximately 90 ms of additional latency while improving the trustworthiness of generated answers.

Research methodology

This study addresses two interconnected optimization problems related to the construction of an efficient Kazakh-language Retrieval-Augmented Generation (RAG) system. The first concerns the identification of an optimal document chunking strategy that produces retrieval units that are simultaneously semantically coherent and appropriately sized for embedding and generation. The second focuses on minimizing end-to-end inference latency while preserving retrieval and answer quality. To investigate these objectives, we design a complete pipeline consisting of corpus construction and preprocessing, chunk generation, chunk quality evaluation, vector indexing and retrieval, and accelerated inference with self-evaluation.

Let $D = \{d^1, d^2, \dots, d^{\mathrm{N}}\}$ denote a corpus of documents, where each document d is represented as a finite sequence of tokens drawn from a vocabulary V. For a document d, let $S(d) = \{s^1, s^2, \dots, s^{\mathrm{n}}\}$ denote its ordered sentence decomposition. A pre-trained embedding model $\phi: V* \to \mathbb{R}^{\mathrm{m}}$ maps each text sequence to a unit-normalized dense vector of dimension $m = 768$. The semantic similarity between two textual units u and v is computed as

$$\sigma(u, v) = \phi(u)^{\mathrm{T}}\phi(v), \quad (1)$$

which corresponds to cosine similarity under $\ell_2 - normalization$. A chunk $c = \langle s_a, s_a{}^{+1}, \dots, s_\beta \rangle$ is defined as a contiguous sequence of sentences within a document, while a chunking function C partitions each document into a finite ordered set of non-overlapping chunks.

To evaluate chunking quality, four complementary criteria are considered. The first is semantic coherence, measured as the average similarity between consecutive sentences within a chunk. The second is coverage, which quantifies the semantic fidelity of a chunk relative to the original document. The third criterion evaluates boundary accuracy by assessing whether chunk boundaries coincide with meaningful topic transitions. Finally, a length score rewards chunks whose token counts fall within an empirically determined optimal interval. These components are combined into a single objective function:

$$Q(c,d) = w^1 Cov(c,d) + w^2 Coh(c) + w^3 Bnd(c) + w^4 Len\big(l(c)\big), \tag{2}$$

where the weights are set to $(0.35, 0.30, 0.20, 0.15)$, respectively. The objective is to identify a chunking function that maximizes the average $ChunkScore$ over the entire corpus. Because the resulting optimization problem is computationally intractable in its exact form, practical solutions are obtained through greedy left-to-right approximation algorithms.

The four components of Eq. (2) are defined as follows, where $\mu(\cdot)$ denotes the $\ell_2 - normalized$ mean (centroid) embedding of a text span and σ the cosine similarity of Eq. (1):

$$Coh(c) = \frac{1}{b-a}\sum_{k=a}^{b-1} \sigma\,(s_k, s_{k+1}) \tag{2a}$$

$$\mathrm{Cov}(c,d) = \sigma\big(\mu(c), \mu(d)\big), \qquad \mu(x) = \frac{\sum_{s\in x}\phi(s)}{\|\sum_{s\in x}\phi(s)\|_2} \tag{2b}$$

$$Bnd(c) = 1 - \sigma(s_b, s_{b+1}) \tag{2c}$$

$$Len\big(l(c)\big) = 1 - L_{penalty}\big(l(c)\big) \tag{2d}$$

The second optimization problem concerns the overall RAG inference process. Given a query q, the retrieval stage identifies the K most relevant chunks by maximizing semantic similarity between the query embedding and chunk embeddings. Retrieved chunks may subsequently be filtered according to a relevance threshold before being passed to the generative model. The generated answer is then evaluated using an attention-based confidence estimator:

$$\Gamma\big(r, \tilde{R}\big) = \left(\frac{1}{|r|}\right) \Sigma\, \alpha_j{}^{\mathrm{ctx}}, \tag{3}$$

where $\alpha_j{}^{\mathrm{ctx}}$ is the final-layer attention mass that the generated answer places on the retrieved context tokens (the set R), averaged over the $|r|$ answer tokens: for each answer token it sums the attention weights from that token to every retrieved context position, and the mean is taken over the answer length $|r|$. The superscript "ctx" is a label (context), not an index; A with superscript L denotes the final-layer attention matrix defined in Table 1. Responses whose confidence score falls below $\theta conf = 0.75$ are regenerated with increased sampling temperature. Consequently, the total response latency can be decomposed as

$$Ttotal(q) = Temb(q) + Tret(q, I) + Tgen\big(q, \tilde{R}\big) + Teval\big(r, \tilde{R}\big). \tag{4}$$

The optimization objective is therefore to determine the implementation configuration that maximizes answer quality while satisfying predefined latency and memory constraints.

To support the experiments, a large-scale corpus was constructed from archival publications of the newspaper Egemen Qazaqstan covering the years 2017-2025. The original collection contained 213,497 JSON-formatted documents with structured metadata fields including title, author, publication date, category, abstract, and full article text. A multi-stage preprocessing pipeline removed duplicate entries through normalized text hashing, filtered documents shorter than 200 characters, and excluded technical pages such as redirects and error stubs. After preprocessing, the final corpus contained 201,202 unique documents. Linguistic normalization was performed using the Stanza toolkit for Kazakh, including sentence segmentation, lemmatization, and morphological analysis. The resulting dataset spans 129 thematic categories, including politics, economics, culture, sports, science, and technology, thereby providing substantial lexical and topical diversity.

Document segmentation constitutes a critical component of the retrieval pipeline. Given a sequence of sentences extracted using Stanza, three chunking strategies are investigated. Token counts are computed using the multilingual-e5-base tokenizer to ensure compatibility with the embedding model. The first approach employs a fixed-size strategy in which sentences are sequentially accumulated until a target length of 512 tokens is reached. Chunks are constrained to remain within the interval $[384, 640]$ tokens, and a new chunk is initiated whenever the addition of another sentence would exceed the upper limit. This approach ignores semantic relationships between adjacent sentences and therefore serves as a baseline.

The second approach introduces semantic awareness by computing sentence embeddings using a Kazakh-optimized embedding model. Chunk boundaries are inserted whenever the cosine similarity between consecutive sentence embeddings falls below a threshold of $\theta = 0.65$:

$$\cos\big(E(s_i), E(s_i + 1)\big) < \theta. \tag{5}$$

The threshold value was determined empirically using a validation set containing 1,000 documents.

The third approach combines semantic information with explicit length regularization. For each potential boundary position, a boundary score is calculated as

$$BoundaryScore(i) = \alpha\big(1 - \cos\big(E(s_i), E(s_i + 1)\big)\big) + \beta Lpenalty(l), \tag{6}$$

where $\alpha = 0.6$ and $\beta = 0.4$. Here l denotes the token length of the currently open chunk, i.e. the sentences accumulated since the last inserted boundary up to and including the current sentence s_i. The length penalty increases when chunk sizes deviate from the preferred range of $384 - 640$ tokens:

$$Lpenalty(l) = 0, if\ l \in [384{,}640], \tag{7}$$

boundary score exceeds 0.7 or when a hard limit of 1024 tokens is reached. The procedure operates greedily from left to right across the document. The resulting chunks are evaluated using the four quality dimensions defined previously: coherence, coverage, boundary accuracy, and length score. Their weighted combination forms the integrated $ChunkScore$ metric, which serves as the primary criterion for comparing chunking strategies. All metrics are computed offline on a representative sample of 50 documents, and the aggregated results are presented in the experimental section. Following chunk generation, each chunk is transformed into a dense vector representation using the multilingual-e5-base embedding model. The resulting ℓ_2-normalized vectors have dimensionality 768, and similarity between a query and a chunk is computed via inner product. To enable efficient retrieval, the embeddings are indexed using FAISS and deployed on an NVIDIA H200 GPU through the exact $GpuIndexFlatIP$ implementation. Approximately 200,000 chunk vectors reside entirely in GPU memory, allowing the system to retrieve the top five most similar chunks for each query. To isolate the effect of GPU acceleration, a CPU-based baseline is implemented using ChromaDB with identical embeddings and HNSW indexing.

The final stage focuses on inference acceleration. Retrieval is performed using FAISS-GPU, which executes exact similarity search in approximately 2.4 ms per query while preserving full retrieval accuracy. Additional gains are achieved through GPU-resident Key–Value cache retention in HBM3 memory, eliminating repeated computation of shared prefixes and substantially reducing generation latency. Retrieved chunks are further processed through relevance filtering and near-duplicate removal before being passed to the language model.

Response generation is performed by a 4-billion-parameter language model quantized to four-bit precision using the $q4_{K_M}$ scheme and executed through the llama.cpp CUDA backend. The model conditions its output on the three highest-ranked chunks after filtering, using a generation temperature of 0.2 and a maximum output length of 256 tokens.

To improve reliability, the system incorporates an attention-based self-evaluation stage inspired by recent confidence-estimation approaches. Attention weights from the final decoder layer are aggregated to estimate the extent to which generated content is grounded in the retrieved evidence. When the confidence score falls below 0.75, the response is regenerated using a higher sampling temperature of 0.8. This lightweight hallucination-detection mechanism introduces approximately 90 ms of additional latency while avoiding the computational overhead of external natural language inference models.

Results of the study

All experiments were conducted on a single-node workstation equipped with an NVIDIA H200 GPU (80 GB HBM3). The experimental design follows a full factorial structure across two factors: chunking strategy (Fixed-size, Semantic, Adaptive) and retrieval index (FAISS-GPU and ChromaDB-CPU), resulting in six configurations and a total of 300 query–configuration evaluations, with 10 repeated executions per query per configuration.

The evaluation dataset consists of 50 Kazakh-language factual queries spanning ten thematic categories, including geography, population statistics, national symbols, independence history, administrative divisions, international organizations, economy, culture, sports, and science. Each query is paired with up to 10 manually annotated relevant chunks. The annotation process was performed independently by two native Kazakh speakers, and disagreements were resolved through consensus. For timing measurements, GPU execution time was recorded using CUDA events, while CPU timing relied on $time.perf_counter$, with explicit CUDA stream synchronization applied prior to all measurements to ensure consistency.

For clarity and reproducibility, Table 1 summarizes the notation used throughout the study, covering dataset structure, chunking functions, embedding space, similarity metrics, retrieval depth, thresholds, and system-level constraints such as GPU memory budget and token length settings.

Table 1. Summary of principal notation.

Symbol	*Definition*
D	*Document corpus*
d, N	*Document; corpus size (N = 201,202)*
$S(d)$	*Sentence decomposition of document d*
c	*Chunk (contiguous sentence subsequence)*
$l(c)$	*Token length of chunk c*
$\phi: V* \rightarrow R^{\mathrm{m}}$	*Text embedding function (m = 768)*
$\sigma(u, v)$	*Cosine similarity (Eq. 1)*
C	*Chunking function (see Methodology)*
I	*Retrieval index (full chunk set)*
$Q(c, d)$	*Integrated chunk quality score (Eq. 2)*
K	*Retrieval depth (K = 5)*
τ^{f}	*Relevance filter threshold (τ^{f} = 0.20)*

θ	*Semantic boundary threshold (θ = 0.65)*
θ_{conf}	*Confidence threshold for self-evaluation (0.75)*
A^{L}	*Final-layer attention matrix*
$\Gamma(r, \tilde{\tilde{R}})$	*Context attribution confidence score (Eq. 3)*
T_{total}	*End-to-end pipeline latency (Eq. 4)*
M_{max}	*GPU memory budget (80 GB HBM3)*
$T*, T_{min}, T_{max}$	*Chunk token lengths (512, 384, 640)*

Moving to chunk-level behavior, quality metrics were computed by averaging over approximately 1,000 chunks per strategy derived from 50 documents. Adaptive comes out ahead on most of these measures, with a ChunkScore of 0.879 – 0.045 above Fixed-size and 0.019 above Semantic. Two components drive the gap: boundary alignment (0.85 versus 0.76 for Fixed-size) and length regularization (0.91 against 0.77). Coverage moves the other way, dropping by 0.007 under Adaptive, which is the expected cost of packing chunks more tightly around a single topic at the expense of how much of the document each one represents. The drop is small enough that retrieval downstream does not suffer. Coherence barely moves across the three strategies – 0.81 for Fixed-size, 0.82 for Semantic, 0.83 for Adaptive.

In terms of retrieval quality, performance across all configurations is summarized in Figure 1. A consistent pattern emerges across both indexing systems, where FAISS-GPU and ChromaDB produce nearly identical results, with differences in F1@5 not exceeding 0.003. This indicates that retrieval effectiveness is governed primarily by embedding quality and chunk structure rather than index implementation. Here too Adaptive chunking gives the best retrieval, with an F1@5 of 0.785 – about 0.065 over Fixed-size and 0.061 over Semantic. This lines up with what we expected going in: when chunks are coherent and kept to a regular length, their embeddings are cleaner, and retrieval gets more accurate as a result.

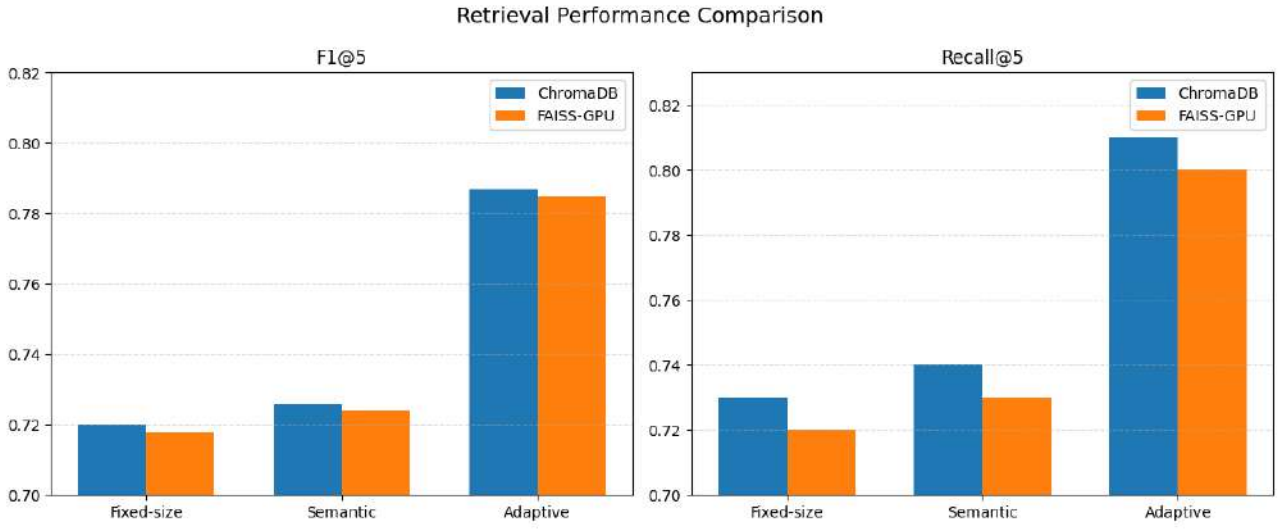


Figure 1. Retrieval performance comparison across chunking strategies and indexing systems (F1@5 and Recall@5).

Latency behavior further highlights the differences between retrieval backends. As shown in Table 2, FAISS-GPU achieves a retrieval latency of only 2.4 ms, which corresponds to a 70.8× speedup compared to ChromaDB running on CPU (0.170 s). However, retrieval is not the dominant contributor to total pipeline cost. Instead, autoregressive generation remains the primary bottleneck, taking approximately 0.810 s for 256 tokens. Generation runs on the same INT4-quantized GPU backend in both configurations, so its latency is identical across the two index conditions; the FAISS-GPU and ChromaDB rows therefore differ only in the retrieval stage. Self-evaluation based on

attention signals introduces an additional 0.090 s, but remains stable and does not significantly affect overall throughput.

Table 2. Latency decomposition (Adaptive + FAISS-GPU on NVIDIA H200; ChromaDB on CPU). All values in seconds. Component latencies are measured independently per stage; the reported Total is the directly measured end-to-end latency and may differ marginally from the component sum (here ≈13 ms) owing to host–device transfer and scheduling overhead not attributable to any single stage.

Stage	*FAISS-GPU (H200)*	*ChromaDB (CPU)*
Query embedding (s)	*0.013 ± 0.002*	*0.013 ± 0.002*
Retrieval (s)	*0.0024 ± 0.0003*	*0.170 ± 0.012*
Generation – TTFT (s)	*0.018 ± 0.003*	*0.018 ± 0.003*
Generation – 256 tokens (s)	*0.810 ± 0.050*	*0.810 ± 0.050*
Self-evaluation (s)	*0.090 ± 0.010*	*0.090 ± 0.010*
Total (s)	*0.920 ± 0.060*	*1.100 ± 0.075*

When all components are considered jointly, end-to-end results across the six configurations are presented in Table 3. A clear monotonic relationship emerges between ChunkScore and final answer accuracy: Fixed-size (0.834, 88%) is followed by Semantic (0.860, 90%), and then Adaptive (0.879, 92%). This indicates that chunk quality is a strong predictor of downstream system performance. Across all strategies, FAISS-GPU consistently reduces total latency by approximately 0.31 seconds compared to ChromaDB, without any observable degradation in accuracy.

Table 3. End-to-end comparison across all six evaluated configurations.

Strategy	*Index*	*Time (s)*	*Acc.*	*F1@5*
Fixed-size	*ChromaDB*	*1.20*	*88%*	*0.719*
Fixed-size	*FAISS-GPU*	*0.89*	*88%*	*0.719*
Semantic	*ChromaDB*	*1.21*	*90%*	*0.725*
Semantic	*FAISS-GPU*	*0.90*	*90%*	*0.725*
Adaptive	*ChromaDB*	*1.23*	*92%*	*0.786*
Adaptive	*FAISS-GPU*	*0.92*	*92%*	*0.786*

Another important aspect of system behavior is KV cache efficiency, particularly under repeated-query conditions. As shown in Table 4, the generation stage requires 0.810 s under cold-cache conditions – consistent with the generation latency reported in Table 2 – and is reduced to 0.190 s under partial cache hits and to 0.100 s under full cache reuse, corresponding to generation-stage speedups of 4.26× and 8.10×, respectively. Translated to the complete pipeline, full cache reuse lowers the mean end-to-end response time from 0.92 s in the cold-cache setting to approximately 0.20 s, an overall ≈4.3× speedup that closely matches theoretical expectations for KV cache optimization.

Table 4. KV cache impact on generation latency (generation stage only; Adaptive + FAISS-GPU).

Condition	*Time (s)*	*Speedup*
Cold cache	*0.810 ± 0.050*	*1.00×*
Partial hit	*0.190 ± 0.020*	*4.26×*
Full hit	*0.100 ± 0.010*	*8.10×*

Finally, when compared against external baselines, the proposed system demonstrates a strong overall advantage. As shown in Figure 2, it outperforms BM25 (Elasticsearch) by a large margin in both accuracy and F1@5, while also surpassing a GPT-3.5 API-based dense retrieval system. Notably, the proposed approach achieves 92% accuracy with a latency of only 0.92 s, compared to 2.40 s for the API-based system, while maintaining full locality without external dependencies.

Although BM25 offers relatively low latency (0.65 s), its performance is significantly lower in terms of retrieval quality and accuracy, particularly in handling morphologically rich Kazakh text.

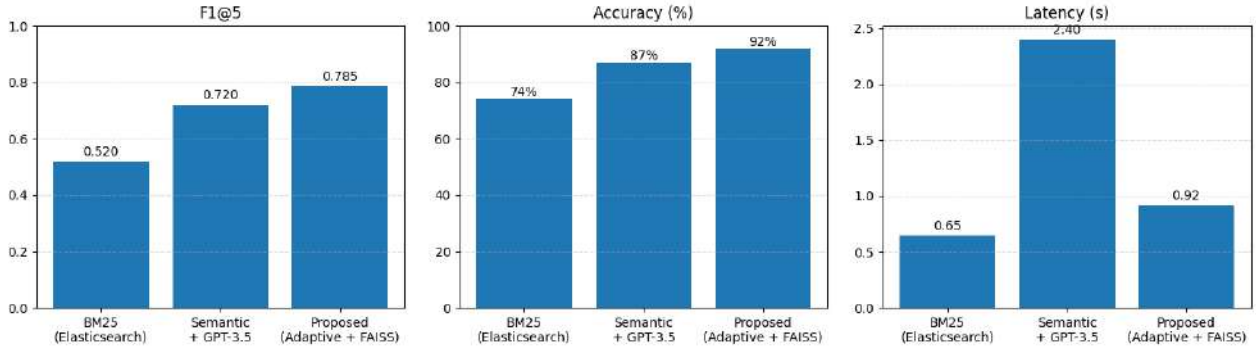


Figure 2. Comparison with baseline systems across retrieval quality and latency.

Discussion

Taken together, these results highlight several consistent patterns. Adaptive chunking consistently improves retrieval quality and downstream accuracy due to stronger boundary detection and better length regularization. FAISS-GPU significantly reduces retrieval latency by approximately 70.8× without any measurable impact on accuracy, confirming that retrieval speedups do not come at the cost of quality at this scale (~200k chunks). INT4 quantization enables sub-second generation on a single GPU while maintaining competitive output quality. KV cache reuse introduces substantial efficiency gains, reaching approximately 4.3× end-to-end speedup with full cache reuse, which is particularly beneficial in interactive or session-based settings. Among all components, retrieval quality has the strongest influence on downstream accuracy, while index choice primarily affects latency rather than correctness.

The chunking results are consistent with research on semantic boundary selection [6] and length-aware chunk optimization [7], while extending this evidence to morphologically rich Kazakh news text. In contrast to question-aware strategies such as QChunker [8], the Adaptive approach evaluated here uses a greedy, query-independent segmentation rule. Its small reduction in document coverage (0.007 relative to Fixed-size) indicates a trade-off between local semantic coherence and retention of broader article context, which may become more consequential for questions requiring evidence from several passages.

The acceleration results help distinguish retrieval efficiency from end-to-end generation efficiency. FAISS-GPU and ChromaDB produced nearly identical retrieval scores, whereas the exact GPU index reduced search latency; this agrees with the role of optimized similarity search established in FAISS research [12, 13]. Since autoregressive generation remained the largest measured latency component, a 70.8× retrieval-stage acceleration does not imply an equivalent end-to-end improvement. The approximately 0.20 s response time under full KV-cache reuse is likewise specific to repeated-prefix conditions and should not be generalized to cold-cache requests. These findings complement work on quantization [14] and cache management [15, 17].

At the same time, several limitations remain. The evaluation is restricted to 50 manually verified single-hop factual queries and does not cover multi-hop reasoning, conversational memory, or dynamic corpora updates. The attention-based confidence estimation remains heuristic and is not fully calibrated against human uncertainty judgments. Additionally, the system is evaluated on a static corpus with a fixed embedding model, leaving further improvements through contrastive fine-tuning on hard negatives as a promising direction for future work.

Attention-based self-evaluation adds approximately 90 ms and offers a practical way to assess evidence attribution, as explored in related attention-based work [18]. However, the present experiment does not establish whether the confidence scores are calibrated to independently judged factual support. Further study should use larger query samples with uncertainty estimates, multi-hop

and time-sensitive questions, dynamic news corpora, and separate assessments of cold-cache and repeated-query workloads. Tests on related Turkic languages and contrastive fine-tuning with hard negatives would help establish how widely the reported improvements generalize.

Conclusion

This work presented a comprehensive pipeline for Kazakh-language Retrieval-Augmented Generation, addressing two critical bottlenecks: the design of semantically coherent document chunks and the acceleration of multi-stage inference on a single GPU. We built a corpus of 201,202 preprocessed news articles from Egemen Qazaqstan and used it to develop an Adaptive chunking strategy. Rather than optimizing a single objective, it balances four at once: intra-chunk coherence, document coverage, boundary accuracy, and length regularity. Measured against this four-part quality framework, Adaptive chunking scored 0.879 on ChunkScore, 0.045 above the fixed-size baseline. The gains carried through to retrieval and generation: F1@5 rose by a statistically significant 0.067, and end-to-end answer accuracy reached 92% on a 50-query factual benchmark.

On the systems side, we demonstrated that an exact GPU-resident FAISS index ($GpuIndexFlatIP$) provides a 70.8× retrieval speedup over ChromaDB without accuracy loss, and that INT4 quantisation, KV-cache residency in HBM3, and attention-based self-evaluation together deliver a mean end-to-end response time of 0.92 ± 0.06 s (cold cache) and ≈ 0.20 s with full KV cache reuse (4.3× speedup) on a single NVIDIA H200 GPU without external API dependencies. These results establish the feasibility of sub-second, fully local Kazakh RAG systems suitable for interactive applications.

The contributions also include a formal mathematical problem statement for chunking quality and inference latency optimization, a reproducible evaluation protocol, and the release of the largest documented Kazakh-language corpus for information retrieval research (201,202 documents, 129 categories).

Future work will extend the evaluation to larger and dynamically updated corpora, explore cross-lingual transfer for other Turkic languages (e.g., Uzbek, Kyrgyz), and investigate learned chunking policies as well as advanced RAG architectures with multi-hop reasoning capabilities. Improving the underlying Kazakh embedding model through contrastive fine-tuning on hard negatives remains a priority for raising the performance ceiling of the entire pipeline.

Acknowledgement and Conflict of Interests

This research was funded by the Science Committee of the Ministry of Science and Higher Education of the Republic of Kazakhstan, grant number BR24993001 “Creation of a large language model (LLM) to maintain the implementation of the Kazakh language and increase the technological progress”. The authors declare no conflicts of interest.

The contribution of the authors

Conceptualization: Assel Ospan, Talshyn Sarsembayeva and Imre Felde; Methodology: Assel Ospan and Aisha Sailau; Software: Aisha Sailau; Validation: Assel Ospan and Talshyn Sarsembayeva; Formal analysis: Assel Ospan and Aisha Sailau; Investigation: all authors; Resources: Assel Ospan; Data curation: Aisha Sailau and Aidana Temirkhanova; Writing – original draft preparation: Assel Ospan and Talshyn Sarsembayeva; Writing – review and editing: all authors; Visualization: Aisha Sailau. All authors reviewed and approved the final version of the manuscript.

References

[1] Lewis P. et al. Retrieval-augmented generation for knowledge-intensive nlp tasks //Advances in neural information processing systems. – 2020. – T. 33. – C. 9459-9474.

[2] Gao Y. et al. Retrieval-augmented generation for large language models: A survey //arXiv preprint arXiv:2312.10997. – 2023.

[3] Huang Y., Huang J. X. A survey on retrieval-augmented text generation for large language models //ACM Computing Surveys. – 2026. – Т. 58. – №. 12. – С. 1-38.

[4] Fan W. et al. A survey on rag meeting llms: Towards retrieval-augmented large language models //Proceedings of the 30th ACM SIGKDD conference on knowledge discovery and data mining. – 2024. – С. 6491-6501.

[5] Baitenova L. et al. Hybrid artificial intelligence architectures for automatic text correction in the Kazakh language //Frontiers in Artificial Intelligence. – 2025. – Т. 8. – С. 1708566.

[6] Allamraju A. et al. Breaking It Down: Domain-Aware Semantic Segmentation for Retrieval Augmented Generation //arXiv preprint arXiv:2512.00367. – 2025.

[7] Kiss C., Nagy M., Szilágyi P. Max–Min semantic chunking of documents for RAG application //Discover Computing. – 2025. – Т. 28. – №. 1. – С. 117.

[8] Zhao J. et al. QChunker: Learning question-aware text chunking for domain RAG via multi-agent debate //Proceedings of the ACM Web Conference 2026. – 2026. – С. 2207-2218.

[9] Qi P. et al. Stanza: A Python natural language processing toolkit for many human languages //Proceedings of the 58th annual meeting of the association for computational linguistics: system demonstrations. – 2020. – С. 101-108.

[10] Wang L. et al. Improving text embeddings with large language models //Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). – 2024. – С. 11897-11916.

[11] Chen J. et al. M3-embedding: Multi-linguality, multi-functionality, multi-granularity text embeddings through self-knowledge distillation //Findings of the association for computational linguistics: ACL 2024. – 2024. – С. 2318-2335.

[12] Johnson J., Douze M., Jégou H. Billion-scale similarity search with GPUs //IEEE transactions on big data. – 2019. – Т. 7. – №. 3. – С. 535-547.

[13] Douze M. et al. The faiss library //IEEE Transactions on Big Data. – 2025.

[14] Lin Y. et al. Qserve: W4a8kv4 quantization and system co-design for efficient llm serving //Proceedings of Machine Learning and Systems. – 2025. – Т. 7.

[15] Li H. et al. A survey on large language model acceleration based on kv cache management //arXiv preprint arXiv:2412.19442. – 2024.

[16] Zhu Y. et al. Accelerating inference of retrieval-augmented generation via sparse context selection //International Conference on Learning Representations. – 2025. – Т. 2025. – С. 25965-25981.

[17] Ge D. et al. HA-RAG: Hotness-Aware RAG Acceleration via Mixed Precision and Data Placement //arXiv preprint arXiv:2510.20878. – 2025.

[18] Matys P. et al. AggTruth: Contextual Hallucination Detection using Aggregated Attention Scores in LLMs //International Conference on Computational Science. – Cham : Springer Nature Switzerland, 2025. – С. 227-243.

IRSTI 28.23.37

https://doi.org/10.51889/2959-5894.2026.95.3.022

M. Sakypbekova[1], R. Aubakirova[1*]

[1]Al-Farabi Kazakh National University, Almaty, Kazakhstan
*e-mail: aubakirova.ramilya@gmail.com

CORPUS AND BASELINE MODELS FOR KAZAKH TOXIC COMMENT DETECTION

Abstract

Public discourse in Kazakhstan is moving online and into the Kazakh language, yet resources for detecting toxic comments in that language remain scarce. The Kazakh corpora that exist were assembled from lists of hostile keywords within an extremism or model-safety frame, few of them are openly released, and fewer still document how they were annotated. This paper presents Kaz-ToCo, a corpus of 84,940 comments in Kazakh and in mixed Kazakh-Russian code-switched text. The comments were retrieved in May 2026 through the YouTube Data API v3 from beneath 30 Kazakhstani talk-show videos, with no keyword-based filtering at any stage of collection, and were labelled on a binary toxicity scale by an open-weight large language model deployed on local hardware. Agreement between the machine labels and the annotations of a native speaker of Kazakh, measured on a stratified sample of 100 comments, reached a Cohen kappa coefficient of 0.602, with a 95 percent confidence interval from 0.446 to 0.758. The training, validation and test splits were built at the level of the source video, which prevents comments from one video from appearing in more than one split. Two baseline classifiers are established on the corpus: term frequency and inverse document frequency features over word and character n-grams combined with logistic regression, and a fine-tuned Kazakh conversational encoder of the RoBERTa architecture. Macro-averaged F1 on the held-out test set reached 0.8086 and 0.8198 respectively. These figures fall well below the near-ceiling accuracies published on keyword-collected Kazakh corpora, and are offered as a realistic reference point. The labels, per-comment metadata, splitting protocol, random seeds and hyperparameters are released without restriction. The comment texts cannot be redistributed under the YouTube API Services Developer Policies and are available from the corresponding author upon reasonable request for non-commercial research use.

Keywords: Kazakh language; toxic comment detection; low-resource language processing; corpus construction; language model annotation; code-switching; text classification; encoder fine-tuning.

М. Сакыпбекова[1], Р. Аубакирова[1*]

[1]әл-Фараби атындағы Қазақ ұлттық университеті, Алматы, Қазақстан

ҚАЗАҚ ТІЛІНДЕГІ УЫТТЫ ПІКІРЛЕРДІ АНЫҚТАУДЫҢ КОРПУСЫ МЕН БАЗАЛЫҚ МОДЕЛЬДЕРІ

Аңдатпа

Қазақстандағы жария коммуникация барған сайын интернетке ауысып, қазақ тіліне көшуде, алайда осы тілдегі уытты пікірлерді автоматты түрде анықтауға арналған ресурстар әлі де аз. Қолда бар қазақ тіліндегі корпустар экстремистік әрі деструктивті мазмұнды іздеп, түйінді сөздер тізімі арқылы жиналған, олардың аз бөлігі ғана ашық жарияланған, ал белгілеу рәсімі одан да сирек құжатталған. Мақалада Kaz-ToCo, яғни қазақ тілінде және аралас қазақ-орыс кодында жазылған 84 940 пікірден тұратын корпус ұсынылады. Пікірлер 2026 жылдың мамыр айында YouTube Data API v3 интерфейсі арқылы 30 қазақстандық ток-шоу бейнежазбаларының пікірлер бөлімінен, түйінді сөздер бойынша сүзгіден өткізілместен жиналды. Оларды екілік уыттылық шкаласымен жергілікті жабдыққа орнатылған ашық салмақты тілдік модель белгіледі. Стратификацияланған 100 пікірден тұратын іріктемеде машиналық белгілеу қазақ тілін ана тілі ретінде білетін маманның бағасымен үйлесті. Коэн каппа коэффициенті 0,602 құрады, 95 пайыздық сенімділік аралығы 0,446-дан 0,758-ге дейін. Оқыту, валидация және тест іріктемелері бейне деңгейінде құрылды, сондықтан бір бейненің пікірлері әртүрлі іріктемеге түспейді. Корпус үшін екі базалық классификатор құрылды: сөздік және таңбалық n-граммаларға негізделген терминдер жиілігі мен құжаттардың кері жиілігі белгілері логистикалық регрессиямен және қосымша оқытылған қазақ тіліндегі RoBERTa сәулетіндегі энкодер. Кейінге қалдырылған тест іріктемесіндегі макроорташаланған F1 өлшемі сәйкесінше 0,8086 және 0,8198 болды.

Алынған мәндер түйінді сөздер бойынша жиналған корпустарда жарияланатын шекке жуық көрсеткіштерден айтарлықтай төмен және оптимистік емес, шынайы бастапқы нүкте ретінде ұсынылады. Белгілер, пікір деңгейіндегі метадеректер, іріктемелерге бөлу хаттамасы, кездейсоқ сандар генераторының бастапқы мәндері мен гиперпараметрлер шектеусіз жарияланады. Пікірлердің мәтіндерін YouTube API Services Developer Policies ережелеріне сәйкес қайта таратуға болмайды, олар коммерциялық емес ғылыми мақсатта хат-хабар авторынан негізді сұраныс бойынша алынады.

Түйін сөздер: қазақ тілі; уытты пікірлерді анықтау; ресурсы аз тілдерді өңдеу; корпус құру; тілдік модельмен белгілеу; код ауыстыру; мәтіндерді жіктеу; энкодерді қосымша оқыту.

М. Сакыпбекова[1], Р. Аубакирова[1*]
[1]Казахский национальный университет имени аль-Фараби, Алматы, Казахстан

КОРПУС И БАЗОВЫЕ МОДЕЛИ ВЫЯВЛЕНИЯ ТОКСИЧНЫХ КАЗАХОЯЗЫЧНЫХ КОММЕНТАРИЕВ

Аннотация

Публичная коммуникация в Казахстане всё активнее уходит в интернет и переходит на казахский язык, однако ресурсов для автоматического выявления токсичных комментариев на этом языке по-прежнему мало. Имеющиеся казахоязычные корпуса собраны по спискам враждебных ключевых слов, в логике поиска экстремистского и деструктивного контента, лишь немногие из них открыты, а процедура разметки документирована ещё реже. В статье представлен Kaz-ToCo, корпус из 84 940 комментариев на казахском языке и на смешанном казахско-русском коде. Комментарии получены в мае 2026 года через интерфейс YouTube Data API v3 под 30 казахстанскими ток-шоу, без фильтрации по ключевым словам. Разметку по бинарной шкале токсичности выполнила языковая модель с открытыми весами, развёрнутая на локальном оборудовании. На стратифицированной выборке из 100 комментариев машинная разметка согласуется с разметкой носителя казахского языка на уровне 0,602 по коэффициенту каппа Коэна, при 95-процентном доверительном интервале от 0,446 до 0,758. Обучающая, валидационная и тестовая выборки построены на уровне исходного видео, поэтому комментарии одного видео не попадают в разные выборки. Для корпуса построены два базовых классификатора: признаки на основе частоты терминов и обратной частоты документов по словным и символьным n-граммам с логистической регрессией, а также дообученный казахоязычный энкодер архитектуры RoBERTa. Макроусреднённая мера F1 на отложенной тестовой выборке достигла 0,8086 и 0,8198 соответственно. Полученные значения заметно ниже околопредельных показателей, публикуемых на корпусах, собранных по ключевым словам, и предлагаются как реалистичная, а не оптимистичная точка отсчёта. Метки, покомментарийные метаданные, протокол разбиения на выборки, начальные значения генератора случайных чисел и гиперпараметры публикуются без ограничений. Сами тексты комментариев повторно распространять нельзя по правилам YouTube API Services Developer Policies, их предоставляет автор для корреспонденции по обоснованному запросу для некоммерческого научного использования.

Ключевые слова: казахский язык; выявление токсичных комментариев; обработка малоресурсных языков; построение корпуса; разметка языковой моделью; переключение кодов; классификация текстов; дообучение энкодера.

Introduction

Kazakh-language research on harmful content has been dominated by keyword-driven collection within an extremism or model-safety frame, which has left the field without a general-purpose corpus of naturally occurring toxic comments. Kaz-ToCo fills that gap with 84,940 comments retrieved exhaustively through the YouTube Data API v3 from beneath 30 talk-show videos, with no content-based selection at the collection stage, and covering both monolingual Kazakh and Kazakh-Russian code-switching. Labelling by a locally deployed open-weight language model agreed with a native speaker at a Cohen kappa of 0.602, a moderate value that matches the cautious picture reported for toxic content in low-resource settings rather than the optimistic one reported for English-language social-science coding. Under a video-level group split, a linear model over word and character n-grams reaches a macro-averaged F1 of 0.8086 and a fine-tuned Kazakh encoder reaches 0.8198, a gap of 0.0112 that no significance test supports. The labels, the per-comment metadata, the splitting

protocol and the full training configuration are released without restriction. The comment texts, which the YouTube API Services Developer Policies do not permit to be redistributed, are available from the corresponding author upon reasonable request for non-commercial research use.

A nationwide survey of 21,000 households carried out in 2025 found that 93.4% of Kazakhstan's population aged six and older uses the internet, and that 69.5% of those users named participation in social networks among their online activities [1]. Public communication is at the same time shifting toward Kazakh. The government Commission for the Implementation of State Language Policy reported that as of June 2026 the share of Kazakh-language content in the media had reached 78%, and that roughly 60% of internet resources now operate in the state language [2]. What aggressive speech does in that space is already on record. The international Health Behaviour in School-aged Children survey found that 13.4% of Kazakhstani adolescents aged 11 to 15 had been subjected to cyberbullying [3]. A large share of the country's public conversation now happens online and in Kazakh, and official statistics already register the harm.

Automatic toxicity detection depends on labelled data in the target language, and labelled data is exactly what Kazakh lacks. Research on harmful content in Kazakh dates to the mid-2010s, but it settled early into an information-security frame. The target phenomenon gets operationalized as extremism, destructive content or cyberbullying. Corpora are assembled from lists of hostile keywords. Annotation procedures often go undocumented, and open releases remain the exception [4, 5, 6].

Three limitations follow from that framing. First, corpus construction in nearly every Kazakh-language study begins with a keyword list and automatic platform scraping [5, 6, 7]. The design produces lexical bias and topic leakage. Accuracies occasionally reported near the ceiling, up to 98% for religious-extremism detection [5], are more plausibly explained by that leakage than by the task actually being solved, a reading consistent with the broader finding that offensive-speech classifiers degrade sharply once evaluation conditions become realistic [8]. Second, annotation protocols are sparsely documented. One clear exception exists: in [4], five trained native speakers worked from written guidelines and reached a Cohen kappa of 0.82. Most of the remaining Kazakh toxicity work reports neither annotator instructions nor any agreement figure. Third, open releases are rare. Three are available. The first is a corpus of roughly three thousand religious extremist texts pulled from a social network by keyword search [9]. The second is the 10,190-text destructive-content dataset published as supplemental material with [4]. The third is a 17,827-sample profanity dataset built to align language models, which pairs every toxic input with a de-escalating model response and so serves response generation rather than comment-level classification [10]. All three were keyword-collected or safety-framed. None contains naturally occurring platform comments gathered without content-based selection, and none implements the graded, multi-facet taxonomy covering profanity, insult, threat and identity attack, nor the severity gradation whose feasibility has been demonstrated for low-resource languages [11]. The open-access practice that KazSAnDRA established for Kazakh sentiment analysis [12] has no counterpart in toxicity.

A second and methodologically stronger strand has grown up around the safety evaluation of large language models in Kazakh. The Qorgau benchmark assessed twelve models in Kazakh and Russian across more than ninety thousand responses and, in a step that matters here, validated its automated judge against human annotations of a thousand examples per language before scaling up, recording high agreement between machine and human in both languages [13]. KZ-SafetyPrompts extends the same logic to prompt collection: 5,717 natively Kazakh prompts across eleven risk categories, with a documented compilation protocol, decision rules for edge cases and deduplication procedures [14]. That a safety benchmark, rather than a toxicity corpus, offers one of the most detailed published accounts of a harmful-content annotation procedure for Kazakh says a good deal about the state of the field.

Both strands converge on a design question any new Kazakh toxicity resource has to answer: who, or what, does the labelling. Toxicity annotation is subjective. An annotator's beliefs and identity systematically shift the labels produced, and low agreement often reflects a plurality of defensible

readings rather than error [15]. Against that background, model-assisted annotation has emerged as a way to cut cost and annotator burden. The headline finding is that a zero-shot commercial chat model outperformed trained crowdworkers on both accuracy and inter-coder agreement, at roughly one-thirtieth of the cost [16]. Evidence specific to toxic content is more guarded: the same model reaches around 80% agreement with human reference labels on non-harmful content and identifies toxicity reliably at a coarse level, yet it struggles to separate hateful, offensive and toxic subcategories, and the outcome is sensitive to prompt wording [17]. Whether any of this survives the move to a low-resource language is unsettled. A direct evaluation for Marathi found a substantial quality drop and concluded that model output cannot stand in for human annotation [18], while a weak-supervision pipeline for Kazakh sequence labelling produced strong taggers from machine annotations that were confidence-filtered and backed by a small, manually verified subset [19]. One further caveat applies to the whole quantitative base: it rests on proprietary models reached through a vendor interface. Whether the findings carry over to open-weight models running on local hardware has not been tested separately.

This study was conducted to close the first of those gaps and to put evidence under the second and the third. Its aim is to build a general-purpose corpus of Kazakh-language toxic comments collected without keyword-driven selection, to document every step of the procedure, to release everything that platform policy permits to be released, and to establish reproducible baselines on the result. Three questions were put to the data. Can an open-weight language model running on local hardware label Kazakh toxicity at a level of agreement with a native speaker that makes the labels usable? Does a corpus collected without keyword filtering and evaluated under a video-level group split reproduce the near-ceiling accuracies reported elsewhere in the Kazakh literature, or noticeably lower ones? And does a Kazakh-specific pretrained encoder beat a sparse n-gram baseline by a margin that matters in practice?

Research methodology

Data source and collection.

The corpus is named Kaz-ToCo after Bangla-ToCo, a corpus of toxic comments in Bengali [20]. Its source material is Kazakhstani talk shows hosted on the YouTube video platform. Talk shows were chosen because the comments beneath them capture spontaneous, unmoderated reactions to whatever the panel was arguing about, rather than text elicited by a search for harmful content. This is the point at which the present collection design parts company with the keyword-built Kazakh corpora reviewed above.

Comments were retrieved in May 2026 through the YouTube Data API v3. The commentThreads.list method was called for each video with the videoId parameter, textFormat set to plain text and maxResults set to the permitted maximum of 100, and the nextPageToken field was followed until the API returned no further page. Top-level comments and their replies were both taken from the returned threads. Thirty videos were processed. Collection was exhaustive at the video level: every comment thread the API returned for a video was retained, with no content-based filtering at the collection stage, so the corpus's size and topical composition were set by the discussions themselves rather than by a per-class or per-topic quota. Raw collection yielded 152,244 comments.

Normalization and filtering.

Comment text was normalized without altering lexical content. Line breaks and non-breaking spaces (U+00A0) were replaced with ordinary spaces. Invisible characters were removed: the zero-width space (U+200B), the word joiner (U+2060) and the byte order mark (U+FEFF). The zero-width joiner (U+200D) was kept, since removing it would break composite emoji. Runs of spaces were collapsed, and leading and trailing spaces were stripped.

Platform artefacts were not deleted but replaced with placeholder tokens, which preserves positional information: URLs became [URL], user mentions became [USER], and timecodes became [TIMECODE].

Filtering ahead of annotation proceeded in three steps. The collection was deduplicated. Language identification was then applied with the fastText classifier, and comments identified as purely Russian were dropped, while monolingual Kazakh comments and Kazakh-Russian code-switched comments were kept. Finally, entries containing no letters or digits at all, comments consisting only of emoji or punctuation for instance, were removed. These three steps reduced the collection from 152,244 raw comments to the 84,940 that make up the final corpus.

Retaining code-switched material was deliberate. Mixing Kazakh and Russian is characteristic of Kazakhstani platform discourse, and until now only the Qorgau safety benchmark had explicitly addressed a bilingual setting [13]. No dedicated resource existed for toxic code-switched comments.

Data labeling.

Comments were labelled into two classes, toxic and non-toxic, by the open-weight model cyankiwi/Qwen3.6-35B-A3B-AWQ-4bit, served through the vLLM inference engine, version 0.21.0, and deployed locally on an ASUS DGX Spark server, whose configuration is detailed under Training procedure below. Local deployment of an open-weight model was chosen for three reasons. User comments are never sent to a third-party service. The weight version and the runtime configuration are fixed and remain available for re-running the process. And annotation cost does not scale with corpus size. Published quantitative assessments of harmful-content annotation, however, come from proprietary models [13, 16, 17], so their conclusions cannot simply be carried over to the model used here. Agreement with human annotation was therefore measured separately.

A comment was assigned to the toxic class (label 1) when it contained profanity, insults, aggression, incitement to hatred or degradation. The neutral class (label 0) covered neutral or positive text, questions and constructive criticism. Judgement was made on the meaning of the Kazakh-language utterance, taking context into account. These criteria summarize the system prompt, which was written in Russian, the working language of the bilingual environment the corpus comes from. In English translation the prompt reads:

You are an expert content moderator specializing in the Kazakh language. Classify user comments for toxicity. Rules: profanity, insults, aggression, hatred, degradation give label 1; neutral or positive text, a question, constructive criticism give label 0; judge the meaning in the Kazakh language, taking context into account; respond only with a JSON object containing a single field label, whose value is 0 or 1.

Each request carried exactly one comment. The model saw nothing else: no video title, no like counts or other metadata, no parent comment and no other comments from the thread. The word context in the instruction therefore refers to the internal context of the utterance itself. Comments whose text was empty after normalization were assigned label 0 without a model call.

Generation ran at a temperature of 0.0, capped at 64 output tokens, in structured JSON output mode. The model's reasoning mode was disabled. Response validity and typing were checked against a Pydantic schema, and requests were retried on any format violation.

Validation of the machine labels.

A hundred comments were selected at random under stratification by machine label, preserving the corpus's class proportions, and were labelled independently by a native speaker of Kazakh. Agreement between the two label sources was measured with the Cohen kappa coefficient, reported with an asymptotic 95% confidence interval and checked against a nonparametric bootstrap over 10,000 resamples. The procedure follows the validate-first-then-scale architecture used in Kazakh safety evaluation [13] and in weak supervision for Kazakh sequence labelling [19].

Dataset organization.

The corpus is a single comma-separated file with the fields video_id, text, like_count, is_reply, is_pinned, is_favorited and label. The author_id field, returned by the API at the collection stage, was excluded for privacy reasons.

The corpus is distributed in two parts, for a reason that is worth stating plainly. The YouTube API Services Developer Policies do not permit the redistribution of comment texts retrieved through the API, so the text column cannot be posted in a public repository. Everything else can be, and is: the

labels, the per-comment metadata, the video-level split assignment, the annotation and training code, the random seeds and the hyperparameters are released without restriction, keyed to the same row order, so that a reader who obtains the texts can rebuild the corpus exactly. The comment texts are available from the corresponding author upon reasonable request for non-commercial research use.

Splitting protocol.

Splitting was carried out at the video level, grouped by video_id, so that comments from one video cannot land in more than one split. Comments under a single video share a topic, participants and conversational context, so a random split at the level of the individual comment would let a model lean on video-specific cues that are unavailable under deployment conditions.

To keep class balance comparable across splits, the toxic share was computed for each video, videos were binarized at the median into low-toxicity and high-toxicity strata, and a stratified split was applied at the video level. Splitting proceeded in two stages, 85/15 and then 82.4/17.6 within the larger portion, which yields roughly a 70/15/15 split by video count at a random seed of 42. The result is 20 videos in the training set and five each in validation and test.

Classification models.

Two models were evaluated. The first pairs sparse lexical features with a linear classifier. The representation concatenates two matrices of term frequency and inverse document frequency weights. The word-level matrix uses n-grams of order 1 to 2, a vocabulary capped at 100,000 features, a minimum document frequency of 2, sublinear term-frequency scaling and lowercasing. The character-level matrix uses character n-grams of order 2 to 5 computed within word boundaries, a cap of 200,000 features, a minimum document frequency of 2 and sublinear scaling. Character features were included because of Kazakh's rich morphology: a single lexeme surfaces as many word forms, and word-level features fragment across them. The resulting representation has 287,946 dimensions, 100,000 of them word-level and 187,946 character-level. The classifier is logistic regression with the liblinear solver, balanced class weights, a cap of 2,000 iterations and a fixed random seed of 42. Regularization strength was tuned by grid search over the values 0.1, 0.5, 1.0, 3.0 and 10.0 for the inverse regularization parameter, selected by macro-averaged F1 on the validation set. The value 1.0 was chosen.

The second model is the Kazakh conversational encoder kz-transformers/kaz-roberta-conversational, a model of the RoBERTa architecture of roughly 334 MB, fitted with a freshly initialized classification head over two labels. Dropout was raised to 0.25 for both the hidden and the attention layers. Input texts were tokenized with the model's own tokenizer at a maximum length of 128 tokens, padded to that length and truncated beyond it.

Training procedure.

Fine-tuning followed a two-stage freeze-then-unfreeze schedule. In the first stage the encoder was frozen for one epoch while only the classification head trained, at a learning rate of 1e-4 with no warmup. In the second stage all parameters were unfrozen and training continued for up to three epochs at a learning rate of 1e-5, with a linear scheduler and warmup over 10% of the steps.

Optimization used the AdamW algorithm with separate weight decay, 0.01 for weight matrices and 0 for biases and layer-normalization parameters. Batch size was 16 with gradient accumulation over two steps, giving an effective batch size of 32. Training ran in mixed precision with gradient clipping at a global norm of 1.0. The loss was cross-entropy with label smoothing of 0.1. Model selection used macro-averaged F1 on the validation set, with a checkpoint saved after every epoch and early stopping at a patience of two. The random seed was fixed at 42 across the random, numpy and torch generators, including the CUDA generator.

For reproducibility, the software and hardware environment is stated here. Annotation, training and inference all ran on the same machine: an ASUS DGX Spark server built on the NVIDIA GB10 Grace Blackwell superchip, which pairs a 20-core ARM64 (aarch64) CPU with an integrated Blackwell-generation GPU over 128 GB of unified memory, under Ubuntu 24.04 LTS. Data processing and the sparse baseline used numpy 2.4.6, pandas 3.0.5, scipy 1.17.1, scikit-learn 1.9.0 and joblib 1.5.3; encoder fine-tuning used torch 2.13.0 and transformers 5.16.1; the annotation model

was served through the vLLM inference engine, version 0.21.0. The exact package list is published together with the training code.

Evaluation metrics.

Models were evaluated on the held-out test set with four metrics: macro-averaged F1, micro-averaged F1, weighted F1, and the area under the receiver operating characteristic curve. Macro-averaged F1 is the primary metric and drives model selection, because it averages F1 across classes without weighting by class size and so treats the toxic and the non-toxic class equally. Micro-averaged and weighted F1 are reported for comparability with prior work that uses them, and the area under the curve captures ranking quality independently of the decision threshold. No significance test was run on the difference between the two models, a point the discussion returns to.

Results

Characteristics of the corpus.

Table 1 summarizes the corpus. Kaz-ToCo contains 84,940 comments from 30 videos in two classes, with a vocabulary of 136,977 distinct tokens. Comments are short: mean length is 14.38 words, the median is 10, and the range runs from 1 to 96 words. Engagement metadata is heavily skewed, with a mean like count of 5.17 against a median of 0. Replies, rather than top-level comments, account for 17.56% of the corpus.

Table 1. Overall statistics of the Kaz-ToCo corpus

Metric	*Value*
Total comments	*84,940*
Unique videos	*30*
Number of classes	*2*
Vocabulary size	*136,977*
Mean words per comment	*14.38*
Median words per comment	*10.0*
Minimum words per comment	*1.0*
Maximum words per comment	*96.0*
Mean like count	*5.17*
Median like count	*0.0*
Reply share, %	*17.56*

The class breakdown in Table 2 shows a corpus close to balanced: 47,614 non-toxic comments, or 56.06%, against 37,326 toxic ones, or 43.94%. The classes differ systematically along every dimension recorded. Toxic comments run longer, 16.08 words on average against 12.45 for non-toxic ones, and 12 against 9 by median. They draw fewer likes, 4.18 against 5.94 on average. And they are less often replies: the reply share among toxic comments is 14.37% against 20.06% for non-toxic ones.

Table 2. Statistics by class, where label 0 denotes the non-toxic class and label 1 the toxic class

Label	*Count*	*Share, %*	*Mean words*	*Median words*	*Mean likes*	*Reply share, %*
0	*47,614*	*56.06*	*12.45*	*9.0*	*5.94*	*20.06*
1	*37,326*	*43.94*	*16.08*	*12.0*	*4.18*	*14.37*

The class difference in length shows up across the whole distribution, not just in the tail. Figure 1 shows that non-toxic comments cluster in a sharp mode around 50 characters, while the toxic distribution has a lower, flatter mode, is shifted to the right, and carries more mass beyond 150 characters.

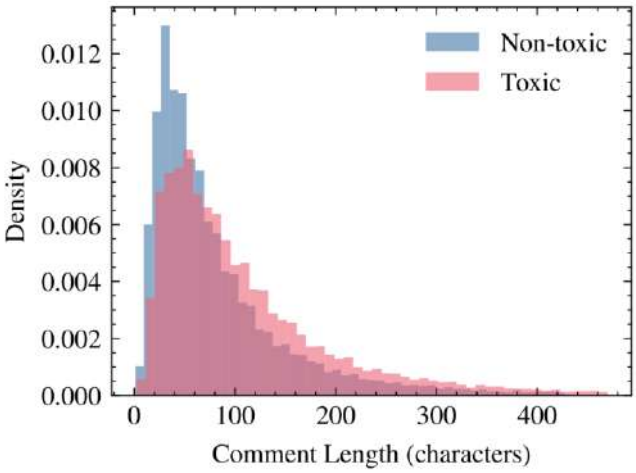


Figure 1. Density distribution of comment length in characters for the non-toxic and the toxic class

Composition of the splits.

Because videos vary in comment count, the per-comment proportions do not mirror the video-level split: the training set holds 62,152 comments, validation 7,780 and test 15,008. The class distribution across the three splits is shown in Figure 2. The toxic share is 43.9% in training, 44.3% in validation and 43.9% in test. Stratifying by each video's internal toxicity share keeps the balance nearly constant despite the grouping constraint, and none of the three splits is heavily imbalanced.

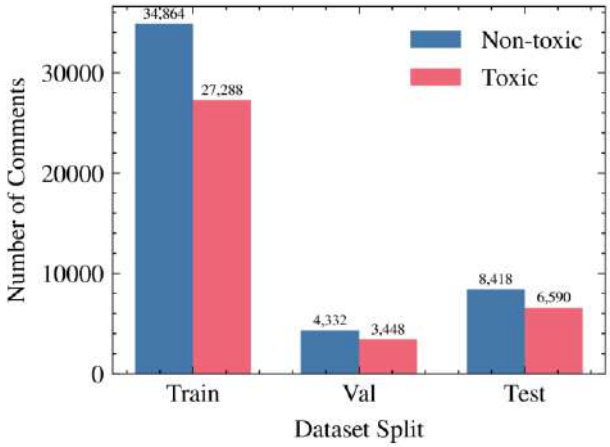


Figure 2. Class distribution across the training, validation and test splits, with counts labelled above the bars

Comment-length distributions were compared across the three splits to confirm that the video-level split had not introduced a systematic difference in text length. The distributions in Figure 3 match closely in mode, spread and right tail.

Agreement between machine and human labels.

On the stratified sample of 100 comments, 44 of which carried a machine toxic label, agreement between the model's labels and the native speaker's reached a Cohen kappa of 0.602, with a 95% confidence interval of 0.446 to 0.758 and an asymptotic standard error of 0.080. The nonparametric bootstrap over 10,000 resamples gives a near-identical interval of 0.443 to 0.755.

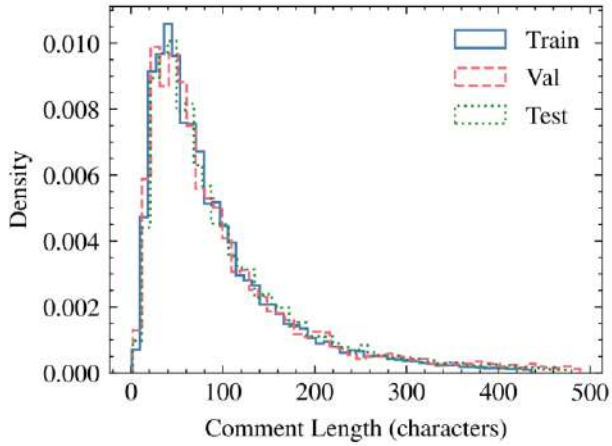


Figure 3. Density distribution of comment length in characters for the training, validation and test splits

Table 3 breaks the same 100 comments down as a contingency table. The two label sources agree on 80 comments. The 20 disagreements are asymmetric: 14 comments the human marked toxic went unflagged by the model, against 6 flagged in the opposite direction. On this sample the model assigns the toxic label less often than the native speaker does, 44 times against 52. The direction of that asymmetry bears on how the corpus's class balance should be read. A toxic share of 43.94% is high next to toxicity corpora for other languages, but the disagreement pattern indicates that the figure is not produced by the model over-assigning the toxic class. If anything, the human reference would have put it higher. The more plausible sources of the high share are the material itself, since talk-show discussions of contested social topics draw heated reactions, and the breadth of the toxic definition in the prompt, which folds aggression and degradation in with profanity and insults.

Table 3. Contingency table of human and model labels on the 100-comment validation sample

Human label	*Model: non-toxic*	*Model: toxic*	*Total*
Human: non-toxic	*42*	*6*	*48*
Human: toxic	*14*	*38*	*52*
Total	*56*	*44*	*100*

Classification quality.

Table 4 reports both baseline models on the test set. The fine-tuned Kazakh encoder scores higher on all four metrics. For both models the three F1 variants diverge by no more than 0.004, which follows from the near-uniform class distribution shown in Table 2 and indicates that neither model has collapsed into predicting the majority class.

Table 4. Performance of the two baseline models on the held-out test set

Model	*Macro-F1*	*Micro-F1*	*Weighted F1*	*ROC-AUC*
kaz-roberta-conversational	*0.8198*	*0.8233*	*0.8228*	*0.9063*
TF-IDF with logistic regression	*0.8086*	*0.8111*	*0.8113*	*0.8982*

Composition of the errors.

Tables 5 and 6 present confusion matrices over the 15,008 test comments. Relative to the linear model, the transformer produces 290 fewer false positives at the cost of 107 additional missed toxic

comments. Its edge on the aggregate metrics therefore comes from a more cautious assignment to the toxic class rather than from uniformly better separation.

Table 5. Confusion matrix on the test set for term frequency and inverse document frequency features with logistic regression

Actual class	*Predicted non-toxic*	*Predicted toxic*
Actual non-toxic	*6,939*	*1,479*
Actual toxic	*1,356*	*5,234*

Table 6. Confusion matrix on the test set for the fine-tuned kaz-roberta-conversational encoder

Actual class	*Predicted non-toxic*	*Predicted toxic*
Actual non-toxic	*7,229*	*1,189*
Actual toxic	*1,463*	*5,127*

Discussion

The gap between the two baselines.

The sparse representation with a linear classifier trails the pretrained Kazakh transformer, fine-tuned through a multi-stage schedule, by 0.0112 macro-averaged F1. No significance test was run, so this difference can be called neither established nor noise. What can be said is that it is descriptively small, and that it matches the result found for destructive-content classification in Kazakh, where n-gram features with classical classifiers achieved high accuracy while a multilingual encoder combined with a recurrent layer proved competitive but not decisively better [4], a picture corroborated by comparisons from the same group [5]. Two independently built Kazakh-language corpora converging on the same qualitative outcome is worth noting, because they differ exactly along the dimension that should have mattered: the corpus in [4] was assembled by keyword search within an extremism frame, whereas Kaz-ToCo was pulled from entire videos with no content filtering at the collection stage.

Relation to the gaps identified in the literature.

The corpus closes the gap it was built to close. Kaz-ToCo is a general-purpose Kazakh-language toxicity resource, and it differs from the Kazakh datasets that precede it [4, 9, 10] in collection design and in scale rather than in availability. Those resources were gathered by keyword search or built for model-safety alignment, and the largest comment-level one holds 10,190 texts against 84,940 here. On availability this corpus is in one respect more constrained than they are, and the constraint is worth being explicit about: retrieval through the YouTube Data API v3 brings the comment texts under the platform's developer policies, which forbid redistributing them, so the texts are shared on request while the labels, the metadata, the splits and the code are posted openly. The open-access practice set by KazSAnDRA in sentiment analysis [12] can therefore be followed only in part. At the same time the corpus inherits a limitation from the same literature: the taxonomy is binary. The graded, multi-facet scheme covering profanity, insult, threat and identity attack, along with the severity gradation whose feasibility has been shown for low-resource languages [11], is not implemented here.

On annotation transparency, this work reports the name and the value of the agreement metric and documents the normalization, filtering and selection steps in full. It does not yet meet the standard set by KZ-SafetyPrompts, whose documented compilation protocol and edge-case rules remain among the most detailed published accounts of a harmful-content annotation procedure for Kazakh [14].

On model-assisted annotation of Kazakh toxicity specifically, this work provides the first direct evidence, and that evidence is ambiguous in exactly the way the prior literature would predict. The Cohen kappa obtained, 0.602, sits noticeably below the human-machine agreement recorded for coarse Kazakh safety judgements by the validated judge in Qorgau [13], and it does not replicate the

crowdworker-beating result reported for zero-shot annotation in English-language social-science coding [16]. It sits closer to the more cautious picture found specifically for toxic content, where coarse detection is reliable but finer distinctions degrade and the outcome is sensitive to prompt wording [17], and it fits the failure mode documented for a low-resource setting in Marathi [18]. The reading that transfer is task-dependent, and that coarse binary safety judgements survive low-resource settings better than culturally loaded distinctions, accommodates this result. A single kappa on a single corpus does not settle the dispute between the competing positions. The comparison is imprecise in a further respect: the cited figures come from proprietary models, while the annotation here used an open-weight model, and the contribution of that difference cannot be isolated.

Two caveats apply to the kappa value itself. Toxicity annotation is inherently subjective, and disagreement often reflects a plurality of defensible readings rather than error [15], so a moderate coefficient does not by itself indict the machine labels. At the same time, no calibration study exists for Kazakh-Russian bilingual annotators against which the human reference itself could be checked.

As for how realistic the evaluation is, the video-level group split removes the most obvious source of inflation in this setup. The figures obtained here, macro-averaged F1 around 0.82, fall well short of the near-ceiling accuracies reported on keyword-collected Kazakh corpora, up to 98% [5]. That is exactly what one would expect if those figures partly reflect lexical bias and topic leakage rather than genuine task difficulty [8]. The comparison spans different corpora built under different annotation schemes, so it is suggestive rather than probative.

Two gaps flagged at the outset remain open. The corpus includes Kazakh-Russian code-switched comments, which at the level of the data addresses the absence of a dedicated bilingual resource beyond the safety benchmark in [13]. But no analysis separating quality on the monolingual subset from quality on the code-switched subset was carried out, so no claim is made here about how the models handle code-switching. Nor is the encoder-choice question settled, that is, whether Kazakh-specific pretraining beats multilingual transfer on the toxicity task, since only the Kazakh-language encoder was fine-tuned, with no multilingual counterpart trained under comparable conditions.

Limitations.

The domain is narrow. All 84,940 comments come from 30 talk-show videos on a single platform. Talk-show discourse covers a particular range of topics and participants, and transfer to news comments, to other platforms, or to other Kazakh-language communities was not tested.

Labels are predominantly machine-generated. Only 100 comments carry a human label, and these were used for validation rather than as a gold-standard training subset in the style of the weak-supervision scheme in [19]. One consequence deserves to be stated outright: since the test set carries the same machine labels as the training set, the classification metrics reported above measure agreement with the annotator model's labels, not with human judgement, and a classifier that reproduced that model's systematic errors would be rewarded rather than penalized for it. A sample of 100 also keeps the confidence interval around the kappa wide. The interval from 0.446 to 0.758 spans values conventionally read as moderate agreement at the lower end and substantial at the upper. Human validation rests on a single annotator, so no human-to-human agreement estimate exists, and the reported kappa measures agreement between the model and one person.

Annotation used a single model and a single prompt formulation. The sensitivity of toxicity labels to prompt wording is documented [17] but was not measured here: neither alternative prompts nor a second annotator model for cross-checking was used. Evaluation is limited to aggregate metrics on a single test set. Breakdowns by video, by comment length and by the presence of code-switching, along with robustness checks against masked or obfuscated lexical content, were not carried out, so nothing can be said about model behaviour beyond the averaged figures. Finally, the redistribution restriction imposed by the platform's developer policies puts a step between a reader and the data. Obtaining the comment texts requires a request to the corresponding author rather than a download, which slows independent verification, and comments deleted by their authors or removed by the platform after May 2026 cannot be recovered from the API by anyone attempting to rebuild the collection from scratch.

Prospects for further research.

The nearest task is to refine the binary label into toxicity subtypes with a severity gradation [11], which would bring Kaz-ToCo in line with taxonomies adopted for other languages at the cost of substantially harder annotation. Next comes expanding the human-annotated subset to a size where it can serve as a gold-standard training set and running a controlled comparison of Kazakh-specific and multilingual encoders under comparable conditions. A practical extension of this work is a browser extension that flags toxic comments in real time.

Conclusion

The Kazakh-language toxicity literature has not, until now, offered a general-purpose corpus of naturally occurring comments. The resources that exist were collected via keyword lists within an extremism or model-safety frame, documented unevenly, and only rarely released. This paper presents Kaz-ToCo, a corpus of 84,940 comments in Kazakh and in mixed Kazakh-Russian code-switched text, retrieved through the YouTube Data API v3 from beneath 30 Kazakhstani talk-show videos in May 2026 and labelled on a binary toxicity scale. The corpus was collected without relying on harmful keywords, normalized and filtered under a documented protocol, labelled by a locally deployed open-weight model, and checked against native-speaker annotation on a stratified sample, where the Cohen kappa came to 0.602. Splitting was performed at the video level, stratified by each video's internal share of toxic comments. Within this protocol two baselines were established. The fine-tuned kaz-roberta-conversational encoder reached a macro-averaged F1 of 0.8198 and an area under the receiver operating characteristic curve of 0.9063. The model over word and character n-grams with logistic regression reached 0.8086 and 0.8982. The gap of 0.0112 in macro-averaged F1 is small, and without a significance test it cannot be taken as establishing the transformer's advantage. It consists of 290 eliminated false positives bought at the cost of 107 additional missed detections.

The practical contribution is a resource several times larger than the few Kazakh datasets released so far, and the first collected without keyword-driven selection. The labels, the per-comment metadata, the splitting protocol, the random seeds and the hyperparameters needed to reproduce the figures reported here are published without restriction, while the comment texts, which the platform's developer policies do not permit to be redistributed, are available from the corresponding author upon reasonable request for non-commercial research use. Those figures fall noticeably short of the near-ceiling results published on keyword-collected corpora and are offered as a realistic rather than an optimistic baseline. The main limitations are the binary taxonomy, the predominantly machine-generated labels, and a human validation step carried out by a single annotator. Expanding the taxonomy, growing the human-annotated subset, comparing Kazakh-specific and multilingual encoders under comparable conditions, and testing robustness to masked text are the natural next steps. Information about the publication material. The labels, the per-comment metadata, the video-level split assignment, the annotation and training code, the random seeds and the hyperparameters that support the findings of this study are openly available at https://github.com/cestramilya/kaz-toco. Comment texts cannot be redistributed under the YouTube API Services Developer Policies and are available from the corresponding author upon reasonable request for non-commercial research use.

References

[1] On the use of information and communication technologies by households in the Republic of Kazakhstan, 2025 [Electronic resource] / Bureau of National Statistics of the Agency for Strategic Planning and Reforms of the Republic of Kazakhstan. – Astana, 2026. – URL: https://stat.gov.kz/en/industries/business-statistics/stat-it/ (date of access: 27.08.2026).

[2] Share of Kazakh-language content in media reaches 78% [Electronic resource] // KazTAG. – 2026. – URL: https://kaztag.kz/en/news/share-of-kazakh-language-content-in-media-reaches-78 (date of access: 27.08.2026).

[3] Key findings from the Health Behaviour in School-aged Children survey in the Republic of Kazakhstan, HBSC 2022: Executive summary [Electronic resource] / National Center for Public Health of the Ministry of Health of the Republic of Kazakhstan. – Almaty, 2023. – URL: https://hls.kz (date of access: 27.08.2026).

[4] Bolatbek M., Sagynay M., Mussiraliyeva S., Yeltay Z. Detection of offensive content in the Kazakh language using machine learning and deep learning approaches // PeerJ Computer Science. – 2025. – Vol. 11. – Art. e3027. – https://doi.org/10.7717/peerj-cs.3027

[5] Bolatbek M., Sagynay M., Mussiraliyeva S., Zhastay Y. Machine learning methods for identifying destructive texts in the Kazakh language // Proceedings of the 2024 3rd International Conference on Algorithms, Data Mining, and Information Technology. – New York: ACM, 2024. – P. 99–103. – https://doi.org/10.1145/3701100.3701121

[6] Mussiraliyeva S.Z., Bolatbek M.A., Baispay G.B. Investigating long short-term memory approach for extremist messages detection in Kazakh language // Expert Systems. – 2025. – Vol. 42, No. 1. – Art. e13595. – https://doi.org/10.1111/exsy.13595

[7] Bolatbek M., Sagynay M., Mussiraliyeva S., Baisylbayeva K., Yeltay Z. Kazakh language dataset for hate speech detection on social media text // 2024 IEEE 9th International Conference on Computational Intelligence and Applications (ICCIA). – Piscataway: IEEE, 2024. – P. 94–98. – https://doi.org/10.1109/ICCIA62557.2024.10719327

[8] Jahan M.S., Oussalah M. A systematic review of hate speech automatic detection using natural language processing // Neurocomputing. – 2023. – Vol. 546. – Art. 126232. – https://doi.org/10.1016/j.neucom.2023.126232

[9] Mussiraliyeva S., Bolatbek M., Omarov B. Corpus of religious extremist texts [Electronic resource]. – Mendeley Data, 2020. – V1. – https://doi.org/10.17632/h272z7xv9w.1

[10] Kadyrbek N., Tuimebayev Z., Mansurova M., Viegas V. The development of small-scale language models for low-resource languages, with a focus on Kazakh and direct preference optimization // Big Data and Cognitive Computing. – 2025. – Vol. 9, No. 5. – Art. 137. – https://doi.org/10.3390/bdcc9050137

[11] Saeed R., Afzal H., Rauf S.A., Iltaf N. Detection of offensive language and its severity for low resource language // ACM Transactions on Asian and Low-Resource Language Information Processing. – 2023. – Vol. 22, No. 6. – P. 1–27. – https://doi.org/10.1145/3580476

[12] Yeshpanov R., Varol H.A. KazSAnDRA: Kazakh sentiment analysis dataset of reviews and attitudes. – arXiv preprint arXiv:2403.19335. – 2024. – https://doi.org/10.48550/arXiv.2403.19335

[13] Goloburda M., Laiyk N., Turmakhan D. et al. Qorgau: Evaluating safety in Kazakh-Russian bilingual contexts // Findings of the Association for Computational Linguistics: ACL 2025. – Stroudsburg: ACL, 2025. – P. 9765–9784. – https://doi.org/10.18653/v1/2025.findings-acl.507

[14] Zaghouani W., Ibrahim S.A., Muratbek A., Zhakenov O., Akhmetzhanova A. KZ-SafetyPrompts: A Kazakh safety evaluation prompt dataset for large language models. – arXiv preprint arXiv:2605.26947. – 2026. – https://doi.org/10.48550/arXiv.2605.26947

[15] Sap M., Swayamdipta S., Vianna L., Zhou X., Choi Y., Smith N.A. Annotators with attitudes: How annotator beliefs and identities bias toxic language detection // Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. – Stroudsburg: ACL, 2022. – P. 5884–5906. – https://doi.org/10.18653/v1/2022.naacl-main.431

[16] Gilardi F., Alizadeh M., Kubli M. ChatGPT outperforms crowd workers for text-annotation tasks // Proceedings of the National Academy of Sciences. – 2023. – Vol. 120, No. 30. – Art. e2305016120. – https://doi.org/10.1073/pnas.2305016120

[17] Li L., Fan L., Atreja S., Hemphill L. HOT ChatGPT: The promise of ChatGPT in detecting and discriminating hateful, offensive, and toxic comments on social media // ACM Transactions on the Web. – 2024. – Vol. 18, No. 2. – P. 1–36. – https://doi.org/10.1145/3643829

[18] Jadhav S., Shanbhag A., Thakurdesai A., Sinare R., Joshi R. On limitations of LLM as annotator for low resource languages. – arXiv preprint arXiv:2411.17637. – 2025. – https://doi.org/10.48550/arXiv.2411.17637

[19] Aitim A. LLM-assisted weak supervision for low-resource Kazakh sequence labeling: Synthetic annotation and CRF-refined NER/POS models // Applied Sciences. – 2026. – Vol. 16, No. 8. – Art. 3632. – https://doi.org/10.3390/app16083632

[20] Rupa S.A., Anwar M.M., Ritu N.A., Arifin K.N. Bangla-ToCo: A context-aware dataset for Bengali toxic comment detection // Data in Brief. – 2025. – Vol. 63. – Art. 112277. – https://doi.org/10.1016/j.dib.2025.112277

ИНФОРМАТИКАНЫ ОҚЫТУ ӘДІСТЕМЕСІ. БІЛІМ БЕРУДІ АҚПАРАТТАНДЫРУ

МЕТОДИКА ПРЕПОДАВАНИЯ ИНФОРМАТИКИ. ИНФОРМАТИЗАЦИЯ ОБРАЗОВАНИЯ

METHODS OF TEACHING COMPUTER SCIENCE. INFORMATIZATION OF EDUCATION

ҒТАХР 14.35.09

https://doi.org/10.51889/2959-5894.2026.95.3.023

К.Ж. Ажибеков[1], Н.К. Мадияров[1], Н.С. Утеулиев[1*], Ш.Е. Алтынбеков[1], М.Ж. Ажибеков[1]

[1]М.Әуезов атындағы Оңтүстік Қазақстан университеті, Шымкент қ., Қазақстан
[*]*e-mail:nurgali_u@mail.ru*

ЖАСАНДЫ ИНТЕЛЛЕКТІНІҢ ӨЗІНДІК ЖҰМЫСТАРДА ҚОЛДАНЫЛУЫ: ПЕДАГОГИКАЛЫҚ БАҚЫЛАУДЫҢ РӨЛІ

Аңдатпа

Мақалада студенттердің өзіндік жұмыстарын орындау барысында жасанды интеллект технологияларын қолдану ерекшеліктері және осы үдерістегі педагогикалық бақылаудың рөлі қарастырылады. Зерттеудің мақсаты – студенттердің өзіндік жұмыстарында жасанды интеллект құралдарын қолданудың тиімділігін зерттеу және педагогикалық бақылаудың оқу нәтижелеріне ықпалын анықтау. Зерттеу барысында теориялық талдау, салыстырмалы талдау, сауалнама, педагогикалық бақылау, педагогикалық эксперимент және статистикалық өңдеу әдістері қолданылды. Зерттеудің эмпирикалық базасын педагогикалық экспериментке қатысқан 182 студенттің (n = 182) деректері құрады. Мәліметтер орташа мәндерді салыстырмалы талдау, екі факторлы дисперсиялық талдау және панельдік регрессия әдістері арқылы өңделді. Зерттеу нәтижелері жасанды интеллект құралдарын педагогикалық бақылаумен ұштастыра қолдану студенттердің оқу мотивациясын, академиялық дербестігін және тапсырмаларды орындау сапасын арттыратынын көрсетті. Сонымен қатар педагог тарапынан жүйелі кері байланыс пен бақылау ұйымдастырылған жағдайда жасанды интеллектіні жауапты пайдалану және академиялық адалдық қағидаларын сақтау деңгейі жоғарылайтыны анықталды. Алынған нәтижелер жасанды интеллектінің білім сапасына оң әсері тек жүйелі педагогикалық бақылау жағдайында ғана тұрақты болатынын дәлелдеді. Бақылау жеткіліксіз болған жағдайда студенттердің дербес танымдық белсенділігі мен оқу тапсырмаларын саналы орындау деңгейінің төмендеу үрдісі байқалды. Зерттеудің ғылыми маңыздылығы жасанды интеллектіні студенттердің өзіндік жұмыстарында қолданудың тиімділігі педагогикалық бақылаудың жүйелілігіне тікелей байланысты екенін негіздеумен сипатталады.

Түйін сөздер: жасанды интеллект, жоғары білім, өзіндік жұмыс, педагогикалық бақылау, білім беру технологиялары.

К.Ж. Ажибеков[1] , Н.К. Мадияров[1], Н.С. Утеулиев[1], Ш.Е. Алтынбеков[1] М.Ж. Ажибеков[1]
[1]Южно-Казахстанский университет имени М.Ауэзова, г. Шымкент, Казахстан.

ИСПОЛЬЗОВАНИЕ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА В САМОСТОЯТЕЛЬНОЙ РАБОТЕ: РОЛЬ ПЕДАГОГИЧЕСКОГО КОНТРОЛЯ

Аннотация

В статье рассматриваются особенности использования технологий искусственного интеллекта при выполнении студентами самостоятельной работы и роль педагогического контроля в этом процессе. Цель исследования – изучение эффективности использования инструментов искусственного интеллекта в самостоятельной работе студентов и определение влияния педагогического контроля на

результаты обучения. В ходе исследования использовались методы теоретического анализа, сравнительного анализа, опроса, педагогического контроля, педагогического эксперимента и статистической обработки. Эмпирическую основу исследования сформировали данные 182 студентов (n = 182), участвовавших в педагогическом эксперименте. Данные обрабатывались с использованием методов сравнительного анализа средних значений, двухфакторного дисперсионного анализа и панельной регрессии. Результаты исследования показали, что использование инструментов искусственного интеллекта в сочетании с педагогическим контролем повышает учебную мотивацию студентов, их академическую самостоятельность и качество выполнения заданий. Кроме того, было установлено, что при организации систематической обратной связи и контроля со стороны преподавателя повышается уровень ответственного использования искусственного интеллекта и соблюдения принципов академической честности. Полученные результаты подтвердили, что положительное влияние искусственного интеллекта на качество обучения устойчиво только при условии систематического педагогического контроля. При отсутствии достаточного контроля наблюдалась тенденция к снижению уровня самостоятельной познавательной активности студентов и осознанного выполнения учебных заданий. Научная значимость исследования заключается в обосновании того, что эффективность использования искусственного интеллекта в самостоятельной работе студентов напрямую зависит от систематичности педагогического контроля.

Ключевые слова: искусственный интеллект, высшее образование, самостоятельная работа, педагогический контроль, образовательные технологии.

K.Zh. Azhibekov[1] , N.K. Madiyarov[1], N.S. Uteuliyev[1], Sh.E. Altynbekov[1], M.Zh. Azhibekov[1]
[1]M.Auezov South-Kazakhstan university, Shymkent, Kazakhstan

USE OF ARTIFICIAL INTELLIGENCE IN INDEPENDENT WORK: THE ROLE OF PEDAGOGICAL SUPERVISION

Abstract

The article examines the features of using artificial intelligence technologies in students' independent work and the role of pedagogical supervision in this process. The purpose of the study is to investigate the effectiveness of using artificial intelligence tools in students' independent work and to determine the influence of pedagogical supervision on learning outcomes. The study employed methods of theoretical analysis, comparative analysis, survey, pedagogical control, pedagogical experiment and statistical processing. The empirical basis of the study consisted of data from 182 students (n = 182) who participated in the pedagogical experiment. The data were processed using methods of comparative analysis of mean values, two-way analysis of variance, and panel regression. The results of the study showed that the use of artificial intelligence tools in combination with pedagogical control increases students' learning motivation, academic independence, and the quality of task completion. In addition, it was established that systematic feedback and control by the teacher improve the level of responsible use of artificial intelligence and adherence to the principles of academic integrity. The findings confirmed that the positive impact of artificial intelligence on the quality of learning is sustainable only under conditions of systematic pedagogical supervision. In the absence of sufficient control, a tendency toward a decrease in students' independent cognitive activity and conscious completion of learning tasks was observed. The scientific significance of the study lies in substantiating the fact that the effectiveness of using artificial intelligence in students' independent work directly depends on the systematic nature of pedagogical supervision.

Keywords: artificial intelligence, higher education, independent work, pedagogical supervision, educational technologies.

Кіріспе

Негізгі ережелер. Білім беру және педагогикалық тәжірибелер жүргізу соңғы уақытта елеулі әрі ауқымды технологиялық өзгерістердің ықпалына ұшырады, оның айқын мысалы ретінде осы салаларға жасанды интеллекттің қарқынды енуімен түсіндіруге болады. Білім беру үдерісінде жасанды интеллект құралдарының белсенді қолданылуы педагогикалық әдістерді тиімді ұсынуға арналған жаңа зерттеулер жүргізуді қажет етеді [1].

Қазақстан Республикасының Президенті Қасым-Жомарт Тоқаев атап өткендей: «Біздің міндетіміз – жаһандық үрдістерге бейімделу. Ол үшін білім беру жүйесінің цифрлық

трансформациясын жеделдету аса маңызды. Жасанды интеллектіні енгізу білім сапасын арттыруға және теңсіздікті азайтуға ықпал етеді» [2].

Жасанды интеллект технологиялары жоғары білім беру жүйесінде кеңінен қолданылып, студенттердің өзіндік жұмыстарын орындау тәсілдерін айтарлықтай өзгертуде. ChatGPT, Gemini, Copilot сияқты құралдар ақпаратты іздеу, мәтін құрастыру, есептерді шешу және бағдарламалау тапсырмаларын орындауда кеңінен пайдаланылуда. Алайда бұл құралдарды бақылаусыз пайдалану студенттердің дербес ойлау қабілетінің, сыни талдау дағдыларының және академиялық адалдық қағидаларын сақтауының төмендеуіне әсер етуі мүмкін. Сондықтан жасанды интеллектіні оқу үдерісіне енгізуде педагогикалық бақылаудың мазмұны, формалары мен тиімділігі ерекше ғылыми және практикалық маңызға ие.

Қазіргі зерттеулердің басым бөлігі жасанды интеллектінің білім беру мүмкіндіктері мен артықшылықтарын қарастырғанымен, студенттердің өзіндік жұмыстарын орындауда педагогикалық бақылаудың жасанды интеллектіні тиімді қолдануға ықпалын кешенді түрде қарастырған зерттеулер жеткіліксіз. Осы қайшылық аталған зерттеудің өзектілігін айқындайды.

Қазіргі ғылыми зерттеулер жасанды интеллекттің ақпаратты жылдам өңдеуге, мәтіндерді құрылымдауды жақсартуға және күрделі академиялық материалдарға қолжетімділікті жеңілдетуге мүмкіндік беретінін көрсетеді [3]. Сонымен қатар, оның бақылаусыз қолданылуы білімді үстірт меңгеру, сыни ойлаудың әлсіреуі және цифрлық тәуелділіктің қалыптасуы сияқты қауіптерді де туындататындығы туралы айтылады [4].

Жасанды интеллект (ЖИ) технологияларының қарқынды дамуы білім беру кеңістігінің мазмұны мен құрылымына елеулі өзгерістер енгізуде. Әсіресе білім алушылардың өзіндік жұмыстарында бұл өзгерістер айқын байқалады, себебі цифрлық құралдар бұрын толықтай адамның когнитивтік қызметіне тиесілі болған бірқатар функцияларды өзіне ала бастады [5].

Jovanović және Campbell еңбектерінде ChatGPT сияқты генеративті жасанды интеллект модельдері (GAI) - статистика мен ықтималдықтарды қолдана отырып, адам жасаған мазмұндық артефактілерді біріктіретін, бақыланбайтын немесе орташа бақыланатын машиналық оқыту жүйесі деп тұжырымдайды [6]. Ал S. Jain және т.б. ғалымдар тобы генеративті жасанды интеллект модельдерін тек инновациялық жолмен ғана емес, сонымен қатар қауіпсіз, этикалық және сенімді түрде қалай пайдалану керектігі белгісіз болып тұрғандығын атап өтеді [7]. Генеративті жасанды интеллект модельдеріндегі бұл маңызды олқылықтар әсіресе білім беру саласында анықталуы мүмкін көптеген қиындықтардың толық шешілмегендігін көрсетеді.

Алайда ChatGPT бүкіл әлемнің назарын аударғанымен, адамдар арасындағы өзара әрекеттесудің азаюы сияқты қазіргі таңдағы туындап отырған көптеген қиындықтардың пайда болуына алып келуде [8].

Генеративті жасанды интеллект модельдері мен ChatGPT сияқты қолданбалардың тілдік модельдермен салыстыруға болатындай адамдардың өзара әрекеттесуін модельдеу мүмкіндігі жоқ. Олар мұғалімнің ізгілік көзқарасы идеясын ескермейді. Технологиялық қолданбалардың, әсіресе жасанды интеллекттің кеңінен қолданылуы барлық білім беру ұйымдары үшін айтарлықтай күрделі мәселе болып тұр. Технологиялар оқыту әдістері және ақпаратқа қол жеткізу жолдарын түбегейлі өзгерткенімен, жақсы ұйымдастырылған оқу жүйесінің негізгі бөлігі болып табылатын өзара жеке қарым-қатынас пен ынтымақтастықтың құндылығын ескермейді [9].

J. Gao жүргізген зерттеулерінде мұғалімдерден жеке кері байланыс пен қолдау алған студенттер автоматтандырылған, цифрландырылған академиялық бағдарламаларды қолданғандармен салыстырғанда білім алуда жоғары жетістіктер мен оқуда белсенділік танытқандықтарын көрсеткен [10].

Осылайша, технологияның гуманитарлық ғылымдарға әсерін түсінудің маңызды аспектісінде - білім алу үдерісіндегі адамдардың өзара әрекеттесу компоненті маңызды рөл атқарады. Бабер зерттеулерінде онлайн курстардағы студенттердің сол курсты дәстүрлі түрде

сыныпта оқыған курстастарына қарағанда нашар нәтиже көрсеткенін анықтаған. Бұл студенттерді тұтас оқу процесіне тарту мен ынтымақтастық пен әлеуметтік өзара әрекеттесуді күшейту қажеттілігін көрсетеді [11].

Білім беруді цифрландыру дәуірінде мұндай мәселелер бүкіл әлем бойынша педагогтар үшін ортақ проблемаға айналды [12]. J.P. Bernius және т.б. ғалымдар тобы зерттеулеріне сүйенетін болсақ аралас оқыту формасы, яғни күндізгі және онлайн оқытудың үйлесімі, тек технологиялар арқылы оқытуға қарағанда қатысушылардың белсенділігі мен қызығушылықтарын арттыратындығын анықтаған [13].

S. Diederich және т.б. ғалымдардың айтуы бойынша технологиялар оқу үдерісін жеңілдетіп, ақпаратты қосымша көмекші құрал ретінде пайдалана алғанымен, олар жеке оқыту мен білім беру үдерісі, оқу барысындағы адами қарым-қатынасты алмастыра алмайды деп тұжырымдайды [14].

S. Seufert, J. Guggemos, M. Sailer зерттеулерінде мұғалімдер математика сыныптарында жасанды интеллектті пайдалануды жеңілдетуде шешуші рөл атқарады деп қарастырады. Зерттеулер мұғалімдердің қабылдауы, сенімдері, педагогикалық дайындығы және көпжылдық тәжірибесі жасанды интеллект құралдарының іс жүзінде қалай қабылданатынына және қолданылатынына айтарлықтай әсер ететінін көрсетеді [15].

S. Forsstrom M. Nja, E. Munthe және т.б ғалымдар математикада робототехника, модельдеу, динамикалық геометрияға арналған бағдарламалық жасақтау, сандық ойындар және бағдарламалау орталары сияқты технологиялар қызығушылықты, түсінуді және жетістікке жетуді жақсартатынын дәлелденген [16]. B.G. Acosta-Enriquez, M.A. Ballesteros, C.R.V. Pérez және т.б.еңбектерінде жасанды интеллекттің тез дамуы мен таралуы этикалық және академиялық адалдыққа қатысты алаңдаушылық тудыратындығын және жасанды интеллект көмегімен оқуды өз бетінше оқумен біріктіру қажеттілігін көрсетеді және бұл жасанды интеллект құралдарын жауапкершілікпен пайдалану үшін нақты нұсқауларды талап етеді деп тұжырым жасайды [17]. H. Fakhar, M. Lamrabet, S. Sangheethaa, A. Korath ғалымдар тобының еңбектерінде жасанды интеллекттің сәтті интеграциялануы көбінесе мұғалімнің көзқарасы мен дайындығына байланысты, ал жасанды интеллектке дайындығы жоғары мұғалімдерде төмен тәуекелдер мен шығармашылық қабілеттерінің жоғары болатынын көрсеткен [18, 19].

Hartley, Hayak және Ko (2024) зерттеуінде ChatGPT құралын студенттердің өзіндік оқу әрекетін қолдау мақсатында пайдалану мүмкіндіктері бағдарламалауды меңгеру үдерісі негізінде қарастырылған. Зерттеу нәтижелері жасанды интеллектінің оқу материалын түсіндіру, тапсырмаларды орындауды жоспарлау, туындаған мәселелерді шешу және өзін-өзі бағалау барысында әр студенттің оқу қажеттілігіне сәйкес көмек көрсетуге мүмкіндік беретінін көрсетеді. Авторлар ChatGPT-дің студенттердің өзіндік оқу әрекетін ұйымдастыруға ықпал ететінін атап өтіп, оны тиімді пайдалану үшін студенттердің сұраныстарды дұрыс құрастыруы, алынған ақпараттың дұрыстығын тексеруі және өзінің оқу үдерісін бақылауы маңызды екенін көрсетеді [20].

Хіа және т.б. (2026) генеративті жасанды интеллект құралдарын университет студенттерінің өзіндік оқу әрекетін қолдауда қолдану мүмкіндіктерін зерттеген. Зерттеу нәтижелері ЖИ құралдарын педагогикалық қолдаумен қолдану студенттердің өзіндік тиімділігі мен жоғары деңгейлі ойлау дағдыларының дамуына оң ықпал ететінін көрсетті. Авторлар ЖИ құралдарын тиімді пайдалану педагогикалық қолдау мен оқу үдерісін дұрыс ұйымдастыруға байланысты екенін атап көрсетеді [21].

Зерттеу барысында психологиялық-педагогикалық, ғылыми-әдістемелік әдебиеттер зерделеніп, оларға талдау жасай келе жасанды интеллектіні қолдануды педагогикалық тұрғыдан реттеу мәселесі өзекті мәселе екендігін көрсетті.

Осындай негізгі мәселелерді ескере отырып, берілген мақаланың мақсаты жасанды интеллект құралдарын қолдану білім алушылардың өзіндік жұмысының нәтижелеріне қалай әсер ететінін және бұл үдерісте педагогикалық бақылаудың рөлін анықтау болып табылады.

Зерттеу әдіснамасы

Білім алушылардың өзіндік оқу жұмыстарын дайындауда жасанды интеллект құралдарын пайдаланудың тиімділігін анықтау мақсатында М.Әуезов атындағы Оңтүстік Қазақстан университетінің базасы «Компьютерлік инженерия және энергетика» жоғары мектебі кафедраларында эксперименттік зерттеу жұмыстары жүргізілді.

Экспериментке 6В06120 - Ақпараттық жүйелер, 6В06130 - Есептеу техникасы және бағдарламалық қамтамасыз ету, 6В06150 - Ақпараттық-коммуникациялық технологиялар және қауіпсіздік, 6В06121 - Жасанды интеллект технологиялары, 6В06141 - Виртуалды шындықты модельдеу және құру, 6В06140 - Математикалық және компьютерлік модельдеу білім беру бағдарламалары бойынша барлығы 182 білім алушы қатысты. Ондағы эксперименттік топта 92 бақылау тобында 90 білім алушыны құрады. Екі топтың бастапқы білім деңгейлері шамалас болды және олар арнайы іріктеу арқылы таңдалған жоқ.

Эксперимент басталғанға дейін білім алушылардың жасанды интеллект құралдарын пайдалану тәжірибесін анықтау мақсатында екі топтан да Google Forms платформасы арқылы сауалнама алынды. Сауалнама сұрақтары 1 – кестеде келтірілді. Сонымен қатар білім алушылармен сұхбат жүргізіліп, олардың зерттеу тақырыбына қатысты көзқарастары мен тәжірибелері анықталды.

Студенттердің оқу үлгерімі пән оқытушылары тарапынан бекітілген бағалау критерийлері, оқу жұмыстарын мазмұндық – талдау әдістері негізінде бағаланды.

Педагогикалық эксперимент бір академиялық кезең бойы 15 апта көлемінде өткізілді. Студенттердің нәтижелері 1, 3, 5, 7, 10 және 15-апталарда бақылауға алынды.

Зерттеу нәтижелерінің статистикалық мәндерінің арасындағы айырмашылықтарын анықтау мақсатында Стьюденттің t-критерийі, екі факторлы дисперсиялық талдау (ANOVA) және панельдік регрессия әдістері қолданылды.

Кесте 1. Студенттердің жасанды интеллект құралдарын пайдалануды бағалауға арналған саулнамадан үзінді.

№	*Сұрақ*	*Жауап түрі*	*Өлшеу мақсаты*
1	*Өзіндік жұмыстарды орындауда ЖИ қаншалықты жиі қолданасыз?*	*(ешқашан-әрдайым) шкала бойынша*	*Қолдану жиілігі*
2	*ЖИ қандай мақсатта қолданасыз? (эссе, іздеу, талдау және т.б.)*	*Көптеген таңдау*	*Когнитивті белсенділік түрі*
3	*Материалды ЖИ көмегінсіз түсінесіз бе?*	*Иә/толық емес/ жоқ*	*Түсіну тереңдігі*
4	*ЖИ жауаптарын тексересіз бе?*	*Шкала бойынша*	*Сыни ойлау*
5	*ЖИ негізгі ақпарат көзі ретінде қолданасыз ба?*	*Иә/жоқ*	*Тәуелділік*
6	*Сіздің жұмыстарыңыздың сапасы жақсарды ма?*	*Шкала бойынша*	*Тиімділікті өзіндік бағалау*
7	*ЖИ көмегінсіз жұмыс жасағанда қиындықтар туындайды ма?*	*Иә/жоқ*	*Когнитивті автономия*

Зерттеу нәтижелері

Зерттеу жұмысының нәтижесі тәжірибеге қатысқан білім беру бағдарламалары бойынша білім алушылардың өзіндік жұмыстарын ұйымдастыру, бағалау және бақылау жүйесі жетілдіріліп, оқу үдерісіне енгізілді. Мысалы: «6В06130 - Есептеу техникасы және бағдарламалық қамтамасыз ету» білім беру бағдарламасы бойынша «Схематехника», «Киберфизикалық жүйелердегі процестерді қорғау» және т.б. пәндері аясында білім алушылардың өзіндік жұмыстары жасанды интеллект құралдарын қолдану арқылы орындалды. «Киберфизикалық жүйелердегі процестерді қорғау» пәні бойынша білім

алушылар басқару жүйелерінің (ICS) қауіпсіздігі, SCADA жүйелеріндегі қауіпсіздік мәселелері, киберфизикалық жүйелерге жасалатын желілік шабуылдардың түрлері, машиналық оқыту әдістерін кибершабуылдарды анықтауда қолдану, ақылды өндіріс (Smart Factory) жүйелеріндегі қауіпсіздік және т.б. тапсырмалар беріліп, олардың орындалу нәтижелері бағалау жүйесіне сәйкес есепке алынып отырды.

Білім алушылардың өзіндік жұмыстарының нәтижелерін бағалау әр пән бойынша әзірленген жұмыс оқу бағдарламасындағы (Syllabus) көрсетілген тапсырмаларды орындау және тапсыру кестесіне сәйкес бағаланып отырды.

Өзіндік жұмыс жасау деңгейі. Білім алушылардың өзіндік жұмыс жасау деңгейі олардың білім беру тапсырмаларын басқалардың көмегінсіз, өз бетінше орындау қабілетінің дәрежесін көрсететін интегралды көрсеткіш ретінде бағаланды. Студенттердің өзіндік жұмыс жасау көрсеткіштерін өлшеу үшін сауалнама, сараптамалық бағалаудан тұратын кешенді әдіс қолданылды. Өзіндік жұмыс жасау көрсеткіштері 2 – кестедегі негізгі компоненттерден құралды.

Кесте 2. Өзіндік жұмыс жасау көрсеткіштерінің негізгі компоненттері.

№	Өзіндік жұмыс жасау көрсеткіштерінің компоненті	Компоненттің мазмұны
1	Когнитивтік дербестік	Оқу материалын оқытушының көмегінсіз түсіну қабілеті
2	Ақпаратты іздеу дағдылары	Тиісті дереккөздерді табу мен таңдай алуы
3	Сыни ойлау	Ақпаратты талдау мен тексере білуі
4	Бастамашылдық	Тапсырмаларды өз еркімен орындауға деген ынтасы
5	Өзін-өзі бақылау	Өзіндік нәтижелерді тексеру мен қатесін жөндеу қабілеті

Студенттердің өзіндік жұмыс жасау көрсеткіштерінің компоненттерін негізге ала отырып оқу тапсырмаларын меңгеру деңгейлері анықталды және деңгей көрсеткіштері: төмен, орташа, ортадан жоғары және жоғары деп 10 балдық шкала бойынша белгіленді (3-кесте).

Кесте 3. Оқу тапсырмаларын меңгеру деңгейін бағалау шкаласы.

№	Балл	Деңгей
1	1-3	Төмен
2	4-6	Орта
3	7-8	Ортадан жоғары
4	9-10	Жоғары

Жалпы балл барлық көрсеткіштердің орташа мәні ретінде келесі формула бойынша есептелді:

$$S = (C_1 + C_2 + C_3 + C_4 + C_5) / 5 \quad (1)$$

Студенттерді бағалау білім алушылардың өзіндік бағалауы (сауалнама), оқытушының бақылауы, өзіндік жұмыстарды орындауын талдауы, тапсырмаларды ауызша қорғау нәтижелері негізінде құрылды.

Өзіндік жұмыс жасау деңгейінің көрсеткіштері бақылау тобында 6,2 балдан 6,5 балға, ал тәжірибе тобында 6,3 балдан 8,4 балға дейін артты. Тәжірибе тобындағы көрсеткіштердің (6,3-тен 8,4-ке дейін) өсуі ақпаратты өңдеу құралдарын кеңінен қолданумен және тапсырмаларды

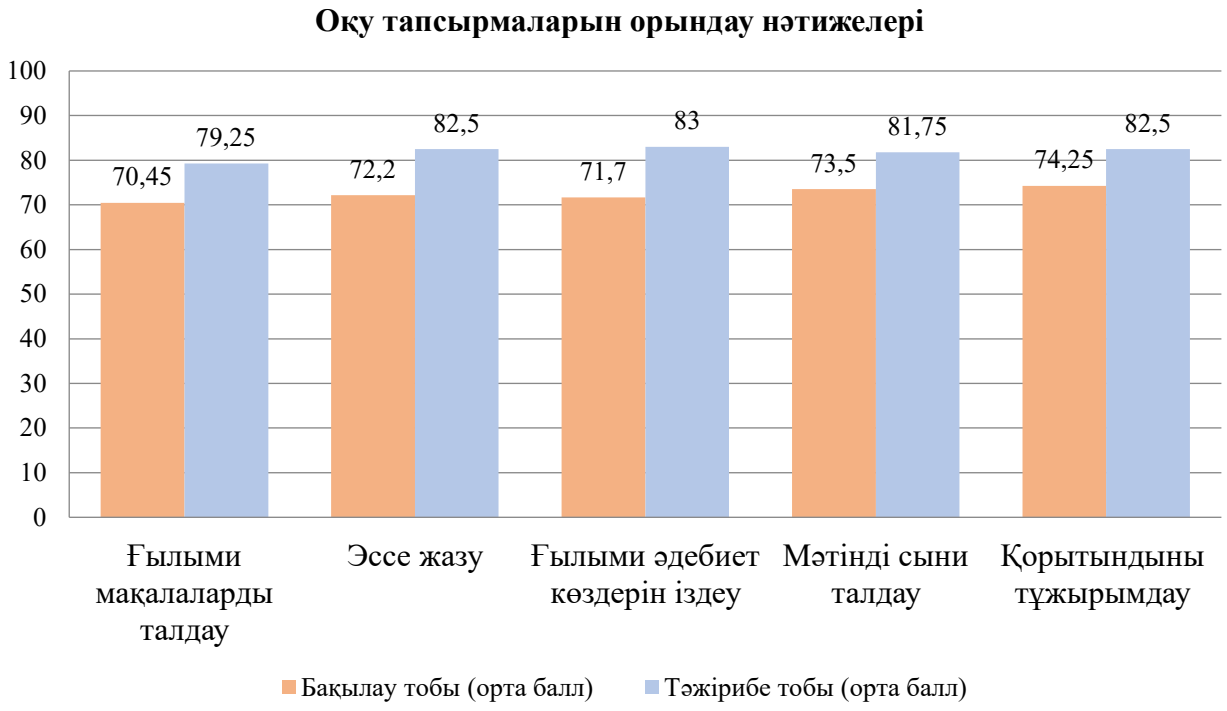
орындау тиімділігінің артуымен байланысты, бұл тұрақты өзіндік оқу дағдыларының қалыптасуын көрсетеді. Көрсеткіштердің бұл өсуі когнитивтік автономияның нақты дамуы мен педагогикалық бақылауды қажет ететін жасанды интеллектті қоса алғанда, құралдарды пайдалану тиімділігінің артуын да көрсетуі мүмкін екенін ескеру маңызды.

Корреляциялық талдау. Білім алушылардың өзіндік жұмыстарын орындау барысында жасанды интеллект технологиясын оқытушы бақылауында қолдану деңгейі мен оқу нәтижелері арасындағы өзара байланысты анықтау мақсатында корреляциялық талдау жүргізілді. Талдау барысында жасанды интеллект құралдарын оқытушы бақылауында пайдалану және тапсырмаларды орындау сапасы көрсеткіштері арасында байланыс зерттелді.

Білім алушылардың өзіндік жұмыстарын орындауда жасанды интеллект технологиясын оқытушы бақылауында қолдану мен оқу жұмыстарын орындау сапасының арасында оң байланыс бары анықталды: $r = 0,68$ ($p < 0,01$).

1- суретте білім алушылардың ең жоғары көрсеткіштері ғылыми әдебиет көздерін іздеу (бақылау тобында 71,7 балл, эксперименттік топта 83 балл), эссе жазу (бақылау тобында 72,2 балл, эксперименттік топта 82,5 балл) және қорытындыларды тұжырымдау дағдылары (бақылау тобында 74,25 балл, эксперименттік топта 82,5 балл) кезінде байқалғаны, ал мәтінді сыни талдау (бақылау тобында 73,5 балл, эксперименттік топта 81,75 балл) және ғылыми мақалаларды талдау (бақылау тобында 70,45 балл, эксперименттік топта 79,25 балл) онша көп өзгермегендігін көрсетті. Топтар арасындағы айырмашылықтар 8 балдан 12 балға дейінгі аралықта болды, бұл дисперстік талдауда анықталған «тапсырма түрі» факторының орташа әсерін растайды.

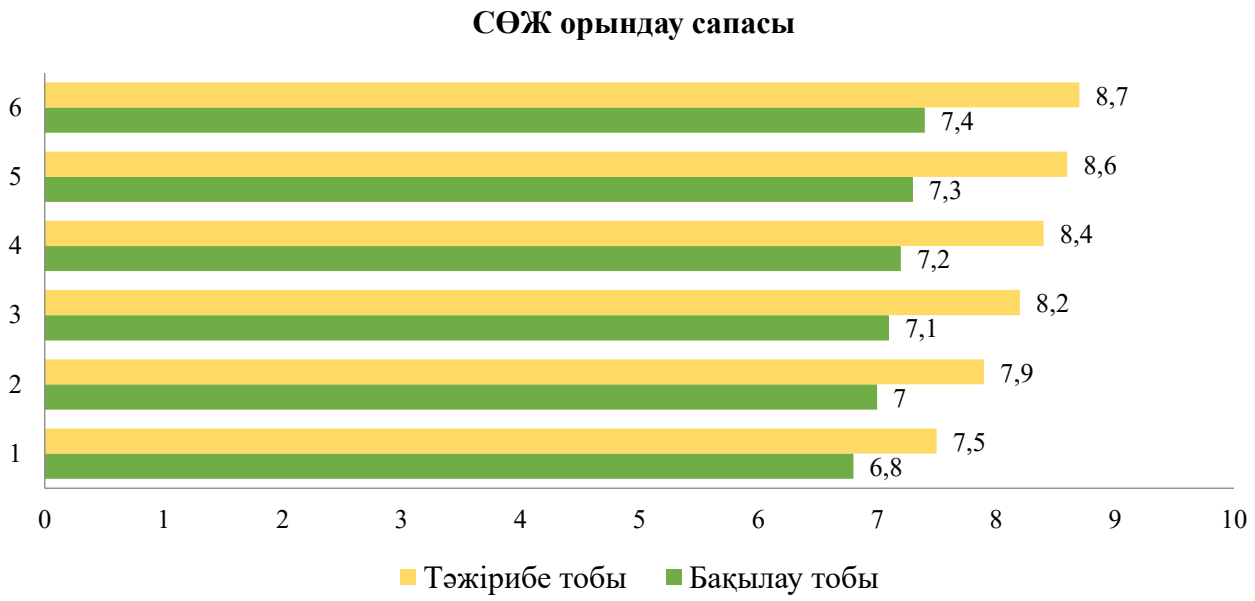
Бұл білім алушылардың жасанды интеллект құралдарын оқытушы бақылауында тиімді қолдану арқылы ақпаратты өңдеу, тапсырмаларды жоспарлау және оқу материалдарын меңгеру мүмкіндіктерінің артатындығын білдіреді.



Сурет 1. Оқу тапсырмаларын меңгеруін салыстыру

Бақылау және тәжірибе топтары арасындағы айырмашылықтардың әртүрлі болуы, бұл тапсырма түріне байланысты жасанды интеллекттің әртүрлі әсер еткендігін көрсетеді.

2-суретте екі топта да тапсырманы орындаудың тұрақты сызықтық өсу сапасы байқалады, алайда тәжірибелік топтағы өсу қарқыны жоғары. Бұл үрдіс регрессиялық талдау нәтижелерімен сәйкес келеді (β = 0,45; p < 0,01).



Сурет 2. СӨЖ орындау сапасының динамикасы (10 - балдық шкала бойынша)

4-кестеде жасанды интеллекттің білім алушылардың когнитивтік көрсеткіштеріне әсері көрсетілген. Жасанды интеллект өнімділікті арттыратынын, бірақ педагогикалық қолдаусыз терең когнитивтік процестерді әлсіретуі мүмкін екенін көрсетеді.

Кесте 4. Жасанды интеллекттің білім алушылардың когнитивті сипаттарына әсері.

№	Көрсеткіштер	Жағымды әсер	Қадағалау болмағандағы қауіп
1	Орындау жылдамдығы	Әлдеқайда артады	Шала-шарпы орындау
2	Мәтін құрылымы	Жақсарады	Шаблондық
3	Ақпаратты іздеу	Жеңілдейді	Іздеу қабілетінің жойылуы
4	Сыни ойлау	Қадағалауда дамиды	Төмендейді
5	Өзіндік әрекет ету	Инструменталдық артады	Когнитивтілік төмендейді

Д.Б. Кумахованың [22] жүргізген зерттеу нәтижелері: жасанды интеллект дұрыс интеграцияланған кезде сыни ойлау мен метакогнитивті дағдыларды дамыта алатынын, сонымен бірге дұрыс интеграцияланбаған жағдайда бірқатар қауіптер туындайтынын көрсеткен, ол біздің зерттеу нәтижелерімізді растайды.

Сапалық талдау. Білім алушылардың өзіндік жұмыстарын мазмұндық талдау мәтіннің дұрыс құрылуының артуын, аналитикалық сипаттаулардың көлемінің артуын және қателер санының азайғандығын көрсетті. Дегенмен, шектеулер де анықталды: дайын шешімдерді пайдалануға бейімділік, тәуелсіз талдау тереңдігінің төмендеуі, жасанды интеллектсіз ақпаратты іздеу дағдыларының ішінара жоғалуы. Бұл теріс әсерлер негізінен педагогикалық бақылау жеткіліксіз болған жағдайларда байқалғанын атап өту маңызды.

Зерттеу нәтижелерінің бақылау және тәжірибелік топтар арасындағы айырмашылықтарының статистикалық маңыздылығын бағалау үшін Стьюденттің t-критерийі қолданылды.

Есептеу формуласы:

$$t = \frac{\overline{x}_1 - \overline{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \quad (2)$$

мұндағы: $\overline{x}_1$, $\overline{x}_2$ - топтардағы көрсеткіштердің орташа мәндері, S_1^2, S_2^2 - дисперсиялар, n_1, n_2 - айнымалылардың көлемі. Қорытынды орташа мәндер: Тәжірибе тобы: $\overline{x}_1$= 82,6; Бақылау тобы: $\overline{x}_1$= 73,1

Қорытынды нәтижелерді салыстыру топтар арасында статистикалық тұрғыдан маңызды айырмашылықты көрсетті (t = 7,87; p < 0,01).

Дисперсиялық талдау (ANOVA). Зерттеуде екі факторды дисперсиялық талдау (ANOVA) қолданылды. Бұл әдіс оқу нәтижелеріне бірнеше фактордың ықпалын бір уақытта бағалауға мүмкіндік берді. Талдауда мынадай факторлар қарастырылды: А факторы - оқыту форматы (дәстүрлі оқыту/ЖИ + педагогикалық бақылау), В факторы – оқу тапсырмаларының түрі (5-кесте).

Кесте 5. Екі факторлы ANOVA

Вариация көздері	*SS*	*df*	*MS*	*F*	*p*
А Фактор (тобы)	*4410*	*1*	*4410*	*61,9*	*<0,001*
В Фактор (тақырыптар)	*1820*	*4*	*455*	*6,39*	*<0,01*
А × В	*960*	*4*	*240*	*3,37*	*<0,05*
Қате	*10800*	*1070*	*10,1*	-	-

Дисперсиялық талдау оқу факторының маңызды әсерін растады (f = 61,9; p < 0,001) және нәтижелердің тапсырма түріне тәуелділігін анықтады.

Бұл көрсеткіштер жасанды интеллект әсері барлық оқу тапсырмаларында бірдей емес екендігін дәлелдейді. Ең жоғары нәтиже ақпаратты іздеу және материалды құрылымдау тапсырмаларында байқалды. Осылайша, дисперсиялық талдау зерттеудің негізгі гипотезасын растайды: жасанды интеллект тек педагогикалық бақылаумен бірге қолданылған жағдайда тиімді нәтиже береді.

Панельдік регрессия моделі. Апталық динамика мен тапсырма түрлері арасындағы айырмашылықтарды есепке алу үшін панельдік регрессия моделі пайдаланылды:

$$Yit = \beta_0 + \beta_1 AI_it + \beta_2 Control_it + \beta_3 Week_t + \beta_4 Task_i + \varepsilon_it \quad (3)$$

мұндағы: Y - білім алушының нәтижесі, AI - жасанды интеллектіні пайдалану деңгейі, Control - педагогикалық бақылау деңгейі, Week - апта нөмірі, Task - тапсырма түрі. Панельдік регрессия нәтижелері 6 – кестеде келтірілді.

Кесте 6. Панельдік регрессия нәтижелері

Айнымалы	*Коэффициент (β)*	*t*	*p*
Тұрақты	*50,6*	*11,8*	*<0,001*
ЖИ (AI)	*2,8*	*6,2*	*<0,001*
Бақылау	*2,3*	*4,9*	*<0,001*
Апта	*0,45*	*3,8*	*<0,01*
Тапсырма түрі	*1,2*	*2,7*	*<0,05*

Модель сапасы: R^2 = 0,67; F = 64,8 (p < 0,001)

Нәтижелер көрсеткендей, жасанды интеллектті қолдану оқу үлгеріміне үлкен әсерін тигізеді (β = 2,8), ал педагогикалық бақылау (β = 2,3) бұл әсерін айтарлықтай күшейтеді. «Апта» айнымалысының оң коэффициенті (β = 0,45) нәтижелердің динамикалық тұрақты өсуін көрсетеді, бұл өзіндік жұмыстарды (СӨЖ) орындау кестесіндегі деректерге сәйкес келеді.

«Тапсырма түрі» факторы (β = 1,2) статистикалық тұрғыдан маңызды, бұл жасанды интеллекттің оқу іс-әрекеттерінің сипатына қарай әсері өзгереді деген қорытындыны растайды. Осылайша, регрессиялық модель жасанды интеллекттің тиімділігі оны қолданудың қарқындылығымен ғана емес, сонымен қатар педагогикалық бақылау дәрежесімен де анықталатынын растайды.

Зерттеу нәтижелері жасанды интеллект құралдарын студенттердің өзіндік жұмыстарында қолдану оқу тапсырмаларын орындау сапасы мен академиялық дербестікке оң әсер ететінін көрсетті. Сонымен қатар бұл әсердің деңгейі педагогикалық бақылаудың ұйымдастырылуына тікелей тәуелді екені анықталды. Педагогикалық бақылау жүйелі жүзеге асырылған жағдайда студенттердің оқу мотивациясы, тапсырмаларды сапалы орындауы және академиялық адалдық қағидаларын сақтауы айтарлықтай жоғары болды. Ал педагогикалық бақылау жеткіліксіз болған жағдайда студенттердің жасанды интеллектке тәуелділігі артып, дербес танымдық белсенділігі мен оқу тапсырмаларын саналы орындау деңгейінің төмендеу үрдісі байқалды. Статистикалық талдау нәтижелері педагогикалық бақылау жағдайында және бақылау әлсіреген жағдайда алынған көрсеткіштер арасында мәнді айырмашылық бар екенін көрсетті. Бұл нәтижелер жасанды интеллект құралдарын тиімді пайдалану педагогикалық бақылаумен ұштасқанда ғана тұрақты оң нәтиже беретінін дәлелдейді.

Дискуссия

Зерттеу нәтижелері жасанды интеллекттің оқу үлгеріміне әсері статистикалық тұрғыдан маңыздылығын, педагогикалық бақылау негізгі фактор болып табылатындығын растайды, оңтайлы әсерге жасанды интеллектті оқытушының қадағалауы арқылы пайдаланға қол жеткізуге болатындығын көрсетті. Динамикада нәтижелердің өсуі байқалды және ол соңғы кезеңдерде тұрақталатындығын көрсетті.

Осылайша, ANOVA талдауы зерттеудің негізгі гипотезасын растайды: Білім алушылардың СӨЖ жұмыстарын орындауда ЖИ қолдану тек педагогикалық тұрғыдан қадағаланатын түрде қолданылған кезде ғана тиімді оқыту құралы бола алады.

Зерттеу нәтижелері ЖИ-дің білім беру процесіне әсерінің қос сипатын растайды. Бір жағынан, интеллектуалды жүйелерді пайдалану: оқу өнімділігін арттырады; жұмыстардың құрылу сапасын жақсартады және материалды тезірек игеруге ықпал етеді. Екінші жағынан, оқытушының бақылауынсыз пайдалану кезінде: түсіну тереңдігінің төмендеуі; сандық құралдарға деген тәуелділіктің пайда болуы мен сыни ойлау қабілеті әлсірейді. Осылайша, ЖИ-ді мағыналы пайдалануды қамтамасыз ететін, тиімділіктің негізгі факторы - педагогикалық бақылау болып табылады.

Қорытынды

Жүргізілген зерттеу жұмыстары бізге келесідей қорытынды жасауға мүмкіндік берді: жасанды интеллект - оқу үлгерімін жақсарту және оқу іс-әрекеттерін оңтайландырудың тиімді құралы болып табылады, алайда оң әсерге тек жүйелі педагогикалық қолдау және бақылау арқылы қол жеткізілетіндігі анықталды. ЖИ-ді бақылаусыз пайдалану білім алушылардың когнитивтік әрекетінің сапасының төмендеуіне әкелді.Осылайша, тиімділіктің негізгі факторына технологияның өзі емес, оны білім беру ортасында тиімді, дұрыс қолдану шарттары жатады. Сонымен қатар, қазіргі таңдағы жасанды интеллект кеңінен таралған замандағы оқытушының рөлі модерациялау, сыни ойлауды бақылау мен дамытуға қарай өзгеруі қажет деп есептейміз.

Жүргізілген аналитикалық шолу мен зерттеу нәтижесі барысында алынған мәліметтер болашақта мынадай мәселелерді қарастырудың қажеттілігін айқындады:

- жасанды интеллект құралдарын білім беру процесіне нақты әдістемелік шектеулермен енгізу;
- білім алушылардың бойында жасанды интеллекттің берген жауаптарын сыни тұрғыдан бағалау дағдыларын дамыту;
- біріктірілген бағалау әдістерін (жазбаша тапсырмалар және ауызша қорғау сияқты) қолдану;
- оқытушылардың цифрлық құзыреттілігін арттыру;
- білім беруде жасанды интеллектті жауапкершілікпен пайдалану стандарттарын қалыптастыру.

Пайдаланылған әдебиеттер тізімі

[1] Y.K. Dwivedi, L. Hughes, E. Ismagilova, G. Aarts, C. Coombs, T. Crick, M.D. Williams Artificial intelligence (AI): multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. International Journal of Information Management, 57, 2021, Article 101994. https://doi.org/10.1016/j.ijinfomgt.2019.08.002

[2] https://www.gov.kz/memleket/entities/edu/press/news/details/1051555?lang=ru

[3] N.Y. Motlagh, M. Khajavi, A. Sharifi, M. Ahmadi. The Impact of Artificial Intelligence on the Evolution of Digital Education: A Comparative Study of Open AI Text Generation Tools including ChatGPT, Bing Chat, Bard, and Ernie. arXiv:2309.02029, 2023, v1. https://doi.org/10.48550/arXiv.2309.02029

[4] Alfarwan A. Generative AI use in K-12 education: a systematic review. Frontiers in Education. 2025. DOI: 10.3389/feduc.2025.1647573

[5] UNESCO. Guidance for generative AI in education and research. Paris, 2023.

[6] M. Jovanović, M. Campbell. Generative artificial intelligence: trends and prospects. Computer, 55 (10), 2022, pp. 107-112. DOI: 10.1109/MC.2022.3192720

[7] S. Jain, S. Basu, A. Ray, R. Das. Impact of irritation and negative emotions on the performance of voice assistants: netting dissatisfied customers' perspectives. International Journal of Information Management, 72, 2023, Article 102662. https://doi.org/10.1016/j.ijinfomgt.2023. 102662

[8] S. Diederich, A.B. Brendel, S. Morana, L. Kolbe. On the design of and interaction with conversational agents: an organizing and assessing review of human-computer interaction research. J. Assoc. Inf. Syst., 23 (1), 2022, pp. 96-138. DOI: 10.17705/1jais.00724

[9] C. Rapanta, L. Botturi, P. Goodyear, L. Guàrdia, M. Koole. Online university teaching during and after the covid-19 crisis: refocusing teacher presence and learning activity. Postdigital Science and Education, 2 (3), 2020, pp.924-945. https://doi.org/10.1007/s42438-020-00155-y

[10] J. Gao. Exploring the feedback quality of an automated writing evaluation system Pigai. Int. J. Emerg. Technol. Learn, 16 (11), 2021, p. 322. https://doi.org/10.3991/ijet.v16i11.19657

[11] H. Baber. Social Interaction and Effectiveness of the Online Learning - A Moderating Role of Maintaining Social Distance During the Pandemic COVID-19. Asian Education and Development Studies 2022, 11 (1): 159–171. https://doi.org/10.1108/AEDS-09-2020-0209

[12] E. Kasneci, K. Sessler, S. Küchemann, M. Bannert, D. Dementieva, F. Fischer, U. Gasser, G. Groh, S. Günnemann, E. Hüllermeier, S. Krusche, G. Kutyniok, T. Michaeli, C. Nerdel, J. Pfeffer, O. Poquet, M. Sailer, A. Schmidt, T. Seidel, G. Kasneci. ChatGPT for good? On opportunities and challenges of large language models for education. Learn. Individ. Differ., 103, 2023, Article 102274. https://doi.org/10.1016/j.lindif.2023.102274

[13] J.P. Bernius, S. Krusche, B. Bruegge. Machine learning based feedback on textual student answers in large courses. Computers and Education: Artificial Intelligence, 3, 2022. https://doi.org/10.1016/j.caeai.2022.100081

[14] S. Diederich, A.B. Brendel, S. Morana, L. Kolbe. On the design of and interaction with conversational agents: an organizing and assessing review of human-computer interaction research. J. Assoc. Inf. Syst., 23 (1), 2022, pp. 96-138. DOI: 10.17705/1jais.00724

[15] S. Seufert, J. Guggemos, M. Sailer. Technology-related knowledge, skills, and attitudes of pre- and in-service teachers: The current situation and emerging trends. Computers in Human Behavior, 115, 2020, Article 106552. https://doi.org/10.1016/j.chb.2020.106552

[16] S. Forsstrom, M. Nja, E. Munthe, J.-L. Alvarez-Galvan, L. Houldsworth. The impact of digital technologies on students' learning: Results from a literature review (OECD education working papers, no. 335) OECD Publishing, 2025. https://doi.org/10.1787/ab309c32-en

[17] B.G. Acosta-Enriquez, M.A. Ballesteros, C.R.V. Pérez, O.H. Jordan, J.A.M. Vergara, R.M. Acosta, C.G.A.P. Vargas, J.C.A. Castillo. AI in academia: How do social influence, self-efficacy, and integrity influence researchers' use of AI models? Social Sciences & Humanities Open, 11, 2025, Article 101274. https://doi.org/10.1016/j.ssaho.2025.101274

[18] H. Fakhar, M. Lamrabet, N. Echantoufi, K. El Khattabi, L. Ajana. Artificial intelligence from teachers' perspectives and understanding: Moroccan study. International Journal of Information and Education Technology, 14 (6), 2024, pp. 856-864. https://doi: 10.18178/ijiet.2024.14.6.2111

[19] S. Sangheethaa, A. Korath. Impact of AI in education through a teachers perceptive. Educational Administration: Theory and Practice, 30 (4), 2024, pp. 3196-3200. https://doi.org/10.53555/kuey.v30i4.1349

[20] K.Hartley, M. Hayak, U.H. Ko. Artificial Intelligence Supporting Independent Student Learning: An Evaluative Case Study of ChatGPT and Learning to Code. Educ. Sci. 2024,14, 120. https://doi.org/10.3390/educsci14020120

[21] Q. Xia, Y. Yang, W. Wang, H. Yin. Promoting Interdisciplinary Learning With Generative AI Through Self-Regulated Scaffolding. Journal of Computer Assisted Learning, 2026, Vol. 42(2). https://doi.org/10.1002/jcal.70214

[22] Д.Б. Кумахова. Влияние искусственного интеллекта на когнитивные способности студентов. Наука и Образование. Том 8. №4. 2025

References

[1] Y.K. Dwivedi, L. Hughes, E. Ismagilova, G. Aarts, C. Coombs, T. Crick, M.D. Williams Artificial intelligence (AI): multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. International Journal of Information Management, 57, 2021, Article 101994. https://doi.org/10.1016/j.ijinfomgt.2019.08.002

[2] https://www.gov.kz/memleket/entities/edu/press/news/details/1051555?lang=ru

[3] N.Y. Motlagh, M. Khajavi, A. Sharifi, M. Ahmadi. The Impact of Artificial Intelligence on the Evolution of Digital Education: A Comparative Study of Open AI Text Generation Tools including ChatGPT, Bing Chat, Bard, and Ernie. arXiv:2309.02029, 2023, v1. https://doi.org/10.48550/arXiv.2309.02029

[4] Alfarwan A. Generative AI use in K-12 education: a systematic review. Frontiers in Education. 2025. DOI: 10.3389/feduc.2025.1647573

[5] UNESCO. Guidance for generative AI in education and research. Paris, 2023.

[6] M. Jovanović, M. Campbell. Generative artificial intelligence: trends and prospects. Computer, 55 (10), 2022, pp. 107-112. DOI: 10.1109/MC.2022.3192720

[7] S. Jain, S. Basu, A. Ray, R. Das. Impact of irritation and negative emotions on the performance of voice assistants: netting dissatisfied customers' perspectives. International Journal of Information Management, 72, 2023, Article 102662. https://doi.org/10.1016/j.ijinfomgt.2023.102662

[8] S. Diederich, A.B. Brendel, S. Morana, L. Kolbe. On the design of and interaction with conversational agents: an organizing and assessing review of human-computer interaction research. J. Assoc. Inf. Syst., 23 (1), 2022, pp. 96-138. DOI: 10.17705/1jais.00724

[9] C. Rapanta, L. Botturi, P. Goodyear, L. Guàrdia, M. Koole. Online university teaching during and after the covid-19 crisis: refocusing teacher presence and learning activity. Postdigital Science and Education, 2 (3), 2020, pp. 924-945. https://doi.org/10.1007/s42438-020-00155-y

[10] J. Gao. Exploring the feedback quality of an automated writing evaluation system Pigai. Int. J. Emerg. Technol. Learn, 16 (11), 2021, p. 322. https://doi.org/10.3991/ijet.v16i11.19657

[11] H. Baber. Social Interaction and Effectiveness of the Online Learning - A Moderating Role of Maintaining Social Distance During the Pandemic COVID-19. Asian Education and Development Studies 2022, 11 (1): 159–171. https://doi.org/10.1108/AEDS-09-2020-0209

[12] E. Kasneci, K. Sessler, S. Küchemann, M. Bannert, D. Dementieva, F. Fischer, U. Gasser, G. Groh, S. Günnemann, E. Hüllermeier, S. Krusche, G. Kutyniok, T. Michaeli, C. Nerdel, J. Pfeffer, O. Poquet, M. Sailer, A. Schmidt, T. Seidel, G. Kasneci. ChatGPT for good? On opportunities and challenges of large language models for education. Learn. Individ. Differ., 103, 2023, Article 102274. https://doi.org/10.1016/j.lindif.2023.102274

[13] J.P. Bernius, S. Krusche, B. Bruegge. Machine learning based feedback on textual student answers in large courses. Computers and Education: Artificial Intelligence, 3, 2022. https://doi.org/10.1016/j.caeai.2022.100081

[14] S. Diederich, A.B. Brendel, S. Morana, L. Kolbe. On the design of and interaction with conversational agents: an organizing and assessing review of human-computer interaction research. J. Assoc. Inf. Syst., 23 (1), 2022, pp. 96-138. DOI: 10.17705/1jais.00724

[15] S. Seufert, J. Guggemos, M. Sailer. Technology-related knowledge, skills, and attitudes of pre- and in-service teachers: The current situation and emerging trends. Computers in Human Behavior, 115, 2020, Article 106552. https://doi.org/10.1016/j.chb.2020.106552

[16] S. Forsstrom, M. Nja, E. Munthe, J.-L. Alvarez-Galvan, L. Houldsworth. The impact of digital technologies on students\' learning: Results from a literature review (OECD education working papers, no. 335) OECD Publishing, 2025. https://doi.org/10.1787/ab309c32-en

[17] B.G. Acosta-Enriquez, M.A. Ballesteros, C.R.V. Pérez, O.H. Jordan, J.A.M. Vergara, R.M. Acosta, C.G.A.P. Vargas, J.C.A. Castillo. AI in academia: How do social influence, self-efficacy, and integrity influence researchers\' use of AI models? Social Sciences & Humanities Open, 11, 2025, Article 101274. https://doi.org/10.1016/j.ssaho.2025.101274

[18] H. Fakhar, M. Lamrabet, N. Echantoufi, K. El Khattabi, L. Ajana. Artificial intelligence from teachers\' perspectives and understanding: Moroccan study. International Journal of Information and Education Technology, 14 (6), 2024, pp. 856-864. https://doi: 10.18178/ijiet.2024.14.6.2111

[19] S. Sangheethaa, A. Korath. Impact of AI in education through a teachers perceptive. Educational Administration: Theory and Practice, 30 (4), 2024, pp. 3196-3200. https://doi.org/10.53555/kuey.v30i4.1349

[20] K.Hartley, M. Hayak, U.H. Ko. Artificial Intelligence Supporting Independent Student Learning: An Evaluative Case Study of ChatGPT and Learning to Code. Educ. Sci. 2024,14, 120. https://doi.org/10.3390/educsci14020120

[21] Q. Xia, Y. Yang, W. Wang, H. Yin. Promoting Interdisciplinary Learning With Generative AI Through Self-Regulated Scaffolding. Journal of Computer Assisted Learning, 2026, Vol. 42(2). https://doi.org/10.1002/jcal.70214

[22] D.B. Kumahova. Vliyanie iskusstvennogo intelekta na kognitivnye sposobnosti studentov. Nauka i Obrazovanie. Tom 8. №4. 2025

МРНТИ 14.35.07

https://doi.org/10.51889/2959-5894.2026.95.3.024

Г.Б. Камалова[1*], Ш.П.Турашова[1], М.Ерсин[1]

[1]Казахский национальный педагогический университет имени Абая, Алматы, Казахстан
[*]*e-mail: g.kamalova@abaiuniversity.edu.kz*

КОНЦЕПТУАЛЬНАЯ МОДЕЛЬ ОРГАНИЗАЦИИ И ЦИФРОВОГО СОПРОВОЖДЕНИЯ ДЕЯТЕЛЬНОСТИ SMART-АУДИТОРИИ В СИСТЕМЕ ПЕДАГОГИЧЕСКОГО ОБРАЗОВАНИЯ

Аннотация

В условиях цифровой трансформации педагогического образования SMART-аудитория рассматривается не только как технологически оснащенное учебное пространство, но и как интегрированная физико-цифровая среда, обеспечивающая организацию образовательной, проектной и исследовательской деятельности будущих педагогов. Несмотря на развитие исследований в данной области, в научной литературе преимущественно представлены отдельные технологические решения и направления применения SMART-аудиторий, тогда как вопросы их организации как целостной образовательной системы остаются недостаточно разработанными. В связи с этим возникает необходимость системного представления структуры SMART-аудитории, ее участников, ресурсов, направлений деятельности и взаимосвязей, а также определения требований к цифровому сопровождению этой деятельности. Цель исследования – разработать концептуальную модель организации и цифрового сопровождения деятельности SMART-аудитории в системе педагогического образования, отражающую ее структуру, участников, ресурсы, основные направления деятельности и взаимосвязи и определяющую функциональную структуру и информационную архитектуру цифровой платформы как технологической основы цифрового сопровождения. Методология исследования включала анализ научных источников, системный и функциональный анализ, концептуализацию SMART-аудитории, структурно-функциональное моделирование и логико-сценарную проверку разработанной модели. В результате определены семь взаимосвязанных компонентов SMART-аудитории, представлен цикл ее деятельности от планирования до анализа и распространения результатов, сформулированы требования к цифровому сопровождению, разработаны функциональная структура и информационная архитектура цифровой платформы. Предложенная концептуальная модель систематизирует структуру и организацию деятельности SMART-аудитории, устанавливает взаимосвязи между ее компонентами и определяет требования к цифровому сопровождению, формируя основу для проектирования единой цифровой платформы.

Ключевые слова: SMART-аудитория, SMART Education, цифровая платформа, IoT, концептуальная модель, интеллектуальная образовательная среда, цифровое сопровождение, информационная архитектура.

Г.Б. Камалова[1], Ш.П. Тұрашова[1], М.Е. Ерсін[1]

[1]Абай атындағы Қазақ ұлттық педагогикалық университеті, Алматы қ., Қазақстан

ПЕДАГОГИКАЛЫҚ БІЛІМ БЕРУ ЖҮЙЕСІНДЕГІ SMART АУДИТОРИЯ ҚЫЗМЕТІН ҰЙЫМДАСТЫРУДЫҢ ЖӘНЕ ЦИФРЛЫҚ СҮЙЕМЕЛДЕУДІҢ ТҰЖЫРЫМДАМАЛЫҚ МОДЕЛІ

Аңдатпа

Цифрлық трансформациясы жағдайында педагогикалық білім берудің SMART-аудиториясы тек технологиялық жабдықталған оқу кеңістігі ретінде ғана емес, сонымен қатар болашақ педагогтердің білім беру, жобалау және зерттеу қызметін ұйымдастыруды қамтамасыз ететін интеграцияланған физикалық-цифрлық орта ретінде қарастырылады. Осы саладағы зерттеулердің дамуына қарамастан, ғылыми әдебиеттерде негізінен SMART-аудиторияларды қолданудың жекелеген технологиялық шешімдері мен бағыттары ұсынылған, ал оларды біртұтас білім беру жүйесі ретінде ұйымдастыру мәселелері жеткіліксіз дамыған. Осыған байланысты SMART-аудитория құрылымын, оның қатысушыларын, ресурстарын, қызмет бағыттары мен өзара байланыстарын жүйелі түрде ұсыну,

сондай-ақ осы қызметті цифрлық сүйемелдеуге қойылатын талаптарды айқындау қажеттілігі туындайды. Зерттеудің мақсаты – педагогикалық білім беру жүйесіндегі SMART-аудитория қызметін ұйымдастыру мен цифрлық сүйемелдеудің тұжырымдамалық моделін әзірлеу, оның құрылымын, қатысушыларын, ресурстарын, қызметі мен өзара байланысының негізгі бағыттарын көрсететін және цифрлық сүйемелдеудің технологиялық негізі ретінде цифрлық платформаның функционалдық құрылымы мен ақпараттық архитектурасын айқындайтын. Нәтижесінде SMART-аудиторияның өзара байланысты жеті компоненті анықталды, оның жоспарлаудан бастап нәтижелерді талдауға және таратуға дейінгі циклі ұсынылды, цифрлық сүйемелдеуге қойылатын талаптар тұжырымдалды, цифрлық платформаның функционалдық құрылымы мен ақпараттық архитектурасы әзірленді. Ұсынылған тұжырымдамалық модель SMART-аудитория қызметінің құрылымы мен ұйымдастырылуын жүйелейді, оның компоненттері арасындағы байланысты белгілейді және бірыңғай цифрлық платформаны жобалау үшін негіз құра отырып, цифрлық сүйемелдеуге қойылатын талаптарды айқындайды.

Түйін сөздер: SMART-аудитория, SMART Education, цифрлық платформа, IoT, тұжырымдамалық модель, интеллектуалды білім беру ортасы, цифрлық сүйемелдеу, ақпараттық сәулет.

G. Kamalova[1], Sh. Turashova[1], M. Yersin [1]
[1]Abai Kazakh National Pedagogical University, Almaty, Kazakhstan

CONCEPTUAL MODEL OF THE ORGANIZATION AND DIGITAL SUPPORT OF SMART CLASSROOM ACTIVITIES IN THE PEDAGOGICAL EDUCATION SYSTEM

Abstract

In the context of the digital transformation of pedagogical education, a SMART classroom is viewed not only as a technologically equipped learning space, but also as an integrated physical-digital environment that supports the organization of educational, project-based, and research activities for future teachers. Despite the progress in research in this field, the scientific literature predominantly presents individual technological solutions and areas of application for SMART classrooms, while the issues of organizing them as a holistic educational system remain underdeveloped. In this regard, there is a need for a systematic representation of the structure of the SMART audience, its participants, resources, areas of activity, and interconnections, as well as for defining the requirements for digital support of this activity. The aim of the study is to develop a conceptual model for organizing and digitally supporting the activities of a SMART audience within the system of pedagogical education, reflecting its structure, participants, resources, main areas of activity, and interrelationships, and defining the functional structure and information architecture of the digital platform as the technological foundation for digital support. The research methodology included an analysis of scientific sources, systemic and functional analysis, conceptualization of the SMART audience, structural-functional modeling, and logical-scenario testing of the developed model. As a result, seven interrelated components of the SMART audience have been identified, the cycle of its activities has been presented from planning to analysis and dissemination of results; requirements for digital support have been formulated; and the functional structure and information architecture of the digital platform have been developed. The proposed conceptual model systematizes the structure and organization of the SMART audience’s activities, establishes the relationships between its components, and defines the requirements for digital support, forming the basis for designing a unified digital platform.

Keywords: SMART Classroom, SMART Education, digital platform, IoT, conceptual model, intelligent educational environment, digital support, information architecture.

Введение

Основные положения

SMART-аудитория рассматривается как целостная физико-цифровая образовательная система, в которой физическая инфраструктура выступает ресурсной основой образовательной, проектной и исследовательской деятельности, а цифровая платформа обеспечивает ее цифровое сопровождение. Разработанная концептуальная модель отражает структуру SMART-аудитории, ее участников, ресурсы, основные направления деятельности, взаимосвязи и процессы управления, связывая физический и цифровой уровни в единую систему. В рамках модели выделены семь взаимосвязанных компонентов SMART-аудитории,

для цифрового сопровождения которых определены требования и разработаны функциональная структура и информационная архитектура единой цифровой платформы.

Концептуальная модель позволяет рассматривать цифровую платформу как средство не только представления и управления физической инфраструктурой, но и организации и сопровождения образовательной, проектной и исследовательской деятельности SMART-аудитории.

Цифровая трансформация образования постепенно приводит к изменению подходов к организации образовательного процесса. Цифровые технологии перестают использоваться преимущественно как отдельные инструменты и становятся частью взаимосвязанной образовательной среды. Если на начальных этапах цифровизации основное внимание уделялось оснащению учебных аудиторий компьютерами, мультимедийным оборудованием и электронными образовательными ресурсами, то развитие IoT, облачных технологий, искусственного интеллекта (ИИ) и средств анализа данных позволяет объединять оборудование, цифровые ресурсы, участников и данные образовательного процесса в единую систему. Такая интеграция расширяет возможности образовательной среды, она позволяет не только обеспечивать доступ к цифровым ресурсам, но и собирать данные о ходе обучения, автоматизировать отдельные процессы, реагировать на изменения и поддерживать принятие педагогических решений на основе анализа информации. Существенную роль в этом играют IoT-технологии, обеспечивающие получение данных о физических объектах и параметрах образовательной среды, и аналитические инструменты, позволяющие интерпретировать эти данные, выявлять изменения в учебной активности обучающихся и своевременно обнаруживать возможные затруднения.

Переход от использования отдельных цифровых инструментов к взаимосвязанной, адаптивной и интеллектуально поддерживаемой образовательной среде лежит в основе концепции SMART Education. В данном контексте понятие «SMART» характеризует не просто наличие современных технологий, а способность образовательной среды взаимодействовать с участниками и объектами, получать и использовать данные, автоматизировать отдельные процессы, учитывать контекст обучения и поддерживать принятие обоснованных педагогических решений. Такой подход предполагает объединение технологических и педагогических компонентов образовательной среды и создание условий для более гибкой организации обучения. SMART Education, таким образом, можно рассматривать как направление развития цифрового образования, ориентированное на создание более гибкой, персонализированной и управляемой на основе данных образовательной среды.

Одной из форм практической реализации принципов SMART Education становится SMART-аудитория. В данном исследовании она рассматривается не просто как учебное помещение, оснащенное современным оборудованием, а как интегрированная физико-цифровая образовательная система, объединяющая два взаимосвязанных уровня: физическую инфраструктуру и единую цифровую платформу. Физическая инфраструктура включает сенсоры и другие IoT-устройства, сетевую инфраструктуру, интерактивное оборудование, средства визуализации и другие технологические ресурсы, обеспечивающие ресурсную основу образовательной, проектной и исследовательской деятельности. Цифровая платформа обеспечивает цифровое представление и управление ресурсами, данными и процессами, а также цифровое сопровождение деятельности SMART-аудитории. Взаимодействие этих уровней позволяет связать физические объекты и процессы SMART-аудитории с их цифровым представлением, использовать данные о состоянии среды и деятельности участников и объединять различные виды деятельности в едином цифровом пространстве.

В отличие от традиционной аудитории, в которой цифровые устройства преимущественно используются для представления и передачи учебной информации, SMART-аудитория создает условия для практического взаимодействия обучающихся с технологиями, данными и цифровыми сервисами. Она может одновременно выступать образовательной, технологической и проектно-исследовательской средой, в которой обучающиеся не только

используют цифровые инструменты, но и приобретают опыт проектирования, анализа и оценки цифровых решений. Поэтому SMART-аудитория представляет собой не только элемент образовательной инфраструктуры, но и пространство организации различных видов деятельности, объединяющих обучение, практику и исследование.

Особую значимость такой подход приобретает в системе педагогического образования. Подготовка будущего учителя предполагает формирование не только навыков работы с отдельными цифровыми инструментами, но и способности проектировать образовательный процесс в условиях цифровой среды, использовать данные для принятия педагогических решений и интегрировать современные технологии в профессиональную деятельность. SMART-аудитория предоставляет для этого практическую среду, в которой будущие педагоги могут осваивать цифровые технологии в условиях их применения, разрабатывать образовательные сценарии, реализовывать проекты, анализировать результаты и оценивать педагогический потенциал цифровых решений. В этом смысле SMART-аудитория приобретает двойное значение. Она является одновременно средой обучения и средой профессиональной подготовки будущего педагога к деятельности в условиях цифровой трансформации образования.

Исследования SMART-аудиторий и интеллектуальных сред обучения охватывают широкий круг технологических и педагогических вопросов. Рассматриваются возможности IoT, облачных технологий, сенсорных систем, искусственного интеллекта, аналитических инструментов и систем управления обучением для расширения возможностей образовательной среды и поддержки более гибкой организации обучения [1,2]. Hwang рассматривает интеллектуальные образовательные среды с позиции контекстной осведомленности и способности учитывать состояние обучающегося и параметры окружающей среды, тогда как Koper [3] акцентирует внимание на взаимодействии человека с цифровыми и адаптивными компонентами образовательного пространства. Обзор Kaur, Bhatia и Stea [4] показывает, что исследования SMART-аудиторий охватывают широкий спектр технологий – от сенсорных сетей и облачных сервисов до машинного обучения и систем управления обучением. При этом авторы отмечают необходимость более целостного и междисциплинарного подхода, поскольку разрозненное внедрение отдельных технологий не позволяет в полной мере раскрыть потенциал SMART-аудитории. Аналогичные выводы содержатся в работах Tabuenca и соавторов [5] и исследованиях интеллектуальных сред обучения [6], где подчеркивается необходимость одновременного учета педагогического, цифрового и физического компонентов образовательного пространства.

Несмотря на значительное количество исследований, посвященных технологическим возможностям SMART-аудиторий, вопросы их организации как целостных образовательных систем остаются недостаточно разработанными. В научных работах отдельно рассматриваются IoT-устройства, цифровые ресурсы, аналитические инструменты, образовательные сценарии и вопросы подготовки преподавателей, однако недостаточно внимания уделяется тому, как эти элементы объединяются в рамках конкретной SMART-аудитории, какие функции она выполняет, кто участвует в ее деятельности, какие процессы в ней происходят и как связаны между собой ее технологические, образовательные, проектные, исследовательские и организационные компоненты.

Это ограничивает возможности обоснованного цифрового сопровождения SMART-аудитории. Цифровая платформа должна не просто предоставлять доступ к отдельным ресурсам или обеспечивать управление подключенными устройствами. Ее функциональная структура должна соответствовать реальным процессам, участникам, ресурсам и направлениям деятельности SMART-аудитории. Поэтому для ее проектирования необходимо предварительно определить, из каких компонентов состоит SMART-аудитория, какие функции они выполняют, кто участвует в ее деятельности, какие взаимосвязи между ними возникают и какие процессы требуют цифровой поддержки. Без такого целостного

представления существует риск построения платформы как набора разрозненных сервисов, не отражающих реальную организацию деятельности SMART-аудитории.

В связи с этим в данном исследовании SMART-аудитория рассматривается как целостная физико-цифровая образовательная система, в которой взаимосвязаны инфраструктура, участники, ресурсы, виды деятельности и процессы управления. Такой подход позволяет перейти от рассмотрения отдельных технологических элементов к определению общей логики функционирования SMART-аудитории и установить, каким образом ее структура и деятельность должны быть представлены в цифровой среде. На этой основе становится возможным определить требования к цифровому сопровождению, состав функциональных компонентов и взаимосвязи цифровой платформы.

Таким образом, для организации и цифрового сопровождения деятельности SMART-аудитории необходимо целостно представить ее структуру, участников, ресурсы, основные направления деятельности и взаимосвязи, а также определить их отражение в цифровой среде. Такое представление позволяет связать организацию деятельности в физическом пространстве с ее цифровым сопровождением и обосновать функциональную структуру и информационную архитектуру соответствующей цифровой платформы.

Цель исследования – разработать концептуальную модель организации и цифрового сопровождения деятельности SMART-аудитории в системе педагогического образования, отражающую ее структуру, участников, ресурсы, основные направления деятельности и взаимосвязи и определяющую функциональную структуру и информационную архитектуру цифровой платформы как технологической основы цифрового сопровождения.

Методология исследования

Данное исследование носит концептуально-модельный характер и направлено на разработку концептуальной модели организации и цифрового сопровождения деятельности SMART-аудитории в системе педагогического образования. В основу исследования положено понимание SMART-аудитории не как отдельного технически оснащенного учебного помещения, а как целостной физико-цифровой образовательной системы, объединяющей физическую инфраструктуру и единую цифровую платформу, а также участников, ресурсы, виды деятельности, данные и процессы управления. Такой подход позволил рассматривать цифровую платформу как цифровой уровень SMART-аудитории, обеспечивающий интеграцию ее ресурсов, процессов, данных и участников и их цифровое сопровождение.

Исследование проводилось в несколько взаимосвязанных этапов.

На первом этапе осуществлялся анализ научных публикаций и современных подходов к цифровой трансформации педагогического образования, SMART Education, интеллектуальным образовательным средам, использованию IoT и цифровых платформ в образовательном процессе. Анализ позволил определить основные характеристики SMART-образовательных сред, направления их применения в педагогическом образовании и проблемы, связанные с организацией их деятельности. Особое внимание уделялось исследованиям, рассматривающим интеграцию физической и цифровой образовательной среды, использование данных и аналитических инструментов, проектную деятельность и цифровые средства взаимодействия.

На втором этапе осуществлялась концептуализация SMART-аудитории как объекта исследования. Определялись ее назначение, физическая инфраструктура, цифровой уровень, основные ресурсы, участники и направления деятельности. В качестве исходного положения было принято, что наличие IoT-устройств, сенсоров, интерактивного оборудования и сетевой инфраструктуры само по себе не определяет SMART-аудиторию как целостную образовательную систему. Поэтому анализировалась взаимосвязь физической инфраструктуры с цифровыми ресурсами, образовательными, проектными, исследовательскими, аналитическими, коммуникационными и организационно-

управленческими процессами. Такой подход позволил перейти от описания отдельных технических средств к определению структуры и логики функционирования SMART-аудитории.

На третьем этапе с использованием функционального анализа определялись основные виды деятельности, которые должны быть включены в систему организации и цифрового сопровождения SMART-аудитории. Были выделены образовательная, проектно-практическая, исследовательская, аналитическая, коммуникационная и организационно-управленческая деятельность, а также деятельность по представлению и распространению результатов. Для каждого направления определялись основные процессы, необходимые ресурсы, участники и ожидаемые результаты. Анализировалось, какие процессы требуют цифровой поддержки, какие данные возникают в ходе их выполнения, кто должен иметь к ним доступ и какие функции необходимы для их организации, мониторинга и управления. Это позволило связать основные виды деятельности SMART-аудитории с конкретными задачами цифрового сопровождения.

На четвертом этапе проводилась структуризация выявленных компонентов и установление связей между ними. Системный подход использовался для определения места каждого компонента в общей структуре SMART-аудитории и характера его взаимодействия с другими компонентами. В результате были выделены семь взаимосвязанных компонентов: инфраструктурный, образовательный, проектно-практический, аналитический, исследовательский, организационно-управленческий и информационно-презентационный. При этом физическая инфраструктура рассматривалась как ресурсная основа системы, а единая цифровая платформа – как ее цифровой уровень, обеспечивающий представление и интеграцию ресурсов, процессов, данных и участников.

На пятом этапе характеристики и процессы SMART-аудитории преобразовывались в требования к ее цифровому сопровождению. Для каждого компонента и направления деятельности определялись необходимые цифровые функции, данные, информационные связи и механизмы взаимодействия участников. При формировании требований учитывалась необходимость цифровой поддержки работы с физическими и цифровыми ресурсами, подключенным оборудованием и IoT-устройствами, организации образовательной, проектной и исследовательской деятельности, сбора и обработки данных, их визуализации и аналитического представления, коммуникации участников, ведения электронных портфолио и репозиториев, а также представления и распространения результатов деятельности. Дополнительно учитывались требования к интеграции сервисов, обмену данными между компонентами, разграничению доступа и формированию единого информационного пространства. На этой основе определялись функциональный состав цифровой платформы и связи между ее основными разделами.

На шестом этапе на основе установленных компонентов, связей и требований разрабатывалась концептуальная модель организации и цифрового сопровождения деятельности SMART-аудитории. Модель отражает взаимосвязь целей, физической инфраструктуры, цифровой платформы, участников, ресурсов, видов деятельности, процессов управления и средств их цифровой поддержки. При ее построении учитывалась последовательность основных процессов: планирование деятельности, предоставление ресурсов, реализация образовательных, проектных и исследовательских мероприятий, сбор и анализ данных, фиксация результатов, взаимодействие участников и представление полученных результатов. Это позволило представить SMART-аудиторию как динамическую физико-цифровую систему, в которой результаты деятельности используются для последующего анализа, совершенствования процессов и развития образовательных практик.

На заключительном этапе проводилась логико-структурная и сценарная проверка разработанной модели. Ее целостность оценивалась по соответствию цели исследования, полноте представления основных компонентов SMART-аудитории, согласованности связей между ними и соответствию предусмотренных цифровых средств основным процессам и

направлениям ее деятельности. Дополнительно модель проверялась на типовых сценариях деятельности, таких как организация учебного занятия с использованием IoT-оборудования, реализация студентами проектной работы, мониторинг состояния SMART-аудитории, анализ образовательных данных, проведение исследовательского проекта и представление его результатов профессиональному сообществу. Такая проверка позволила определить, охватывает ли модель основные процессы SMART-аудитории и обеспечивает ли согласованное представление их цифровой поддержки в рамках единой платформы.

Таким образом, методология исследования сочетает анализ научных источников, системный и функциональный анализ, концептуализацию объекта, структурно-функциональное моделирование и логико-сценарную проверку разработанной модели. Такой комплекс методов соответствует цели исследования, поскольку позволяет последовательно представить SMART-аудиторию как целостную физико-цифровую образовательную систему, определить структуру и взаимосвязи ее деятельности, сформировать требования к цифровому сопровождению и разработать концептуальную модель, на основе которой обосновываются функциональная структура и информационная архитектура цифровой платформы.

Результаты исследования

Проведенное исследование позволило уточнить представление о SMART-аудитории в системе педагогического образования. Она рассматривается не просто как учебное помещение, оснащенное современными техническими средствами, а как интегрированная физико-цифровая образовательная среда, объединяющая физическую инфраструктуру и цифровую платформу, в рамках которых взаимосвязаны ресурсы, участники, данные и процессы образовательной, проектной и исследовательской деятельности. Это соответствует современному пониманию интеллектуальной образовательной среды как пространства, в котором физические, цифровые и педагогические компоненты функционируют во взаимосвязи [6].

Физическую основу SMART-аудитории составляют учебное помещение и его технологическая инфраструктура, а именно IoT-устройства и сенсоры, контроллеры, исполнительные устройства, сетевое и интерактивное оборудование, серверные и облачные ресурсы. Они обеспечивают мониторинг параметров образовательной среды, управление оборудованием и передачу данных в цифровую среду. Однако техническое оснащение само по себе не обеспечивает целостного функционирования SMART-аудитории. Для этого необходима цифровая среда, объединяющая образовательные сервисы, инструменты проектной и исследовательской деятельности, аналитические средства, репозитории материалов и результатов, коммуникационные средства и механизмы управления и информационного сопровождения деятельности аудитории. Использование IoT расширяет эти возможности за счет получения данных о физических объектах и параметрах образовательной среды, которые могут использоваться в образовательном процессе и аналитике [7,8]. Взаимодействие физического и цифрового уровней позволяет связать реальные объекты и процессы SMART-аудитории с их цифровым представлением и обеспечить сопровождение различных видов деятельности в едином цифровом пространстве.

На этой основе в исследовании выделены семь взаимосвязанных компонентов организации и деятельности SMART-аудитории. Это инфраструктурный, образовательный, проектно-практический, аналитический, исследовательский, организационно-управленческий и информационно-презентационный компоненты. Они отражают основные ресурсы, направления деятельности и процессы, которые должны быть учтены при организации SMART-аудитории и ее цифровом сопровождении.

Инфраструктурный компонент образует технологическую основу SMART-аудитории. Он включает IoT-устройства, сенсоры, контроллеры, исполнительные устройства, локальную сеть и Wi-Fi, серверные и облачные ресурсы, интерактивное оборудование и средства

идентификации пользователей, а также цифровые ресурсы, необходимые для реализации образовательных, проектных и исследовательских сценариев.

Образовательный компонент определяет содержание и средства учебной работы в SMART-аудитории. Он включает электронные курсы, цифровые учебные материалы, виртуальные лаборатории, тренажеры и средства оценивания, обеспечивающие теоретическую подготовку, практическое освоение технологий и оценку результатов обучения. Например, при изучении темы IoT студенты сначала осваивают теоретический материал в электронном курсе и выполняют микро-задание, затем в виртуальной лаборатории моделируют подключение датчика к микроконтроллеру и после этого реализуют аналогичную задачу на реальном оборудовании SMART-аудитории. Результаты выполнения заданий и практической работы фиксируются в цифровой среде и могут использоваться для анализа учебной деятельности. Виртуальные лаборатории при этом позволяют безопасно отрабатывать практические навыки и формировать цифровой след, который может быть использован для последующего анализа результатов обучения [9].

Проектно-практический компонент обеспечивает применение освоенных технологий при решении практических задач. Например, в проекте «Мониторинг параметров SMART-аудитории» студенты разрабатывают систему сбора данных о температуре, влажности и концентрации CO_2, передают их через микроконтроллер в цифровую среду и отображают текущие значения на аналитической панели. Такой проект объединяет программирование микроконтроллеров, работу с IoT-устройствами, передачу и анализ данных, и принятие решений по организации образовательной среды.

Аналитический компонент обеспечивает сбор и анализ данных об образовательной активности, использовании оборудования, физических параметрах среды и результатах учебной и проектной деятельности. Анализ этих данных позволяет выявлять изменения в учебной активности студентов, затруднения при выполнении заданий и своевременно корректировать учебную работу. Например, преподаватель может анализировать посещаемость и выполнение заданий, результаты работы в виртуальной лаборатории и участие студентов в проектах. Данные о температуре, освещенности и концентрации CO_2 могут использоваться для оценки физических условий проведения занятий. Таким образом, аналитический компонент не ограничивается отображением данных, а служит основой для принятия педагогических и организационных решений. Современные исследования learning analytics в высшем образовании также показывают расширение практик использования образовательных данных для анализа учебной активности и результатов обучающихся [10].

Исследовательский компонент обеспечивает использование SMART-аудитории как площадки для учебных и научных исследований. Он включает электронные репозитории проектов, цифровую библиотеку и электронные журналы наблюдений. Накопление и систематизация таких материалов позволяет сохранять результаты исследований и использовать их в последующей образовательной и проектной деятельности.

Организационно-управленческий компонент обеспечивает координацию деятельности и взаимодействие участников. Он включает средства совместной работы, консультаций, планирования и координации проектов, а также инструменты управления ресурсами и доступом к ним.

Информационно-презентационный компонент связан с представлением и распространением результатов деятельности SMART-аудитории. К ним относятся студенческие проекты, цифровые разработки, методические материалы, результаты исследований и успешные образовательные практики. Их представление в профессиональном пространстве создает условия для обмена опытом, распространения успешных решений и взаимодействия с профессиональным сообществом.

Выделенные компоненты функционируют во взаимосвязи и объединяются общими ресурсами, данными, участниками и результатами. Например, при разработке системы мониторинга студенты используют оборудование инфраструктурного компонента, учебные

материалы образовательного компонента, выполняют проектные задачи, анализируют полученные данные, сохраняют результаты в электронном портфолио и исследовательском репозитории и представляют готовое решение профессиональному сообществу. Таким образом, один практический сценарий объединяет ресурсы и процессы всех семи компонентов в единую систему.

Информация, формируемая в ходе деятельности, также перемещается между компонентами. Данные IoT-устройств используются в учебных заданиях и проектах, результаты проектной работы сохраняются в цифровых репозиториях, аналитические показатели применяются для оценки результатов и планирования последующей деятельности, а успешные разработки представляются профессиональному сообществу. Поэтому деятельность SMART-аудитории может быть представлена как взаимосвязанный цикл планирование → организация → реализация → мониторинг → анализ → фиксация результатов → представление и распространение → совершенствование деятельности. Такая последовательность показывает, что цифровое сопровождение должно охватывать не отдельные технические операции, а весь цикл деятельности. При этом результаты одного этапа могут становиться ресурсом для последующих этапов, обеспечивая непрерывность и цикличность процессов.

Для поддержки такого цикла необходима единая цифровая платформа, которая связывает физическую инфраструктуру SMART-аудитории с ее цифровым представлением и объединяет ресурсы, данные, участников и организационные процессы. Она представляет собой цифровой уровень SMART-аудитории, основное ее назначение заключается в цифровой интеграции и сопровождении деятельности, реализуемой в рамках выделенных компонентов.

На этой основе разработана концептуальная модель организации и цифрового сопровождения деятельности SMART-аудитории. Модель объединяет физическую инфраструктуру, цифровую платформу, ресурсы, участников, семь взаимосвязанных компонентов, виды деятельности, данные, результаты и процессы управления. Ее структура отражает взаимосвязь физического и цифрового уровней, а также движение ресурсов, информации и результатов между компонентами. Физическая инфраструктура выступает ресурсной основой образовательной, проектной и исследовательской деятельности, тогда как цифровая платформа обеспечивает ее цифровое представление, взаимодействие участников, обработку данных и поддержку основных процессов. В результате SMART-аудитория представляется как динамическая система, в которой результаты деятельности используются для последующего анализа, совершенствования процессов и развития образовательных практик представлены на рисунке 1.

Разработанная модель позволила определить, какие процессы SMART-аудитории требуют цифровой поддержки и какие функции должна обеспечивать цифровая платформа для организации, сопровождения и взаимосвязи. Функциональная структура платформы формировалась не на основе перечня доступных цифровых технологий, а исходя из реальных видов деятельности, ресурсов и информационных потоков SMART-аудитории. В результате определены следующие основные функции цифрового сопровождения:

– обеспечение доступа к физическим и цифровым ресурсам;
– управление и мониторинг IoT-инфраструктуры;
– организация образовательной, проектной и исследовательской деятельности;
– сбор, хранение и обработка данных;
– визуализация и аналитическое представление информации;
– организация коммуникации и совместной работы;
– хранение результатов и формирование электронных портфолио;
– представление и распространение разработок и успешных практик;
– интеграция перечисленных функций в едином информационном пространстве.

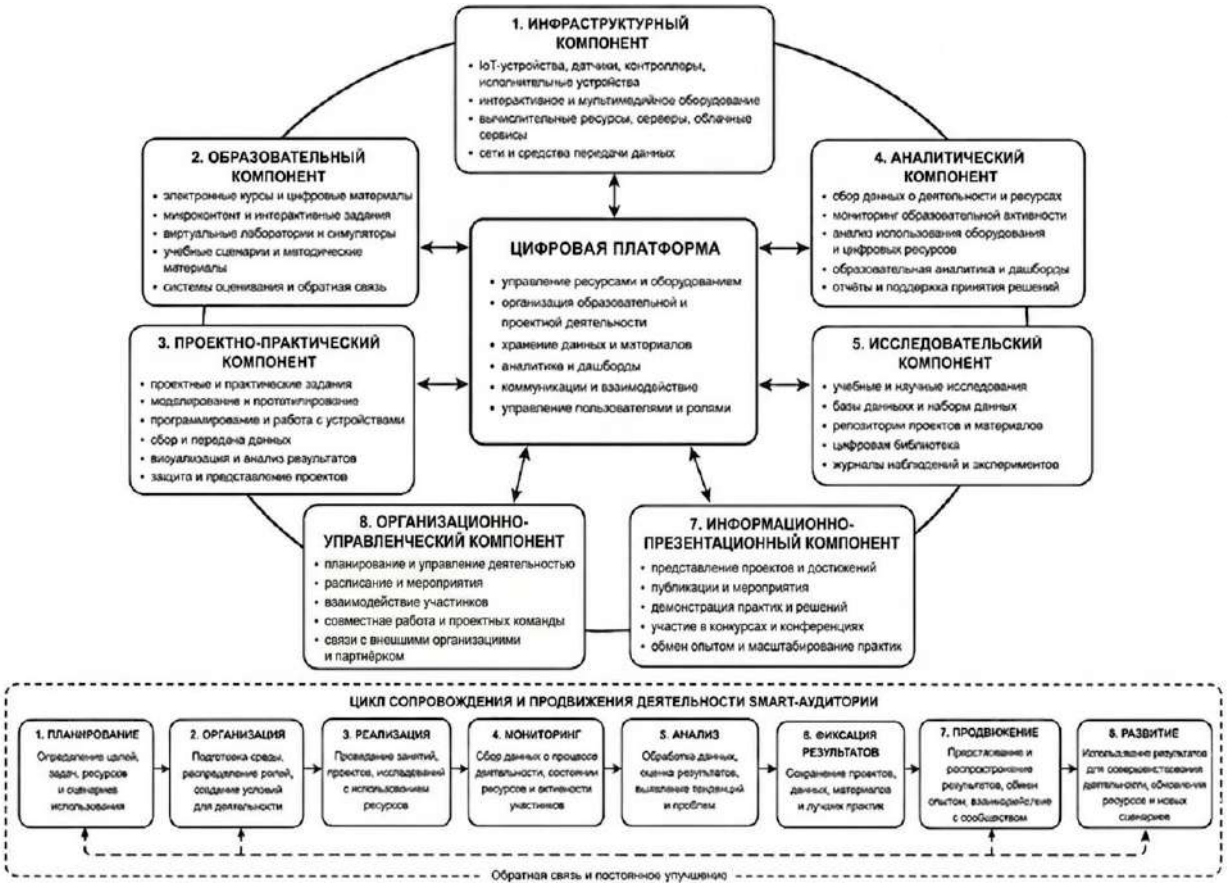


Рисунок 1. Концептуальная структурно-функциональная модель SMART-аудитории

На основе определенных функций сформирована функциональная структура цифровой платформы, включающая семь основных разделов: «SMART-аудитория», «Обучение», «Проекты», «Исследования», «Аналитика», «Сообщество», «Результаты и медиа». Эти разделы соответствуют основным направлениям деятельности SMART-аудитории, но не предполагают линейной последовательности работы. Они образуют взаимосвязанную систему, в которой между разделами осуществляется обмен данными и результатами. Например, при выполнении проекта студент получает задание и необходимые материалы в разделе «Обучение», использует оборудование и данные через раздел «SMART-аудитория», фиксирует этапы и результаты работы в разделе «Проекты», сохраняет данные и материалы эксперимента в разделе «Исследования», а результаты и показатели работы анализируются в разделе «Аналитика». Итоговая разработка может быть представлена в разделе «Результаты и медиа». Такой сценарий демонстрирует функциональную связь разделов и необходимость их объединения в единой цифровой среде.

Раздел «SMART-аудитория» обеспечивает цифровое представление физической инфраструктуры. В нем размещается информация об оборудовании, IoT-устройствах, подключенных сенсорах, доступных ресурсах и состоянии образовательной среды. При наличии соответствующего оборудования и сетевого подключения данные о состоянии среды и устройств могут передаваться в аналитические сервисы.

Раздел «Обучение» объединяет LMS и специализированные образовательные сервисы. LMS обеспечивает размещение электронных курсов, организацию занятий, выдачу заданий, оценивание и отслеживание прогресса. Дополнительные сервисы включают микроконтент, виртуальные лаборатории, симуляторы, тренажеры и интерактивные задания. Таким образом, LMS выступает образовательной подсистемой единой платформы.

Раздел «Проекты» предназначен для сопровождения проектной деятельности от постановки задачи до представления результата. Он включает активные проекты, команды, этапы выполнения, задания, результаты и электронные портфолио.

Раздел «Исследования» объединяет исследовательские проекты, наборы данных, экспериментальные материалы, публикации и электронную библиотеку. Он обеспечивает сохранение и повторное использование результатов в образовательной и проектной деятельности.

Раздел «Аналитика» объединяет данные об образовательной активности, использовании оборудования, IoT-устройств и проектной деятельности. Их совместный анализ позволяет отслеживать деятельность SMART-аудитории и поддерживать принятие педагогических и организационных решений.

Раздел «Сообщество» обеспечивает взаимодействие преподавателей, студентов, исследователей, представителей школ и работодателей. Здесь могут выполняться совместные проекты, проводиться консультации, вебинары, обсуждения и профессиональные мероприятия.

Раздел «Результаты и медиа» предназначен для представления проектов, цифровых разработок, исследовательских результатов, публикаций, мероприятий и успешных педагогических практик. Он обеспечивает представление результатов деятельности SMART-аудитории во внешнем профессиональном пространстве.

Таким образом, функциональная структура платформы непосредственно отражает структуру деятельности SMART-аудитории. Раздел «SMART-аудитория» связывает цифровую среду с физической инфраструктурой, «Обучение» поддерживает образовательный процесс, «Проекты» и «Исследования» – практическую и научную деятельность, «Аналитика» – работу с данными, «Сообщество» – взаимодействие участников, а «Результаты и медиа» – представление и распространение результатов. Их объединение формирует единую цифровую среду организации и сопровождения деятельности SMART-аудитории.

На основе функциональной структуры определена информационная архитектура платформы, объединяющая в едином цифровом пространстве физические ресурсы SMART-аудитории, образовательные сервисы, проектную и исследовательскую деятельность, аналитические инструменты и средства коммуникации. Архитектура предусматривает обмен данными между разделами. Например, данные IoT-устройств могут одновременно использоваться в учебных и проектных заданиях и поступать в аналитические сервисы. Результаты проекта могут сохраняться в электронном портфолио, передаваться в исследовательский репозиторий и после завершения представляться в разделе «Результаты и медиа». Таким образом, архитектура представляет собой не линейную последовательность разделов, а взаимосвязанную систему функциональных модулей, объединенных общими данными и информационными потоками (рис. 2).

Такое построение архитектуры отражает основную идею разработанной модели. Цифровая платформа не заменяет физическую SMART-аудиторию, а дополняет ее цифровым представлением, поддерживает организацию процессов, работу с данными и взаимодействие участников.

Для практической реализации функциональной структуры предусмотрено ролевое разграничение доступа к платформе. При единой архитектуре набор доступных функций и содержание интерфейса определяются ролью пользователя. Студент получает доступ к курсам, заданиям, проектам, собственным результатам, аналитике учебной деятельности и электронному портфолио. Преподаватель работает с образовательными материалами, проектами, ресурсами SMART-аудитории, аналитическими данными и средствами сопровождения обучающихся. Администратор управляет пользователями, оборудованием, расписанием, ресурсами, правами доступа и сводной аналитикой. Такое распределение сохраняет единую структуру платформы и одновременно обеспечивает каждому участнику доступ к функциям, соответствующим его роли в деятельности SMART-аудитории.

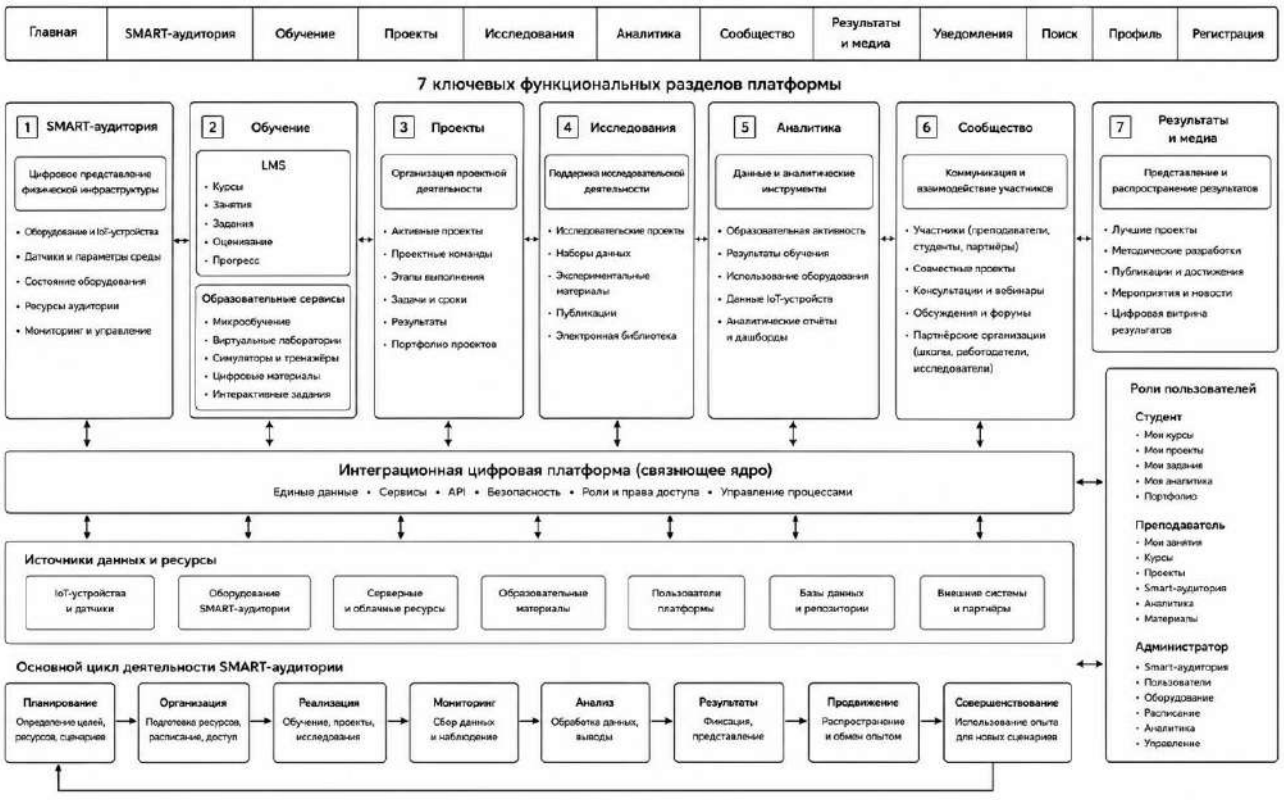


Рисунок 2. Информационная архитектура цифровой платформы сопровождения деятельности SMART-аудитории

Главная страница платформы при этом должна отражать не только ее структуру, но и текущее состояние деятельности SMART-аудитории, включая состояние инфраструктуры, образовательные и проектные процессы, основные аналитические показатели, исследовательскую активность и последние результаты.

Логическая целостность модели была проверена на типовом сценарии разработки обучающимися системы мониторинга параметров SMART-аудитории. В ходе сценария оценивалась возможность последовательного использования ресурсов различных компонентов модели и соответствующих функций цифровой платформы – от доступа к IoT-оборудованию и учебным материалам до передачи и анализа данных, фиксации и представления результата. Проверка показала, что выделенные компоненты не являются изолированными блоками. Один процесс может одновременно задействовать ресурсы нескольких компонентов, а его результаты переходят в другие направления деятельности. Так, данные, полученные в ходе выполнения проекта, используются для анализа, результаты сохраняются в электронном портфолио и исследовательском репозитории, а готовая разработка может быть представлена профессиональному сообществу. Сценарий подтвердил, что предложенная модель позволяет описать полный цикл деятельности SMART-аудитории – от организации образовательной или проектной работы до анализа, фиксации и представления результатов – и связать его с функциями цифровой платформы. В целом результаты исследования позволяют представить SMART-аудиторию как целостную физико-цифровую образовательную систему, в которой физическая инфраструктура и цифровая платформа выполняют взаимодополняющие функции. Выделение семи взаимосвязанных компонентов позволило систематизировать ее структуру и основные направления деятельности, определить процессы, ресурсы и информационные потоки, требующие цифровой поддержки. На этой основе разработаны концептуальная модель, функциональная структура и информационная архитектура цифровой платформы, обеспечивающей интеграцию ресурсов, данных, участников и процессов в едином цифровом пространстве.

Дискуссия

Полученные результаты позволяют уточнить, при каких условиях SMART-аудитория может быть организована как интеллектуальная образовательная среда в системе педагогического образования. Современные исследования подтверждают, что интеграция интеллектуальных технологий, IoT, сетевой инфраструктуры, средств анализа данных и интерактивного оборудования может положительно влиять на результаты обучения. В частности, метаанализ Cao и др. [11], включивший 43 экспериментальных и квазиэкспериментальных исследования, выявил статистически значимый положительный эффект интеллектуальной среды обучения на когнитивные, аффективные и психомоторные результаты. Вместе с тем высокая неоднородность полученных результатов показывает, что эффективность таких сред определяется не только используемыми технологиями, но и условиями их реализации, педагогическими подходами и особенностями образовательного контекста. Поэтому в данном исследовании акцент сделан не на расширении технологического оснащения SMART-аудитории, а на организации взаимосвязанной системы ее деятельности.

Такой подход согласуется с исследованиями, в которых интеллектуальная среда обучения рассматривается как среда, способная учитывать контекст обучения и состояние взаимодействующих с ней объектов и участников. Hwang [2] связывает интеллектуальность образовательной среды с ее контекстной осведомленностью, то есть способностью получать информацию о состоянии обучающегося и окружающего пространства и использовать ее для поддержки обучения. Систематический обзор Bautista и López-Costa [6] также показывает необходимость одновременного учета педагогического, физического и цифрового компонентов интеллектуальной среды обучения. Разработанная в данном исследовании модель развивает этот подход за счет включения организационно-управленческого компонента и представления SMART-аудитории как системы, в которой технологии, участники, ресурсы и данные связаны с образовательной, проектной, исследовательской и коммуникационной деятельностью. Это позволяет сместить акцент с вопроса о том, какие технологии установлены в аудитории, на вопрос о том, каким образом они включены в реальные процессы и обеспечивают достижение образовательных и организационных результатов.

Для педагогического образования такое расширение имеет особое значение. Будущему педагогу недостаточно уметь пользоваться отдельными цифровыми инструментами; ему необходимо понимать, как технологии, данные, физическое пространство и педагогические сценарии могут использоваться совместно для решения образовательных задач. Поэтому SMART-аудитория может одновременно выступать средой освоения технологий и пространством формирования опыта их проектирования, практического применения и педагогической оценки. Подобный комплексный подход отражен и в обзоре Quevedo Piratova и соавторов [12], где при проектировании интеллектуальных аудиторий учитываются физические параметры пространства, технологии, взаимодействие участников и педагогические подходы. В предлагаемой модели эта идея дополнена проектно-практическим и исследовательским компонентами, благодаря чему обучающиеся получают возможность не только использовать готовые технологии, но и разрабатывать, апробировать и анализировать собственные цифровые решения.

Необходимость целостного подхода подтверждается и систематическим обзором Alfoudari A.M. и др. [13], охватившим 105 исследований SMART-аудиторий. Авторы выделяют среди сохраняющихся проблем интеграцию интеллектуальных систем и аналитических инструментов, проектирование образовательной среды и согласование технологических и социальных компонентов. Полученные в данном исследовании результаты показывают один из возможных способов решения этой проблемы – связывать технологические ресурсы с конкретными видами деятельности, участниками, ресурсами, данными и ожидаемыми результатами. В таком случае IoT, аналитика, образовательные

сервисы и другие технологии становятся частью общей системы, а не набором независимо функционирующих решений.

Именно в этом контексте цифровая платформа приобретает значение цифрового уровня SMART-аудитории. В предложенной модели она не выделяется в качестве отдельного структурного компонента наряду с образовательным, проектным или исследовательским компонентами, а выполняет интегрирующую функцию. Платформа связывает физическую инфраструктуру с цифровыми ресурсами, данными, участниками и процессами деятельности. Такое представление позволяет избежать понимания цифровизации как простого добавления набора сервисов к уже существующей инфраструктуре. Цифровое сопровождение в данном случае строится вокруг реальной деятельности SMART-аудитории и обеспечивает ее представление, организацию и взаимосвязь в цифровой среде. Это соответствует пониманию интеллектуальной среды обучения как системы, в которой физическое, цифровое и педагогическое измерения должны рассматриваться во взаимосвязи [6].

Отсюда следует важный принцип формирования функциональной структуры платформы. Ее состав определяется не перечнем доступных технологий, а процессами, ресурсами и информационными потоками SMART-аудитории, которые требуют цифровой поддержки. Поэтому разделы «SMART-аудитория», «Обучение», «Проекты», «Исследования», «Аналитика», «Сообщество» и «Результаты и медиа» отражают основные направления деятельности и соответствующие им функции цифрового сопровождения. При этом они не являются изолированными модулями. Между ними осуществляется обмен данными и результатами. Например, данные IoT могут использоваться в учебных и проектных заданиях, результаты проектов – сохраняться в исследовательских репозиториях и портфолио, а накопленные данные – использоваться для аналитики. Такой принцип позволяет сформировать единую цифровую среду, соответствующую логике деятельности SMART-аудитории.

Особое место в этой системе занимает аналитика. SMART-аудитория формирует данные не только об образовательной активности, но и об использовании оборудования, состоянии физической среды, проектной и исследовательской работе. Это соответствует современному развитию learning analytics, где данные используются для изучения активности пользователей цифровых платформ и результатов обучения [14]. Однако сам по себе сбор данных не гарантирует улучшения образовательного процесса. Их практическая ценность возникает тогда, когда они позволяют отвечать на конкретные педагогические и организационные вопросы – где возникают затруднения, какие ресурсы используются наиболее активно, каким проектам требуется дополнительная помощь и какие образовательные сценарии дают лучшие результаты. Поэтому в предлагаемой модели аналитика включена в общий цикл управления деятельностью планирование → реализация → сбор данных → анализ → принятие решений → совершенствование.

Этот принцип имеет значение и для использования искусственного интеллекта. Исследования ИИ в SMART-аудиториях показывают, что внимание пока во многом сосредоточено на технологических возможностях и роль преподавателя остается важным условием эффективного применения ИИ [15,16]. В предложенной модели интеллектуальные технологии поэтому не рассматриваются как замена преподавателя. Их использование связано прежде всего с обработкой и интерпретацией данных, выявлением значимых изменений и поддержкой педагогических и организационных решений. Такой подход позволяет рассматривать ИИ как один из инструментов цифрового сопровождения деятельности SMART-аудитории, сохраняя ведущую роль преподавателя в интерпретации результатов и принятии педагогических решений.

Еще одной особенностью предлагаемого подхода является включение представления и распространения результатов в общий цикл деятельности SMART-аудитории. Проекты обучающихся, исследовательские результаты, методические разработки и успешные образовательные практики могут использоваться не только внутри аудитории, но и

представляться другим подразделениям университета, школам, работодателям и профессиональному сообществу. Это создает возможности для обмена опытом, получения обратной связи и дальнейшего развития разработанных решений. Таким образом, деятельность SMART-аудитории не завершается получением результата, а предполагает его фиксацию, представление и последующее использование в образовательной, проектной и исследовательской деятельности.

В совокупности это позволяет рассматривать SMART-аудиторию как открытую физико-цифровую образовательную систему, эффективность которой определяется не количеством установленных устройств, а тем, насколько последовательно инфраструктура включена в реальные образовательные, проектные и исследовательские процессы. Такой вывод согласуется с пониманием SMART-аудиторий как социотехнических систем, объединяющих технологические, педагогические и организационные аспекты [13]. Вклад данного исследования состоит в конкретизации этой взаимосвязи на уровне организации деятельности SMART-аудитории и ее цифрового сопровождения – от выделения компонентов и процессов до определения функциональной структуры и информационной архитектуры платформы.

При этом разработанная модель имеет концептуальный характер. Ее проверка осуществлялась посредством анализа структуры, функциональных связей и типовых сценариев использования, поэтому на данном этапе нельзя утверждать, что предложенная архитектура непосредственно приводит к повышению качества обучения или эффективности управления SMART-аудиторией. Для получения эмпирических доказательств необходимы разработка прототипа платформы, интеграция с реальными IoT-устройствами и апробация в образовательном процессе. Перспективными направлениями также являются подключение LMS и аналитических сервисов, исследование вопросов информационной безопасности и защиты образовательных данных, а также оценка влияния платформы на деятельность преподавателей и обучающихся.

Таким образом, результаты исследования показывают, что проектирование цифровой платформы для SMART-аудитории целесообразно начинать не с выбора отдельных технологий, а с анализа ее инфраструктуры, участников, видов деятельности, ресурсов, данных и взаимосвязей. Разработанная концептуальная модель связывает физическую инфраструктуру и цифровую платформу в единую систему и позволяет последовательно представить, каким образом деятельность SMART-аудитории должна поддерживаться в цифровой среде. Это создает основу для дальнейшей разработки и практической апробации платформы, а также для эмпирической оценки эффективности предложенного подхода.

Заключение

Проведенное исследование показывает, что развитие SMART-аудитории в педагогическом вузе связано не столько с расширением технического оснащения, сколько с согласованной организацией его использования в образовательной, проектной и исследовательской деятельности. Поэтому SMART-аудиторию целесообразно рассматривать как целостную физико-цифровую образовательную систему, объединяющую физическую инфраструктуру и единую цифровую платформу, в рамках которых взаимосвязаны участники, ресурсы, данные и процессы деятельности.

Разработанная концептуальная модель систематизирует структуру и основные направления деятельности SMART-аудитории, устанавливает взаимосвязи между ее компонентами и отражает взаимодействие физического и цифрового уровней. На этой основе определены требования к цифровому сопровождению, сформированы функциональная структура и информационная архитектура единой цифровой платформы. Платформа при этом рассматривается не как самостоятельный компонент структуры SMART-аудитории, а как ее цифровый уровень, обеспечивающий представление и интеграцию физических и цифровых ресурсов, данных, процессов и взаимодействие участников. Такой подход позволяет связать

организацию деятельности SMART-аудитории с ее цифровым сопровождением и избежать представления платформы как набора разрозненных сервисов.

Для педагогического образования предложенный подход имеет особое значение, поскольку SMART-аудитория может выступать не только средой обучения, но и практическим пространством профессиональной подготовки будущих педагогов. В ней обучающиеся получают возможность осваивать цифровые и интеллектуальные технологии, разрабатывать и апробировать технологические решения, работать с данными, реализовывать проекты и оценивать педагогический потенциал цифровых инструментов. Это создает условия для формирования целостного практического опыта организации образовательной деятельности в физико-цифровой интеллектуальной среде.

Таким образом, предложенная модель задает целостное представление об организации и цифровом сопровождении деятельности SMART-аудитории и обеспечивает основу для ее дальнейшей практической реализации. Следующим этапом исследования станет разработка прототипа цифровой платформы, интеграция с IoT-инфраструктурой и апробация в реальном образовательном процессе. Это позволит проверить практическую применимость предложенной модели, уточнить функциональные требования и оценить влияние цифрового сопровождения на организацию образовательной, проектной и исследовательской деятельности SMART-аудитории.

Благодарность

Исследование проведено в рамках финансирования Комитетом науки Министерства науки и высшего образования Республики Казахстан (грант №AP23490844 «IoT-трансформация учебного заведения: проектирование, внедрение Smart аудитории и подготовка специалистов в области IoT для образования»).

Список использованных источников

[1]Spector, J.M. Conceptualizing the emerging field of SMART learning environments. SMART Learning Environments. 1, 2. – 2014. https://doi.org/10.1186/s40561-014-0002-7

[2]Hwang, GJ. Definition, framework and research issues of SMART learning environments - a context-aware ubiquitous learning perspective. SMART Learning Environment. 1, 4. – 2014. https://doi.org/10.1186/s40561-014-0004-5

[3]Koper, R. Conditions for effective SMART learning environments. SMART Learning Environment. 1, 5. – 2014. https://doi.org/10.1186/s40561-014-0005-4

[4]Kaur, A., Bhatia, M., & Stea, G. A Survey of SMART Classroom Literature. Education Sciences, 12(2), 86. – 2022. https://doi.org/10.3390/educsci12020086

[5]Tabuenca B.et al. "Affordances and Core Functions of SMART Learning Environments: A Systematic Literature Review," in IEEE Transactions on Learning Technologies, vol. 14, no. 2. – 2021. – pp. 129-145, doi: https://doi.org/10.1109/TLT.2021.3067946

[6]Bautista, G., López-Costa, M. SMART learning spaces considering the integration of the pedagogical, environmental and digital dimensions: a systematic review. Learning Environments Restarch 28. – 2025. – 455-472. https://doi.org/10.1007/s10984-025-09551-2 (дата обращения: 12.08.2026).

[7]Valentina Terzieva, Svetozar Ilchev, Katia Todorova. The Role of Internet of Things in SMART Education, IFAC-PapersOnLine, Volume 55, Issue 11. – 2022. – Pages 108-113, https://doi.org/10.1016/j.ifacol.2022.08.057

[8]Afzal Badshah, Anwar Ghani, Ali Daud, Ateeqa Jalal, Muhammad Bilal, Jon Crowcroft. Towards Smart Education through the Internet of Things: A Review. – 2023 – https://doi.org/10.48550/arXiv.2304.12851 (дата обращения: 12.08.2026).

[9]Elmoazen, R., Saqr, M., Khalil, M. et al. Learning analytics in virtual laboratories: a systematic literature review of empirical research. Smart Learning Environment. 10, 23. – 2023. https://doi.org/10.1186/s40561-023-00244-y (дата обращения: 12.08.2026).

[10]Tao, L., Cukurova, M. & Song, Y. Learning analytics in immersive virtual learning environments: a systematic literature review. Smart Learning Environment. 12, 43. – 2025. https://doi.org/10.1186/s40561-025-00381-6

[11]Cao J, A. Rashid RAB, Zhu H, Lyu B, Huang Y and Li X. The effects of smart learning environments on learning outcomes: a meta-analysis of 43 experimental and quasi-experimental studies. Frontiers in Education. – 2026. – https://doi.org/10.3389/feduc.2026.1829962

[12]Quevedo Piratova, D.A., Sánchez Duarte, M.M., Chiappe, A. and Sáez Delgado, F.M. Intelligent Classrooms: How AI and IoT Can Reshape Learning Spaces. European Journal of Education. – 2025. – https://doi.org/10.1111/ejed.70000

[13]Alfoudari Aisha M., Durugbo Christopher M., Aldhmour, Fairouz M. Understanding socio-technological challenges of smart classrooms using a systematic review, Computers & Education, Volume 173. – 2021. – https://doi.org/10.1016/j.compedu.2021.104282

[14]Hernández-Campos, M., Gonzalez-Torres, A., & García-Peñalvo, F. J. Learning Outcomes Evaluation Through Learning Analytics Systems in Higher Education: A Systematic Literature Review. Sage Open, 15(3). – 2025. – https://doi.org/10.1177/21582440251347374

[15]Llurba, C., Fretes, G., Martínez-Ballesté, A., & Palau, R. Artificial Intelligence in Smart Classrooms: A Systematic Literature Review of Applications, Dimensions, and Teacher Roles. Encyclopedia, 6(6), 138. – 2026. – https://doi.org/10.3390/encyclopedia6060138

[16]Dimitriadou, E., Lanitis, A. A critical evaluation, challenges, and future perspectives of using artificial intelligence and emerging technologies in smart classrooms. Smart Learning Environment. 10, 12. – 2023. – https://doi.org/10.1186/s40561-023-00231-3

МРНТИ 28.23.29

https://doi.org/10.51889/2959-5894.2026.95.3.025

Д.С. Найманова[1*], А.З. Даутова[1], Ж.С. Алимова[1], Н.Н. Оспанова[1]

[1] Торайгыров университет, г.Павлодар, Қазақстан
[*]*e-mail: dinarasakenkyzy83@gmail.com*

РАЗРАБОТКА ИНТЕЛЛЕКТУАЛЬНОЙ АНАЛИТИЧЕСКОЙ СИСТЕМЫ ДЛЯ АНАЛИЗА ОБРАЗОВАТЕЛЬНЫХ ПРОГРАММ НА БАЗЕ LLM

Аннотация

В статье рассматривается разработка интеллектуальной аналитической системы для анализа образовательных программ, предназначенной для автоматизации оценки содержания образовательных документов с использованием технологий искусственного интеллекта. В отличие от существующих решений, ориентированных преимущественно на локальную обработку отдельных документов, предложенная система реализована в виде веб-приложения с клиент-серверной архитектурой, включающей frontend, backend, реляционную базу данных, модуль обработки PDF-документов и модуль искусственного интеллекта. Представлена логика функционирования системы, описывающая взаимодействие пользователей различных ролей (преподавателя, студента и администратора), организацию хранения данных и маршрут их обработки внутри системы. Для интеллектуального анализа содержания образовательных программ используются методы обработки естественного языка (NLP) и большие языковые модели (LLM), обеспечивающие автоматическое выявление универсальных компетенций (soft skills) и формирование рекомендаций по совершенствованию образовательных программ. Результаты анализа сохраняются в базе данных и предоставляются пользователям через веб-интерфейс. Разработанная интеллектуальная аналитическая система обеспечивает автоматизацию анализа образовательных программ, сокращает трудоемкость экспертной оценки, повышает объективность анализа и может быть использована образовательными организациями при мониторинге и совершенствовании образовательных программ в соответствии с современными требованиями рынка труда.

Ключевые слова: анализ образовательных программ; искусственный интеллект; обработка естественного языка (NLP); большие языковые модели (LLM); автоматизация анализа документов; клиент-серверная архитектура; реляционная база данных; образовательные технологии; цифровизация образования.

Д.С. Найманова[1], А.З. Даутова[1], Ж.С. Алимова[1], Н.Н. Оспанова[1]
[1] Торайғыров университеті, Павлодар қ., Қазақстан.

БІЛІМ БЕРУ БАҒДАРЛАМАЛАРЫН ТАЛДАУДЫҢ ВЕБ-ЖҮЙЕСІНІҢ ЖҰМЫС ЛОГИКАСЫ

Аңдатпа

Мақалада жасанды интеллект технологияларын пайдалана отырып, білім беру құжаттарының мазмұнын бағалауды автоматтандыруға арналған білім беру бағдарламаларын талдаудың веб-жүйесін әзірлеу қарастырылады. Жекелеген құжаттарды жергілікті өңдеуге бағытталған қолданыстағы шешімдерден айырмашылығы, ұсынылған жүйе frontend, backend, реляциялық деректер базасы, PDF құжаттарын өңдеу модулі және жасанды интеллект модулінен тұратын клиент–серверлік архитектура негізінде веб-қосымша ретінде жүзеге асырылған. Әртүрлі рөлдегі пайдаланушылардың (оқытушы, студент және әкімші) өзара әрекеттесуін, деректерді сақтау құрылымын және жүйе ішіндегі өңдеу маршрутын сипаттайтын жүйенің жұмыс логикасы ұсынылған. Білім беру бағдарламаларының мазмұнын интеллектуалды талдау үшін табиғи тілді өңдеу (Natural Language Processing, NLP) әдістері мен үлкен тілдік модельдер (Large Language Models, LLM) қолданылады, бұл әмбебап құзыреттерді (soft skills) автоматты түрде анықтауға және білім беру бағдарламаларын жетілдіру бойынша ұсынымдар қалыптастыруға мүмкіндік береді. Талдау нәтижелері деректер базасында сақталып, веб-интерфейс арқылы пайдаланушыларға қолжетімді болады. Әзірленген веб-жүйе білім беру бағдарламаларын талдау үдерісін автоматтандыруды қамтамасыз етеді, сараптамалық бағалаудың

еңбек сыйымдылығын төмендетеді, талдаудың объективтілігін арттырады және білім беру ұйымдарында еңбек нарығының заманауи талаптарына сәйкес білім беру бағдарламаларын мониторингтеу және жетілдіру мақсатында қолданылуы мүмкін.

Түйін сөздер: білім беру бағдарламаларын талдау; жасанды интеллект; табиғи тілді өңдеу (NLP); үлкен тілдік модельдер (LLM); құжаттарды автоматтандырылған талдау; клиент–серверлік архитектура; реляциялық деректер базасы; білім беру технологиялары; білім беруді цифрландыру

D. Naimanova[1*], A. Dautova[1], Zh. Alimova[1], N. Ospanova[1]
[1] Toraighyrov University, Pavlodar, Kazakhstan.

OPERATIONAL LOGIC OF THE EDUCATIONAL PROGRAM ANALYSIS WEB SYSTEM

Abstract

This paper presents the development of a web-based system for educational program analysis designed to automate the assessment of educational document content using artificial intelligence technologies. Unlike existing solutions that primarily focus on local processing of individual documents, the proposed system is implemented as a web application based on a client–server architecture comprising a frontend, backend, relational database, PDF document processing module, and an artificial intelligence module. The paper describes the system's operational logic, including interactions among users with different roles (instructors, students, and administrators), data storage organization, and the internal data processing workflow. Natural Language Processing (NLP) techniques and Large Language Models (LLMs) are employed to perform intelligent analysis of educational program content, enabling the automatic identification of soft skills and the generation of recommendations for improving educational programs. The analysis results are stored in a relational database and made available to users through a web interface. The developed web-based system automates the analysis of educational programs, reduces the workload associated with expert evaluation, improves the objectivity of the analysis, and can be applied by higher education institutions to monitor and enhance educational programs in accordance with current labor market requirements.

Keywords: educational program analysis; artificial intelligence; natural language processing (NLP); large language models (LLMs); automated document analysis; client–server architecture; relational database; educational technologies; digitalization of education.

Введение

В условиях цифровой трансформации высшего образования особую актуальность приобретает обеспечение соответствия образовательных программ современным требованиям рынка труда. Быстрое обновление профессиональных компетенций, появление новых технологий и изменение запросов работодателей требуют регулярного анализа содержания образовательных программ и своевременного внесения изменений в их структуру. Наряду с профессиональными знаниями все большее значение приобретают универсальные компетенции (soft skills), формирование которых рассматривается как один из ключевых факторов успешной профессиональной деятельности выпускников.

В Республике Казахстан развитие цифровых технологий в сфере образования является одним из приоритетных направлений государственной политики. Концепция развития высшего образования и науки Республики Казахстан на 2023–2029 годы предусматривает цифровую трансформацию образовательного процесса, внедрение технологий искусственного интеллекта, развитие цифровой инфраструктуры вузов и использование интеллектуальных информационных систем для повышения качества подготовки кадров. Дополнительным фактором является постоянное обновление требований рынка труда, отраженных в профессиональных стандартах и аналитических материалах, включая «Атлас новых профессий и компетенций Казахстана», которые подчеркивают возрастающую значимость цифровых и универсальных компетенций выпускников.

Традиционный анализ образовательных программ осуществляется экспертами вручную, что требует значительных временных затрат и во многом зависит от субъективной оценки специалистов. При большом количестве образовательных программ и постоянном обновлении профессиональных стандартов такой подход становится недостаточно эффективным. В связи

с этим возрастает необходимость создания интеллектуальных информационных систем, способных автоматически анализировать содержание образовательных документов, выявлять соответствие компетенций требованиям работодателей и предоставлять преподавателям рекомендации по совершенствованию образовательных программ.

Развитие технологий обработки естественного языка (Natural Language Processing, NLP) и больших языковых моделей (Large Language Models, LLM) открывает новые возможности для автоматизированного анализа текстовых документов. Использование подобных технологий позволяет анализировать не только наличие отдельных терминов, но и смысловое содержание образовательных программ, оценивать полноту представления компетенций и формировать рекомендации по их совершенствованию.

Научная новизна исследования заключается в разработке архитектуры интеллектуальной аналитической системы для анализа образовательных программ, обеспечивающей полный цикл обработки данных в многопользовательской среде. В отличие от существующих решений, предложенная система реализует интеграцию клиент-серверной архитектуры, реляционной базы данных, модулей обработки документов и больших языковых моделей с механизмом динамического управления промптами, автоматически адаптирующего запросы к LLM в зависимости от роли пользователя, типа документа и этапа анализа. Это обеспечивает более гибкое, масштабируемое и объективное выполнение интеллектуального анализа образовательных программ.

Цифровизация высшего образования обусловила активное развитие исследований, посвященных интеллектуальному анализу образовательных данных и применению технологий искусственного интеллекта для поддержки принятия решений в образовательной среде. Одними из наиболее интенсивно развивающихся направлений являются *Educational Data Mining (EDM) и Learning Analytics (LA***)**. Согласно обзору Romero и Ventura [1], Educational Data Mining ориентирован на извлечение знаний из образовательных данных с использованием методов интеллектуального анализа, машинного обучения и статистики, тогда как Learning Analytics направлен на сбор, анализ и интерпретацию данных о процессе обучения для повышения его эффективности и качества образовательного процесса. Современные исследования демонстрируют постепенный переход от использования традиционных статистических методов к интеллектуальным прогностическим моделям, основанным на технологиях искусственного интеллекта, обеспечивающим персонализацию обучения и поддержку принятия управленческих решений в образовательных организациях [2].

Развитие методов обработки естественного языка (Natural Language Processing, NLP) и больших языковых моделей (Large Language Models, LLM) значительно расширило возможности автоматизированного анализа текстовых образовательных документов. Современные модели способны выполнять не только поиск ключевых слов, но и анализировать смысловое содержание текстов, выявлять компетенции, определять семантическое соответствие образовательных программ требованиям профессиональных стандартов и рынка труда, а также формировать рекомендации по их совершенствованию. В работе Xu и соавт. [3] показано, что применение архитектуры Retrieval-Augmented Generation (RAG) при анализе учебных программ обеспечивает более высокую точность извлечения профессиональных компетенций по сравнению с традиционными подходами, основанными на стандартных промптах (zero-shot prompting). Аналитические исследования компании Mapademics [4] также подтверждают необходимость автоматизации анализа образовательных программ, указывая, что к 2030 году около 39 % востребованных профессиональных навыков изменятся, что требует регулярного обновления содержания образовательных программ.

Особое значение применение больших языковых моделей приобретает в условиях многоязычной образовательной среды Республики Казахстан. В исследовании Togmanov и соавт. [5] представлен первый специализированный бенчмарк *KazMMLU*, предназначенный для оценки качества работы языковых моделей на казахском, русском и английском языках.

Полученные результаты подтверждают перспективность использования LLM при решении прикладных задач анализа образовательных документов и обработки текстовой информации в национальном образовательном пространстве.

Актуальность разработки интеллектуальных образовательных систем подтверждается и государственными инициативами Республики Казахстан в области цифровой трансформации. В соответствии с государственной программой *Digital Qazaqstan* цифровизация образования рассматривается как одно из приоритетных направлений развития страны, предусматривающее внедрение технологий искусственного интеллекта, развитие интеллектуальных образовательных сервисов и использование аналитических систем для повышения качества подготовки кадров [6]. Практическое применение методов Learning Analytics в казахстанских университетах рассмотрено в исследовании Zhumazhanova и соавт. [7], результаты которого демонстрируют положительное влияние аналитики образовательных данных на повышение вовлеченности студентов и эффективность образовательного процесса. Вместе с тем Klevtsov [8] отмечает, что большинство вузов Казахстана пока не готовы к полномасштабному внедрению технологий искусственного интеллекта вследствие отсутствия современных цифровых регламентов и недостаточной интеграции интеллектуальных сервисов в существующую образовательную инфраструктуру.

Анализ существующих исследований показывает, что большинство разработанных решений ориентировано либо на анализ образовательных данных студентов, либо на локальную обработку отдельных документов. При этом недостаточное внимание уделяется построению комплексных программ, обеспечивающих полный цикл обработки данных, включая управление пользователями, централизованное хранение результатов анализа, маршрутизацию данных между компонентами системы и интеграцию модулей искусственного интеллекта в единую информационную среду. Кроме того, существующие решения, как правило, используют статические сценарии взаимодействия с языковыми моделями и не учитывают различия между ролями пользователей и этапами обработки данных.

В отличие от существующих подходов, в настоящем исследовании рассматривается архитектура и разработка интеллектуальной аналитической системы для анализа образовательных программ, реализующая полный цикл интеллектуальной обработки данных в многопользовательской среде. Система объединяет клиент-серверную архитектуру, централизованную реляционную базу данных, модуль обработки PDF-документов, динамическое управление промптами для больших языковых моделей и механизмы ролевого доступа пользователей, что обеспечивает масштабируемость, воспроизводимость результатов анализа и возможность практического внедрения в образовательных организациях.

Анализ существующих исследований показывает, что большинство разработанных решений представляет собой отдельные программные модули или исследовательские прототипы, ориентированные исключительно на выполнение анализа текста образовательной программы. При этом недостаточно рассматриваются вопросы построения полноценной веб-системы, обеспечивающей совместную работу различных категорий пользователей, централизованное хранение данных, разграничение прав доступа и автоматизированное взаимодействие между компонентами системы.

Отсутствие комплексного подхода к организации серверной логики, базы данных, механизмов обработки документов и интеграции модулей искусственного интеллекта затрудняет практическое внедрение подобных решений в образовательных организациях. В связи с этим возникает необходимость разработки интеллектуальной системы, объединяющей процессы загрузки образовательных программ, интеллектуального анализа документов и предоставления результатов различным категориям пользователей в рамках единой архитектуры.

Целью исследования является разработка интеллектуальной аналитической системы для анализа образовательных программ, обеспечивающей автоматизированную обработку

документов, интеллектуальный анализ содержания образовательных программ с использованием технологий искусственного интеллекта и организацию взаимодействия пользователей в рамках единой клиент-серверной архитектуры.

Для достижения поставленной цели были решены следующие задачи:

– разработана интеллектуальная аналитическая система для анализа образовательных программ;

– описана логика работы интеллектуальной аналитической системы: логика работы преподавателей, студента, администратора; база данных и связи между сущностями; маршрут данных внутри системы.

Методология исследования

Методология исследования основана на применении методов системного анализа, объектно-ориентированного проектирования, разработки веб-приложений, технологий обработки естественного языка (Natural Language Processing, NLP) и больших языковых моделей (Large Language Models, LLM). При проектировании системы использовались принципы клиент-серверной архитектуры, обеспечивающие разделение пользовательского интерфейса, серверной логики, хранения данных и интеллектуальной обработки документов.

В процессе исследования разработана интеллектуальная аналитическая система для анализа образовательных программ, реализованная на основе многослойной клиент-серверной архитектуры. Система включает frontend-интерфейс, backend для обработки пользовательских запросов, реляционную базу данных, модуль обработки PDF-документов и модуль искусственного интеллекта, выполняющий интеллектуальный анализ содержания образовательных программ. Организация обработки данных обеспечивает централизованное хранение информации, поддержку многопользовательского режима работы, ролевое разграничение доступа и сохранение состояния объектов системы на всех этапах анализа.

Предлагаемый метод исследования основан на многоэтапной обработке образовательных программ с использованием технологий NLP и LLM. Исходные документы в формате PDF проходят этап извлечения и подготовки текста с применением специализированных библиотек. Подготовленный текст используется в архитектуре RAG, а запросы к языковой модели динамически формируются с учетом типа документа, роли пользователя и задачи анализа. Далее LLM выполняет семантический анализ образовательной программы, выявляет универсальные компетенции (soft skills), оценивает полноту их представления и соответствие установленным требованиям. На основе результатов анализа система автоматически формирует рекомендации по устранению выявленных пробелов и совершенствованию образовательной программы. Метод позволяет автоматизировать анализ документов, повысить объективность оценки и сократить время обработки.

Для проведения интеллектуального анализа в системе применяется подход, основанный на принципах архитектуры Retrieval-Augmented Generation (RAG), при котором извлеченный из документа текст используется в качестве контекстной информации при формировании запроса к языковой модели. Выбор данного подхода обусловлен результатами исследования Xu и соавт. [3], согласно которым применение контекстно-ориентированной архитектуры RAG позволяет повысить точность извлечения компетенций из образовательных программ по сравнению с традиционными подходами к взаимодействию с большими языковыми моделями.

Таким образом, использование архитектуры Retrieval-Augmented Generation позволяет объединить возможности современных больших языковых моделей с информацией, содержащейся непосредственно в анализируемом документе. Это обеспечивает высокую точность извлечения профессиональных компетенций, уменьшает вероятность возникновения ошибок генерации, повышает надежность результатов анализа и делает разработанную систему пригодной для обработки широкого спектра образовательных документов без дополнительного обучения модели.

Одной из ключевых особенностей разработанной системы является механизм динамического управления промптами (Dynamic Prompt Management), обеспечивающий адаптивное взаимодействие с большой языковой моделью. В отличие от традиционного подхода, при котором используется единый статический запрос для всех сценариев обработки документов, разработанная система автоматически формирует содержимое промпта с учетом текущего контекста выполнения задачи. Это позволяет максимально эффективно использовать возможности языковой модели при анализе документов различного типа и назначения.

Необходимость применения механизма динамического управления промптами обусловлена тем, что эффективность работы больших языковых моделей в значительной степени зависит не только от качества входных данных, но и от структуры, содержания и формулировки самого запроса. Даже при использовании одной и той же языковой модели изменение инструкции может существенно повлиять на полноту, точность и формат получаемого результата. По этой причине современные системы, основанные на технологиях искусственного интеллекта, все чаще используют специализированные механизмы генерации промптов, позволяющие адаптировать запрос к конкретной предметной области и решаемой задаче.

В разработанной системе промпт не является заранее фиксированным текстом. Он формируется автоматически на основании нескольких параметров, характеризующих текущий сценарий работы. Основными факторами, влияющими на содержание запроса, являются роль пользователя в системе, тип анализируемого документа, цель анализа и требуемый формат представления результатов. Такой подход позволяет использовать единую языковую модель для решения широкого круга задач без необходимости изменения ее внутренней архитектуры или повторного обучения.

Первым фактором, учитываемым при формировании промпта, является роль пользователя. В зависимости от категории пользователя система может изменять глубину анализа, степень детализации ответа и характер извлекаемой информации. Если системой пользуется администратор, запрос может быть ориентирован на получение технической информации о документе или диагностике процесса обработки. Таким образом, одна и та же языковая модель используется для различных категорий пользователей, однако содержание запроса автоматически адаптируется к их функциональным обязанностям.

Вторым фактором является тип анализируемого документа. Различные виды образовательной документации обладают собственной структурой, содержанием и терминологией. Рабочие программы дисциплин, учебные планы, профессиональные стандарты, образовательные программы и иные документы содержат различные наборы сведений, которые необходимо извлекать в процессе анализа. По этой причине система изменяет инструкцию, передаваемую языковой модели, в зависимости от категории документа. Благодаря такой адаптации языковая модель получает более точные инструкции и концентрируется только на информации, необходимой для конкретного типа документа.

Третьим фактором выступает цель анализа, определяемая выполняемой пользователем задачей. Даже при обработке одного и того же документа система может формировать различные промпты. Например, в одном случае требуется извлечь перечень профессиональных компетенций, в другом – определить результаты обучения, в третьем – проверить соответствие документа установленному шаблону или сформировать краткое аналитическое резюме. Поскольку каждая из перечисленных задач требует различных инструкций, система автоматически изменяет текст запроса, обеспечивая наиболее эффективное использование возможностей языковой модели.

Дополнительно механизм динамического управления промптами учитывает требуемый формат выходных данных. В зависимости от дальнейшего использования результатов языковой модели ответ может формироваться в виде обычного текста, маркированного списка, структурированной таблицы или объекта в формате JSON. Особенно важным является

формирование структурированных данных при их последующей автоматической обработке программными средствами. В этом случае промпт содержит дополнительные инструкции, определяющие строгую структуру ответа, перечень обязательных полей и правила их заполнения. Это значительно упрощает дальнейший программный анализ результатов и их сохранение в базе данных.

Формирование промпта осуществляется программным модулем непосредственно перед обращением к языковой модели. На первом этапе система определяет параметры текущего запроса, включая сведения о пользователе, выбранной операции, типе документа и необходимом формате результата. Затем из набора шаблонов выбираются соответствующие фрагменты инструкций, которые объединяются с контекстом, полученным после извлечения текста из документа в рамках архитектуры Retrieval-Augmented Generation. После этого формируется итоговый запрос, содержащий системную инструкцию, контекст анализируемого документа и описание задачи, который передается языковой модели для выполнения семантического анализа.

Следует отметить, что применение механизма динамического управления промптами позволяет реализовать принцип разделения логики обработки документов и логики взаимодействия с языковой моделью. При необходимости изменения требований к анализу документов разработчику достаточно скорректировать соответствующий шаблон промпта, не изменяя программный код системы и не выполняя повторное обучение модели. Это существенно упрощает сопровождение программного обеспечения и обеспечивает возможность быстрого расширения функциональных возможностей системы.

Использование динамического формирования промптов также способствует повышению качества семантического анализа. Благодаря тому, что языковая модель получает инструкции, максимально соответствующие конкретной задаче, уменьшается вероятность неоднозначной интерпретации запроса, снижается количество нерелевантной информации в ответе и повышается точность извлечения требуемых сведений. Кроме того, адаптация запроса позволяет эффективнее использовать ограниченный контекст языковой модели, включая в него только наиболее значимые инструкции и текстовые данные.

С архитектурной точки зрения механизм динамического управления промптами обеспечивает масштабируемость разработанной системы. По мере появления новых типов документов, новых ролей пользователей или дополнительных аналитических задач отсутствует необходимость переработки существующего модуля искусственного интеллекта. Достаточно добавить новые шаблоны промптов и определить правила их выбора. Такой подход соответствует современным принципам модульного проектирования интеллектуальных информационных систем и обеспечивает возможность дальнейшего развития программного комплекса без существенного изменения его архитектуры.

Таким образом, применение механизма динамического управления промптами является одним из ключевых элементов разработанной системы интеллектуального анализа документов. Его использование позволяет адаптировать взаимодействие с большой языковой моделью к различным сценариям обработки, повысить точность извлечения информации, обеспечить единообразие формируемых результатов, а также значительно увеличить гибкость, расширяемость и сопровождаемость программного обеспечения.

Результаты интеллектуального анализа автоматически сохраняются в реляционной базе данных и становятся доступными пользователям через веб-интерфейс. Дополнительно реализован механизм генерации персонализированной обратной связи, основанный на совместном анализе результатов оценки образовательной программы и ответов студента. Интеграция модулей обработки документов, динамического управления промптами, больших языковых моделей и централизованного хранения данных обеспечивает автоматизацию анализа образовательных программ, снижает трудоемкость экспертной оценки и повышает объективность анализа содержания компетенций**.**

Результаты исследования

В результате проведенного исследования была разработана интеллектуальная аналитическая система для анализа образовательных программ, обеспечивающая автоматизированную обработку документов и интеллектуальный анализ их содержания с использованием технологий искусственного интеллекта. Реализована логика функционирования системы с учетом сценариев работы преподавателя, студента и администратора. Более подробно остановимся на этом.

1) Разработка интеллектуальной аналитической системы для анализа образовательных программ.

Архитектурной основой системы являются современные веб-технологии и средства искусственного интеллекта. Серверная часть реализована с использованием фреймворка FastAPI, обеспечивающего высокую производительность, поддержку асинхронной обработки запросов и удобное построение REST API для взаимодействия клиентской и серверной частей приложения. Пользовательский интерфейс разработан как веб-приложение, обеспечивающее регистрацию и авторизацию пользователей, работу с личными кабинетами и доступ к функциям системы в соответствии с назначенными ролями.

Для хранения информации используется реляционная база данных, в которой сохраняются сведения о пользователях, образовательных программах, результатах анализа и истории обработки документов. Такой подход обеспечивает целостность данных, поддержку многопользовательского режима работы и возможность последующего анализа накопленной информации.

Обработка образовательных программ осуществляется с помощью специализированных библиотек для работы с PDF-документами, которые выполняют извлечение текстового содержимого и его предварительную обработку. Подготовленный текст передается в модуль интеллектуального анализа, использующий методы обработки естественного языка (Natural Language Processing, NLP) для анализа структуры документа и подготовки контекста.

Интеллектуальный анализ содержания образовательных программ реализован на основе больших языковых моделей (Large Language Models, LLM). Для повышения качества обработки применяется архитектура Retrieval-Augmented Generation (RAG), при которой извлеченный из документа текст используется в качестве контекста запроса к языковой модели. Дополнительно реализован механизм динамического формирования промптов, позволяющий адаптировать инструкции модели в зависимости от типа документа, роли пользователя и выполняемой задачи. Такой подход повышает точность выявления универсальных компетенций (soft skills), снижает вероятность генерации нерелевантной информации и обеспечивает возможность расширения функциональности системы без изменения ее базовой архитектуры (рисунок 1).

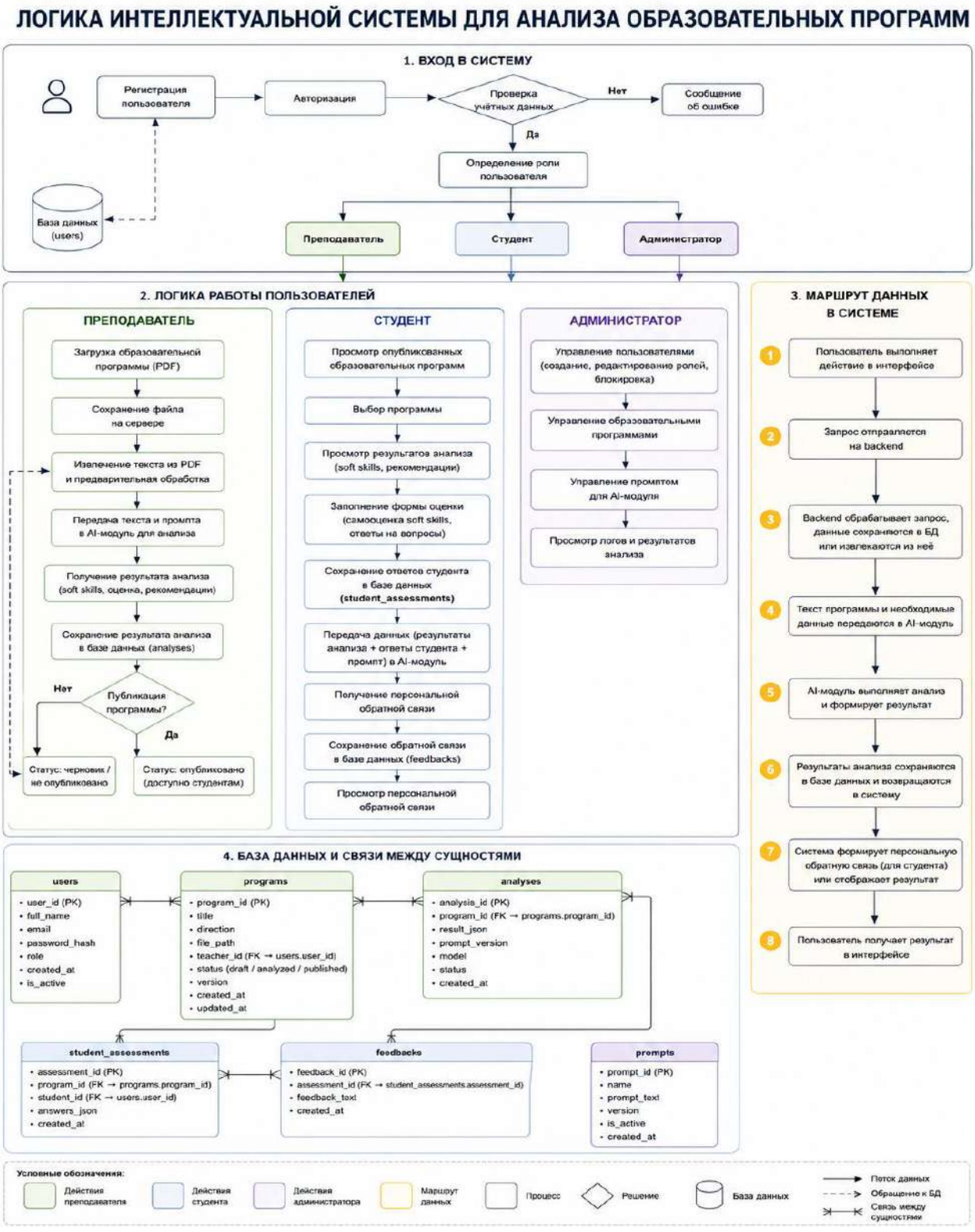


Рисунок 1. Логика работы разработанной интеллектуальной аналитической системы для анализа образовательных программ

2) Сценарии взаимодействия пользователей различных ролей с интеллектуальной системой.

1. Регистрация, авторизация и определение роли

Первый этап работы начинается с регистрации пользователя на сайте. Пользователь заполняет форму, в которую обычно входят имя, электронная почта, пароль и выбранная роль. Frontend передаёт эти данные через HTTP-запрос на backend. Сервер принимает запрос, проверяет корректность формата полей, убеждается, что email ещё не используется, шифрует пароль и создаёт новую запись в таблице пользователей. В базе данных сохраняются идентификатор пользователя, ФИО или имя, email, хэш пароля, роль, дата регистрации и статус активности аккаунта.

После регистрации пользователь проходит авторизацию. Backend сравнивает введённые данные со сведениями из базы, и при успешной проверке создаёт токен доступа или сессионную запись. После этого frontend использует этот токен во всех последующих запросах, чтобы сервер мог определить, кто именно инициирует действие. Это необходимо для разграничения доступа между преподавателем, студентом и администратором.

Данный процесс регистрации, авторизации и разграничения прав доступа пользователей представлен на рисунке 2, который отражает взаимодействие frontend, backend и базы данных при проверке учетных данных и предоставлении доступа к системе.

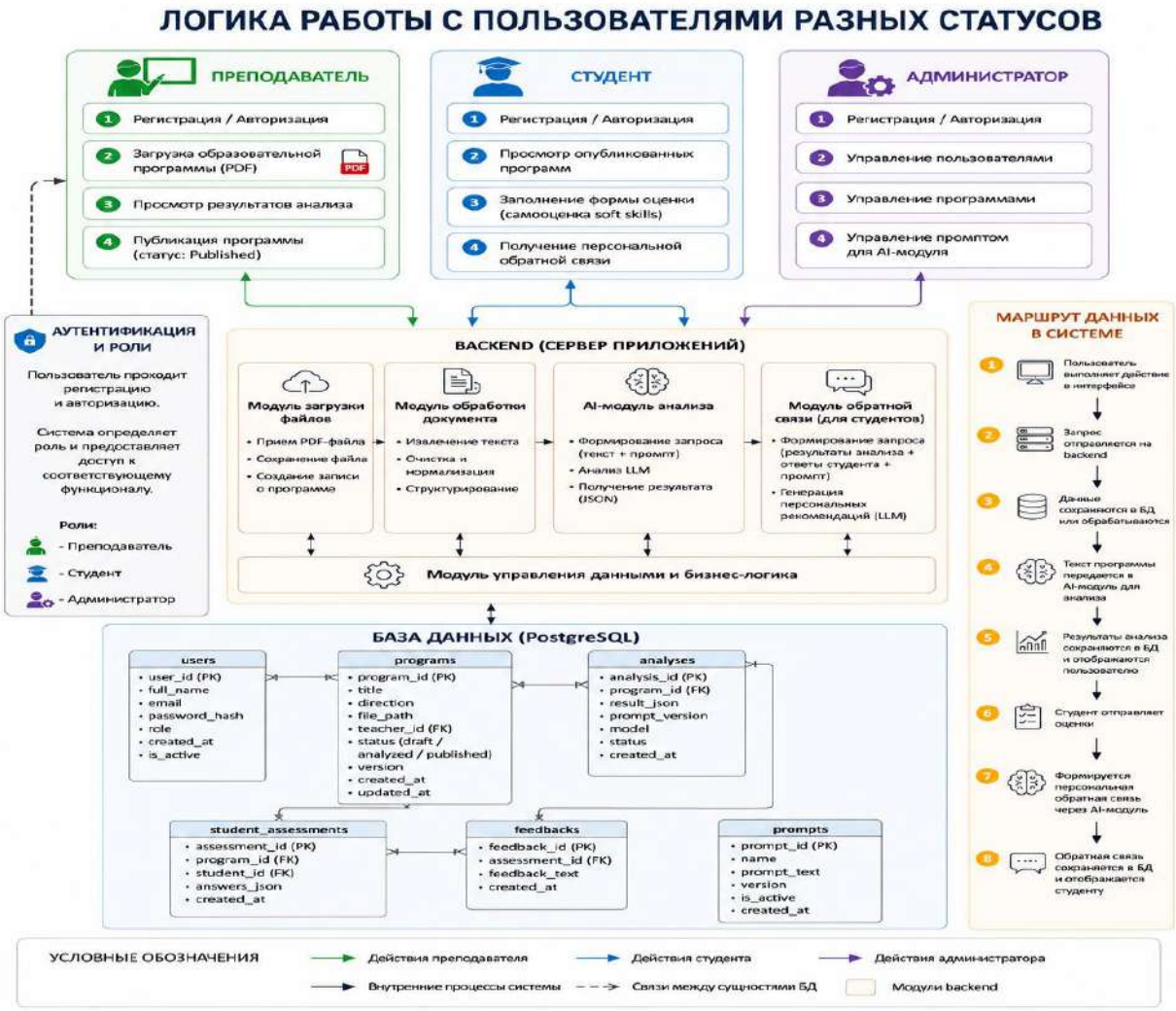


Рисунок 2. Функциональные сценарии интеллектуальной системы анализа образовательных программ

Роль пользователя определяет весь дальнейший сценарий работы. Если в систему вошёл преподаватель, ему открывается интерфейс загрузки образовательных программ и просмотра результатов анализа. Если входит студент, он получает доступ к уже опубликованным

программам, форме оценки и персональной обратной связи. Если входит администратор, он получает полный доступ ко всем сущностям системы, включая изменение промпта, используемого языковой моделью.

2. Логика работы преподавателя

Работа преподавателя начинается с добавления образовательной программы в систему. На сайте он выбирает PDF-файл, при необходимости указывает название программы, направление подготовки, учебный год или комментарий, после чего frontend отправляет файл и сопроводительные данные на backend. Передача файла выполняется как multipart/form-data, потому что в одном запросе нужно передать и двоичные данные файла, и обычные текстовые поля. Сервер получает этот запрос и сначала проверяет, имеет ли пользователь право на выполнение загрузки. Если роль подтверждена, backend создаёт запись о новой образовательной программе в таблице programs. В этой записи могут храниться program_id, название ОП, путь к файлу, идентификатор преподавателя, дата загрузки, статус документа, номер версии и другие служебные поля. Сам PDF-файл обычно сохраняется в файловом каталоге сервера или в отдельном хранилище, а в базу данных записывается путь к этому файлу, чтобы не перегружать БД бинарными данными.

После физического сохранения файла запускается этап его обработки. Специальный модуль открывает PDF-документ, извлекает из него текст, очищает его от лишних символов, служебных переносов, пустых блоков и других элементов, которые не нужны для смыслового анализа. Затем текст может быть разбит по разделам, например по описанию компетенций, результатам обучения, дисциплинам или блокам, где потенциально описаны soft skills. Эта промежуточная текстовая версия документа может сохраняться в базе данных в поле raw_text либо в отдельной таблице, чтобы при необходимости не выполнять повторное извлечение текста из PDF.

Следующим этапом extracted text передаётся в модуль анализа. Здесь backend формирует запрос к AI-модулю, который включает сам текст образовательной программы, активный промпт, системные инструкции и формат ожидаемого ответа. За счёт этого модель получает не просто документ, а чёткую задачу: определить наличие soft skills, выделить найденные навыки, оценить полноту их представления и предложить рекомендации по улучшению образовательной программы.

Ответ модели возвращается на backend уже как структурированный результат. Это может быть JSON-объект, содержащий перечень выявленных soft skills, их категории, текстовые комментарии, итоговую оценку и рекомендации. Сервер сохраняет этот результат в таблицу analyses, где каждая запись связана с конкретной образовательной программой. Также здесь могут фиксироваться версия промпта, дата анализа, статус выполнения и используемая модель.

После завершения анализа frontend получает готовый результат и отображает его преподавателю. На странице анализа он может увидеть структуру найденных soft skills, интерпретацию модели, замечания по содержанию документа и возможные рекомендации. На этом этапе программа ещё не обязательно видна студентам, потому что окончательное решение о её публикации принимает преподаватель.

Если преподавателя устраивает результат, он нажимает кнопку сохранения или публикации. Тогда frontend отправляет на backend команду изменить статус программы. Сервер обновляет запись в таблице programs: например, меняет статус с draft или analyzed на published. С этого момента программа попадает в список доступных для студентов. Если преподаватель не согласен с результатом, он может оставить программу в черновике, удалить её или загрузить исправленную версию для повторного анализа.

3. Логика работы студента

Студент не работает с загрузкой документов и не может изменять образовательные программы. Его сценарий начинается с входа в систему и просмотра опубликованных ОП. Frontend отправляет запрос на backend, сервер выбирает из таблицы programs только те записи,

у которых статус равен published, и возвращает их список. Таким образом студент видит только проверенные и подтверждённые программы, а промежуточные версии и черновики ему недоступны.

После выбора конкретной программы студент открывает её страницу. В этот момент frontend может запросить сразу несколько наборов данных: общую информацию о программе, сохранённый результат анализа из таблицы analyses и форму для ввода собственных оценок или ответов. То есть страница студента строится не только на статических данных программы, но и на результатах уже проведённого анализа.

Далее студент заполняет форму. В зависимости от реализации это может быть самооценка soft skills, выставление баллов по отдельным навыкам, ответы на вопросы или текстовые комментарии. После отправки формы все введённые данные уходят на backend и валидируются: сервер проверяет, что выбранная программа действительно опубликована, что студент имеет право её проходить, и что структура данных соответствует ожидаемому формату. После проверки запись сохраняется в таблицу student_assessments .

В этой таблице могут храниться assessment_id, student_id, program_id, дата отправки, значения оценок, текстовые ответы и статус обработки. Таким образом в базе фиксируется не просто факт взаимодействия студента с программой, а полный набор его ответов, который затем можно использовать для генерации обратной связи, повторного анализа или статистики.

4. Генерация обратной связи студенту

После того как ответы студента сохранены, система переходит к следующему этапу – формированию персональной обратной связи. Для этого backend извлекает из базы несколько связанных наборов данных: саму образовательную программу, ранее полученный результат её анализа, ответы студента и активный промпт администратора. Затем всё это объединяется в один входной пакет для AI-модуля.

На этом этапе особенно важно, что система сравнивает разные источники информации. С одной стороны, у неё уже есть анализ того, какие soft skills отражены в образовательной программе. С другой стороны, есть реальные оценки или ответы студента по этим же навыкам. Модель получает оба источника и может сформировать персонализированную обратную связь: определить сильные стороны студента, выявить пробелы, показать, насколько его самооценка соотносится с содержанием программы, и предложить рекомендации по развитию.

Сгенерированный результат сохраняется в таблицу feedback. В ней могут находиться feedback_id, student_id, program_id, assessment_id, текст обратной связи, дата генерации, версия промпта и, при необходимости, техническая информация о модели. После записи в БД backend возвращает результат на frontend, и студент видит его в своём кабинете или на странице выбранной образовательной программы.

Таким образом, в логике студента происходит важное преобразование данных: сначала он отправляет сырой набор своих ответов, затем эти ответы связываются с анализом образовательной программы, после чего AI-модуль превращает объединённые данные в содержательную обратную связь. Это значит, что обратная связь не существует сама по себе, а возникает как результат объединения нескольких сущностей системы.

5. Логика работы администратора

Администратор является самым привилегированным пользователем системы. Он может выполнять действия преподавателя и студента, а также управлять внутренними настройками сайта. Главная дополнительная функция администратора – изменение промпта, по которому работает языковая модель.

Когда администратор открывает страницу настроек, frontend отправляет запрос на backend для получения текущего активного промпта. Сервер извлекает его из таблицы settings или prompts и возвращает в интерфейс. После редактирования администратор сохраняет изменения, и frontend отправляет новый текст промпта обратно на сервер.

Backend проверяет права пользователя, создаёт новую запись в таблице промптов или обновляет текущую активную запись, после чего помечает её как действующую. Обычно в такой таблице хранятся prompt_id, текст промпта, дата изменения, автор изменения, версия и признак активности. Это позволяет не просто заменить инструкцию, но и хранить историю изменений, чтобы можно было понять, по какой версии промпта выполнялся тот или иной анализ.

После активации нового промпта все последующие обращения к AI-модулю используют уже обновлённую инструкцию. Это касается как анализа вновь загруженных образовательных программ, так и генерации обратной связи студентам. Благодаря этому администратор может изменять критерии, стиль ответа, глубину анализа и формат выводов без изменения исходного кода backend-части.

6. База данных и связи между сущностями

Вся система логически объединяется через единую реляционную базу данных. Основные таблицы можно описать следующим образом: users – данные пользователей; programs – сведения о загруженных образовательных программах; analyses – результаты анализа ОП; student_assessments – ответы студентов; feedback – сформированная обратная связь; settings или prompts – системные настройки и активный промпт.

Связи между таблицами определяют логику работы системы. Один преподаватель из users может быть связан со многими программами из programs через поле uploaded_by. Каждая программа из programs может иметь одну или несколько записей в analyses, если анализ выполнялся повторно или для разных версий документа. Одна опубликованная программа может быть связана со многими студентами через таблицу student_assessments. Каждая запись из student_assessments может иметь соответствующую запись в feedback, в которой хранится итоговая обратная связь. Таблица prompts влияет на analyses и feedback, потому что именно её содержимое используется при запросах к модели.

Такая схема позволяет проследить полный жизненный цикл данных. Например, можно определить, какой преподаватель загрузил конкретную программу, какой текст из неё был извлечён, какой анализ был выполнен, по какому промпту он формировался, какой студент выбрал эту программу, какие оценки он выставил и какую обратную связь в итоге получил. Это особенно важно для прозрачности системы, повторного анализа и аудита результатов.

7. Полный маршрут данных внутри системы

Если описывать систему как непрерывный поток данных, то она работает следующим образом. Пользователь создаёт аккаунт, и его данные сохраняются в users. Преподаватель загружает PDF-файл, который физически сохраняется в файловом хранилище, а информация о нём – в programs. Затем из PDF извлекается текст, который очищается и передаётся в AI-модуль. Модель формирует результат анализа, а backend записывает его в analyses. После одобрения преподавателем программа переводится в статус published.

Далее студент выбирает опубликованную программу и отправляет свои оценки. Эти данные записываются в student_assessments. После этого backend извлекает из базы анализ программы, ответы студента и активный промпт, объединяет их и отправляет в AI-модуль. На выходе получается персональная обратная связь, которая сохраняется в feedback и возвращается студенту через frontend.

Таким образом, данные в системе постоянно меняют состояние. PDF-файл превращается в текст, текст – в анализ, анализ – в опубликованный объект сайта, ответы студента – в основу для персональной обратной связи, а промпт администратора – в управляющий элемент, влияющий на все будущие результаты. Именно эта последовательная трансформация данных и составляет основную логику работы всей программы.

Дискуссия

Полученные результаты согласуются с современными тенденциями развития интеллектуальных образовательных систем, основанных на применении методов Educational

Data Mining, Learning Analytics и технологий искусственного интеллекта. Вместе с тем предлагаемая веб-система отличается от большинства существующих исследований тем, что ориентирована не только на интеллектуальный анализ образовательных документов, но и на организацию полного цикла обработки данных в единой многопользовательской информационной среде.

Результаты исследования Zhumazhanova и соавт. (2026) демонстрируют, что использование методов Learning Analytics в университетах Казахстана, в частности в Казахском национальном университете имени аль-Фараби и Satbayev University, способствует повышению вовлеченности обучающихся и эффективности образовательного процесса [7]. Однако основное внимание в данной работе уделяется анализу образовательных данных студентов и поддержке процесса обучения. В отличие от этого подхода, разработанная интеллектуальная аналитическая система ориентирована на анализ содержания образовательных программ, автоматизированную оценку сформированности универсальных компетенций и поддержку принятия решений при совершенствовании образовательных программ. Таким образом, предлагаемое решение расширяет область применения образовательной аналитики, распространяя ее на этап проектирования и актуализации образовательных программ.

Вместе с тем результаты настоящего исследования подтверждают выводы Klevtsov (2026) о существующих барьерах внедрения технологий искусственного интеллекта в высших учебных заведениях Казахстана [8]. К числу основных ограничений относятся недостаточная готовность цифровой инфраструктуры университетов, отсутствие единых нормативных подходов к использованию генеративного искусственного интеллекта, а также ограниченная интеграция интеллектуальных сервисов с существующими информационными системами вузов. Предложенная интеллектуальная система учитывает указанные ограничения за счет использования клиент-серверной архитектуры, централизованного хранения данных, ролевой модели доступа и модульного построения программного обеспечения, что обеспечивает возможность интеграции с существующими образовательными платформами и дальнейшего расширения функциональности.

Следует отметить, что настоящее исследование имеет ряд ограничений. Предлагаемая система ориентирована на анализ образовательных программ, представленных в формате PDF, а качество интеллектуального анализа зависит от полноты и структуры исходных документов, а также возможностей используемой большой языковой модели. Кроме того, эффективность работы системы требует дальнейшей экспериментальной проверки на более широком наборе образовательных программ различных университетов и направлений подготовки.

Перспективы дальнейших исследований связаны с интеграцией веб-системы с LMS-платформами университетов, использованием технологий Retrieval-Augmented Generation (RAG) с подключением внешних баз знаний, расширением механизмов динамического управления промптами и разработкой инструментов автоматического мониторинга соответствия образовательных программ требованиям профессиональных стандартов и рынка труда.

Заключение

В результате применения предложенных методов была разработана интеллектуальная аналитическая система для анализа образовательных программ, обеспечивающая автоматизированную обработку документов и интеллектуальный анализ их содержания с использованием технологий искусственного интеллекта. Полученные результаты включают описание архитектуры разработанной системы и логики её функционирования, охватывающей взаимодействие пользователей различных ролей, организацию хранения данных и маршрут их обработки внутри системы.

Разработанная система реализована в виде многопользовательского веб-приложения с клиент-серверной архитектурой и включает frontend-интерфейс, backend, реляционную базу

данных, модуль обработки PDF-документов и модуль искусственного интеллекта. Особенностью системы является организация полного цикла обработки данных - от загрузки образовательной программы и извлечения текста до интеллектуального анализа содержания, сохранения результатов и формирования рекомендаций для пользователей различных ролей. В отличие от локальных программных решений, система обеспечивает централизованное хранение данных, сохранение состояния объектов (stateful), разграничение прав доступа и возможность масштабирования для работы нескольких пользователей одновременно.

Интеллектуальный анализ образовательных программ выполняется с использованием методов обработки естественного языка и больших языковых моделей. Реализованный механизм динамического управления промптами позволяет адаптировать запросы к языковой модели в зависимости от роли пользователя, типа анализируемого документа и выполняемой задачи. Это обеспечивает более гибкий и объективный анализ содержания образовательных программ и позволяет автоматически выявлять универсальные компетенции (soft skills), а также формировать рекомендации по их совершенствованию.

Благодарность

Данная статья выполнена в рамках проекта грантового финансирования Министерства образования и науки Республики Казахстан, AP23489805 «Методика интеллектуального анализа данных для выявления факторов, влияющих на профессиональное развитие студентов в технических специальностях».

Список использованных источников

[1]Romero C., Ventura S. Educational Data Mining and Learning Analytics: An Updated Survey. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery. 2020. DOI: 10.1002/widm.1355.

[2]López-Meneses E., et al. Educational Data Mining and Predictive Modeling in the Age of AI. Computers. 2025. DOI: 10.3390/computers14020068.

[3]Xu Z., Yu R., et al. From Course to Skill: Evaluating LLM Performance in Curricular Analytics. 2025. DOI: 10.1007/978-3-031-98465-5_26

[4]Mapademics. AI-Powered Curriculum Mapping for Academic Excellence. White Paper. 2025. URL: https://www.mapademics.com/resources/whitepapers/ai-powered-curriculum-mapping-for-academic-excellence.

[5]Togmanov M., et al. KazMMLU: Evaluating Language Models on Kazakh, Russian, and Regional Knowledge. arXiv:2502.12829. 2025. DOI: 10.18653/v1/2025.acl-long.701

[6]Digital Qazaqstan. Государственная стратегия цифрового развития Республики Казахстан до 2029 года. URL: https://adilet.zan.kz/rus/docs/U2600001311

[7]Zhumazhanova K. I., et al. Learning Analytics and Ergonomic Educational Spaces for Active Learning: A Case Study from Kazakhstan. Frontiers in Education. 2026. DOI: 10.3389/feduc.2026.1801026.

[8]Klevtsov I. AI in Kazakh Universities: Institutions Are Not Ready for the ChatGPT Era. Times of Central Asia. 2026. URL: https://timesca.com/ai-in-kazakh-universities/

ҒТАХР 14.35.09

https://doi.org/10.51889/2959-5894.2026.95.3.026

Б.Ж. Нурбеков[1], А.О.Байарыстан[1], Д.С Майдисарова[2], С.Ә. Сантеева[1], Г.Б. Толегенова[1*]

[1] Л.Н. Гумилев атындағы Еуразия ұлттық университеті, Астана қ., Қазақстан
[2]Торайғыров университеті, Павлодар қ., Қазақстан
[*]*e-mail: gulnaztolegenova@mail.ru*

ЖАСАНДЫ ИНТЕЛЛЕКТ НЕГІЗІНДЕ ЖОҒАРЫ МАТЕМАТИКАНЫ ОҚЫТУДЫҢ ӘДІСТЕМЕЛІК МОДЕЛІ

Аңдатпа

Соңғы жылдары жасанды интеллект технологиялары қарқынды дамып, жоғары білім беру жүйесінде математика пәндерін оқытуда жаңа мүмкіндіктер ашуда. Осыған байланысты жұмыста жасанды интеллект негізінде жоғары математиканы оқытудың әдістемелік моделін жасау және осы модельдің тиімділігін бағалауға арналған диагностикалық-аналитикалық құралдарды ұсыну мақсат етілді. Жұмыс барысында виртуалды есептер жинағы мен интеллектуалды чат-боттың нақты қандай функцияны атқаратыны айқындалып, технологияны қолдану реті мен ішкі құрылымы белгіленді, оқу жетістіктерін өлшейтін критерийлер тұжырымдалды, студенттердің қанағаттануын бағалайтын сауалнама әзірленді, деректерді жинау мен статистикалық өңдеудің тәртібі қалыптастырылды, ал ұсынылған талдамалық аппарат модельдік деректерде тексерістен өтті. Әдістемелік модель бір-бірімен тығыз байланысты төрт блоктан – мақсаттық, мазмұндық-технологиялық, ұйымдастырушылық-әрекеттік және бағалау-нәтижелік блоктардан тұрады, сондай-ақ интеллектуалды қолдаудың үш деңгейі, атап айтқанда бағдарлаушы, бағыттаушы және түсіндіруші деңгейлер әзірленді. Нәтижелерді бағалау үшін оқу жетістіктері мен студенттердің қанағаттануы бойынша екі топ көрсеткіш ұсынылды, ал модельдік деректер негізінде жүргізілген валидация ұсынылған аналитикалық хаттаманың жұмысқа жарамды екенін дәлелдеді. Зерттеудің ғылыми жаңалығы виртуалды есептер жинағы мен чат-ботты бірыңғай оқу циклінде реттелетін ЖИ-қолдау деңгейлерімен және екі компонентті бағалау жүйесімен ұштастыруында жатыр, ал практикалық маңыздылығы әзірленген модель мен құралдарды Қазақстан Республикасының жоғары оқу орындарында қолдану мүмкіндігінде жатыр.

Түйін сөздер: жасанды интеллект, жоғары математика, оқыту технологиясы, виртуалды есептер жинағы, чат-бот, әдістемелік модель, оқу жетістіктері.

Б.Ж. Нурбеков[1]., А.О. Байарыстан[1], Д.С. Майдисарова[2], С.Ә. Сантеева[1], Г.Б.Толегенова[1]
[1] Л.Н. Евразийский национальный университет имени Гумилева, г.Астана, Казахстан
[2] Университет Торайгырова, г. Павлодар, Казахстан

МЕТОДИЧЕСКАЯ МОДЕЛЬ ОБУЧЕНИЯ ВЫСШЕЙ МАТЕМАТИКЕ НА ОСНОВЕ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

Аннотация

В последние годы технологии искусственного интеллекта стремительно развиваются, открывая новые возможности в преподавании математических дисциплин в системе высшего образования. В связи с этим в работе была поставлена цель разработать методическую модель обучения высшей математике на основе искусственного интеллекта, а также предложить диагностико-аналитический инструментарий для оценки эффективности данной модели. В ходе работы были определены конкретные функции, выполняемые виртуальным сборником задач и интеллектуальным чат-ботом, установлены порядок применения технологии и её внутренняя структура, сформулированы критерии измерения учебных достижений, разработана анкета для оценки удовлетворённости студентов, сформирован порядок сбора данных и статистической обработки, а предложенный аналитический аппарат прошёл проверку на модельных данных. Методическая модель состоит из четырёх тесно взаимосвязанных блоков-целевого, содержательно-технологического, организационно-

деятельностного и оценочно-результативного, а также разработаны три уровня интеллектуальной поддержки: ориентирующий, направляющий и разъясняющий. Для оценки результатов предложены две группы показателей – учебные достижения и удовлетворённость студентов, а валидация, проведённая на основе модельных данных, подтвердила работоспособность предложенного аналитического протокола. Научная новизна исследования заключается в сочетании виртуального сборника задач и чат-бота в едином учебном цикле с регулируемыми уровнями ИИ-поддержки и двухкомпонентной системой оценки. Практическая значимость определяется возможностью применения разработанной модели и инструментов в высших учебных заведениях Республики Казахстан.

Ключевые слова: искусственный интеллект, высшая математика, технология обучения, виртуальный задачник, чат-бот, методическая модель, учебные достижения.

B.Zh. Nurbekov[1], A.O. Baiarystan[1], D.S. Maidisarova[2], S.A. Santeeva[1], G.B. Tolegenova[1]
[1] L.N. Eurasian National University named after Gumileva, Astana, Kazakhstan
[2] Toraygyrova University, Pavlodar, Kazakhstan

A METHODOLOGICAL MODEL FOR TEACHING HIGHER MATHEMATICS BASED ON ARTIFICIAL INTELLIGENCE

Abstract

In recent years, artificial intelligence technologies have been developing rapidly, opening up new opportunities for teaching mathematical disciplines within the higher education system. In this regard, the study aimed to develop a methodological model for teaching higher mathematics based on artificial intelligence, as well as to propose diagnostic and analytical tools for evaluating its effectiveness. In the course of the work, the specific functions performed by the virtual problem set and the intelligent chatbot were identified, the sequence of technology application and its internal structure were established, criteria for measuring learning achievements were formulated, a questionnaire for assessing student satisfaction was developed, a procedure for data collection and statistical processing was established, and the proposed analytical apparatus was tested on model data. The methodological model consists of four closely interrelated blocks — target, content-technological, organizational-activity, and evaluation-result — and three levels of intelligent support were developed: orienting, guiding, and explanatory. Two groups of indicators — learning achievements and student satisfaction — were proposed to evaluate the results, and validation performed on model data confirmed the operability of the proposed analytical protocol. The scientific novelty of the study lies in combining the virtual problem set and the chatbot within a unified learning cycle featuring adjustable levels of AI support and a two-component evaluation system. The practical significance is determined by the possibility of applying the developed model and tools in higher education institutions of the Republic of Kazakhstan.

Keywords: artificial intelligence, higher mathematics, teaching technology, virtual problem-solving environment, chatbot, methodological model, academic achievement.

Кіріспе

Негізгі ережелер:

1. Жоғары математиканы оқытуда жасанды интеллект құралдарын педагогикалық негізде пайдалану үшін виртуалды есептер жинағы мен интеллектуалды чат-ботты біріктіретін төрт компонентті әдістемелік модель әзірленді.
2. Интеллектуалды қолдаудың бағдарлаушы, бағыттаушы және түсіндіруші деңгейлері студенттерге мөлшерленген көмек алуға және өзіндік шешімге міндетті түрде қайта оралуға мүмкіндік береді.
3. Нәтижелерді бағалау екі бағытта жүзеге асырылды: оқу жетістіктері (диагностикалық тапсырмалар) және студенттердің қанағаттануы (сауалнама).
4. Модельдік валидация ұсынылған аналитикалық хаттаманың статистикалық өңдеу рәсімдерін қолдануға жарамды екенін көрсетті.

Әзірленген модельді Қазақстан Республикасының жоғары оқу орындарында математикалық пәндерді оқытуда ЖИ құралдарын кіріктірудің әдіснамалық негізі ретінде пайдалануға болады.

Жоғары білім берудегі цифрлық трансформация жасанды интеллект технологияларын оқу үдерісіне енгізудің жаңа бағыттарын ашуда. Бұл процестер математикалық пәндерді оқытуда ерекше маңызды, өйткені студенттердің практикалық дағдылары мен өзіндік жұмысының тиімділігі көбінесе кері байланыс жүйесінің қаншалықты уақтылы әрі сапалы жұмыс істейтініне байланысты. Осыған орай заманауи зерттеулер де адаптивті оқыту жүйелері мен интеллектуалды қолдау құралдарының білім беру үдерісіндегі рөлінің арта түсу қажеттігін көрсетіп отыр [1,2].

Генеративті модельдер мен диалогтық жүйелердің дамуы білім алушылармен интерактивті өзара әрекеттесудің жаңа мүмкіндіктерін ашады. Чат-боттар табиғи тілде сұхбат жүргізе алады, түсініктемелер бере алады, оқу тапсырмаларын орындау барысында көмек көрсете алады[3]. Дегенмен бұл құралдарды пайдалану бірқатар тәуекелмен қатар жүреді: студенттің дайын шешімге тәуелді болуы, алынған ақпаратты сыни тұрғыдан бағаламауы, өзіндік ойлау белсенділігінің төмендеуі осының қатарында [4,5]. Осыған байланысты ЖИ құралдарын педагогикалық негізде пайдалану стратегиясын әзірлеу қажеттілігі туындайды.

Математикалық пәндерді оқытуда интеллектуалды жүйелердің тиімділігі бірқатар зерттеулерде дәлелденген. Атап айтқанда, VanLehn K. [6] интеллектуалды оқыту жүйелерінің дәстүрлі оқытуға қарағанда тиімділігін мета-талдау арқылы көрсетсе, Steenbergen-Hu және Cooper [7] мұндай жүйелердің жоғары оқу орны студенттерінің оқу нәтижелеріне оң әсерін анықтады. және т.б. [8] математиканы оқытудағы адаптивті жүйелердің тиімділігін мета-талдау арқылы зерттеп, олардың дәстүрлі оқытуға қосымша құрал ретінде қолданылғанда айтарлықтай нәтиже беретінін көрсетті. Соңғы жылдары чат-боттарды оқытуда, соның ішінде математиканы оқытуда қолдануға арналған эмпирикалық жұмыстар да пайда болды [9–11]. Бұл зерттеулер ЖИ құралдарының математикалық дайындық тиімділігіне ықпал ететінін көрсетеді, алайда олардың көпшілігі жекелеген аспектілерге (тапсырмаларды автоматты тексеру, кері байланыс) шоғырланған, ал кешенді әдістемелік модель әзірлеу мәселесі жеткіліксіз зерттелген. Жоғары математиканы оқыту студенттерден ұғымдық аппаратты терең меңгеруді, математикалық объектілер арасындағы байланыстарды түсінуді және әртүрлі деңгейдегі есептерді шешу дағдыларын қалыптастыруды талап етеді. Ал студенттердің бастапқы дайындығындағы айырмашылықтар тапсырмаларды саралауды және қиындық туындаған жағдайда қосымша қолдау көрсетуді қажет етеді [12]. Осы тұрғыдан алғанда, математикалық есептерді шешудің құрылымдық ортасын диалогтық ЖИ мүмкіндіктерімен ұштастыру перспективалы бағыт ретінде қарастырылады.

Зерттеудің мақсаты – жасанды интеллект негізінде жоғары математиканы оқыту технологиясын қолданудың әдістемелік моделін әзірлеу және осы модельдің тиімділігін бағалауға арналған диагностикалық-аналитикалық құралдарды ұсыну.

Зерттеу міндеттері:

–виртуалды есептер жинағы мен интеллектуалды чат-боттың функционалдық рөлін айқындау;

–ЖИ негізінде оқыту технологиясын қолданудың құрылымы мен реттілігін белгілеу;

–оқу жетістіктерін бағалаудың өлшемдері мен көрсеткіштерін анықтау;

–студенттердің қанағаттану деңгейін бағалау құралдарын әзірлеу;

–деректерді жинау және статистикалық өңдеу рәсімін қалыптастыру;

–ұсынылған талдамалық құралдардың модельдік валидациясын жүргізу.

Зерттеудің ғылыми жаңалығы – виртуалды есептер жинағы мен интеллектуалды чат-ботты жоғары математиканы оқытудың бірыңғай циклінде реттелетін ЖИ-қолдау деңгейлерімен және екі компонентті бағалау жүйесімен біріктіретін модельді әзірлеу.

Зерттеу әдіснамасы

Зерттеу Қазақстан Республикасының жоғары білім беру жүйесінде жасанды интеллект негізінде жоғары математиканы оқыту технологиясын қолданудың әдістемелік моделін

әзірлеу мақсатында орындалды. Ұйымдастыру базасы – Л.Н. Гумилев атындағы Еуразия ұлттық университеті. Зерттеу төрт кезеңнен тұрды: жобалау, диагностикалық, ұйымдастырушылық-әрекеттік және аналитикалық кезеңдер.

Оқыту технологиясының құрылымы. Технология екі негізгі компоненттен тұрады. Виртуалды есептер жинағы әртүрлі күрделілік деңгейіндегі тапсырмаларды ұсына отырып, студенттердің практикалық әрекетін ұйымдастырады. Интеллектуалды чат-бот контекстік қолдау көрсетеді: ұғымдарды түсіндіру, шарттарды нақтылау, бағыттаушы кеңестер беру, шешім кезеңдерін түсіндіру, қателерді анықтау, ұқсас тапсырмалар ұсыну және өзін-өзі тексеруге арналған сұрақтар қалыптастыру.

Ұйымдастырушылық схема келесі кезеңдерден тұрады: бастапқы диагностика → оқу материалымен жұмыс → есептер жинағында есептерді шешу → чат-ботқа жүгіну → өзіндік шешім → қорытынды диагностика → сауалнама → нәтижелерді талдау.

Оқу жетістіктерін бағалау 100 балдық шкала бойынша орындалады. Бастапқы және қорытынды тапсырмалар салыстырмалы түрде ұсынылады. Бағалау критерийлеріне ұғымды түсіну, әдісті таңдау, түрлендірудің дұрыстығы, дәйектіліктің негізділігі, нәтиженің дұрыстығы және оны түсіндіру жатады.

Қанағаттанушылықты бағалау үшін Лайкерттің бес балдық шкаласына негізделген сауалнама қолданылады. Ол төрт өлшемді қамтиды: пайдалылық, өзара әрекеттесудің ыңғайлылығы, оқу әрекетіне әсері және жалпы қанағаттану; оған қоса ашық сұрақтар да қарастырылған.

Статистикалық өңдеу сипаттамалық статистиканы (орташа мән, медиана, стандартты ауытқу, квартильаралық ауқым), Стьюденттің жұптастырылған t-критерийін немесе Уилкоксон критерийін, әсер мөлшерін, сондай-ақ Кронбахтың α коэффициентін есептеуді қамтиды. Маңыздылық деңгейі $p < 0,05$ болып белгіленді.

Деректермен жұмыс және этикалық қағидаттар тұрғысынан алғанда, зерттеу барысында деректерді иесіздендіру, қатысушыларды ақпараттандыру және деректерді жиынтық түрде талдау принциптері сақталды [13].

Модельдік валидация мақсатында әзірленген аналитикалық хаттаманың жұмысқа қабілеттілігін тексеру үшін бастапқы және қорытынды диагностика көрсеткіштерін, сондай-ақ қанағаттану шкалалары бойынша жауаптарды қамтитын модельдік деректер массиві ($N = 72$) құрылды. Бұл массив көзделген деректерді жинау процедурасына толық сай келді, сондықтан байланысты өлшемдерді талдау әдістерінің қолданылу мүмкіндігін тексеруге болды.

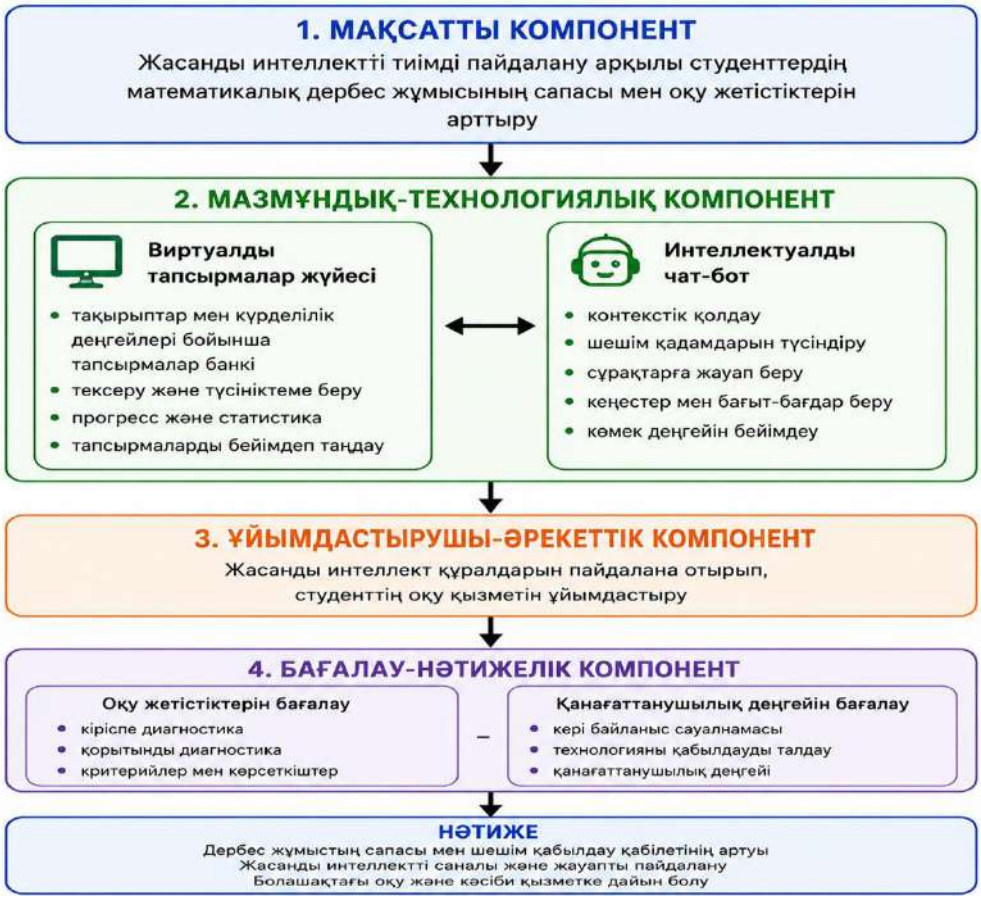
Зерттеу нәтижелері

Зерттеудің негізгі нәтижесі – жасанды интеллект негізінде жоғары математиканы оқыту технологиясын қолданудың әдістемелік моделін әзірлеу болып отыр. Бұл модель өзара байланысты төрт компоненттен тұрады: мақсаттық, мазмұндық-технологиялық, ұйымдастырушылық-әрекеттік және бағалау-нәтижелік. Модельдің құрылымы 1-суретте көрсетілген.

Ескерту: Модель төрт негізгі блокты қамтиды: мақсаттық, мазмұндық-технологиялық, ұйымдастырушылық-әрекеттік және бағалау-нәтижелік.

Мақсаттық компонент студенттің өзіндік математикалық әрекетін ұйымдастыруды және мөлшерленген контекстік қолдау көрсету арқылы оқыту тиімділігін арттыруды көздейді. Осы компонент аясында интеллектуалды қолдаудың үш деңгейі әзірленді, олар студентке көрсетілетін көмек көлемінің біртіндеп артуын қамтамасыз етеді. Әрбір деңгейдің сипаттамасы 2-суретте берілген.

Ескерту: 1-деңгей – бағдарлаушы (есеп шартын түсіндіру); 2-деңгей – бағыттаушы (шешу жолын нұсқау); 3-деңгей – түсіндіруші (толық түсіндірме беру).



Сурет 1. Жасанды интеллект негізінде жоғары математиканы оқыту технологиясын қолданудың әдістемелік моделінің құрылымы

Мазмұндық-технологиялық компонент үш күрделілік деңгейіндегі тапсырмалар банкі мен интеллектуалды қолдау құралдарын біріктіреді, ал технологияның әрбір компоненті өзіне тән функцияны атқарады. Виртуалды есептер жинағы практикалық әрекетті ұйымдастырса, интеллектуалды чат-бот контекстік қолдау көрсетеді. Диагностикалық және аналитикалық модульдер нәтижелерді бағалау мен жинақтауды қамтамасыз етеді. Компоненттердің функционалдық бөлінуі 1-кестеде көрсетілген.

Кесте 1. Технология құрамдастарының функционалдық бөлінуі

Компонент	*Негізгі функция*	*Студенттің әрекеті*	*Нәтиже*
Виртуалды есептер жинағы	*Практикалық әрекетті ұйымдастыру*	*Тапсырмаларды таңдау және шешу*	*Есеп шығару дағдыларын қалыптастыру*
Чат-бот	*Контекстік қолдау*	*Сұрақ құрастыру, кеңесті талдау*	*Қиындықтарды жеңу*
Диагностикалық модуль	*Нәтижелерді бағалау*	*Диагностикалық тапсырмаларды орындау*	*Жетістіктер динамикасы*
Кері байланыс модулі	*Қабылдауды бағалау*	*Сауалнаманы толтыру*	*Қанағаттану*
Аналитикалық модуль	*Нәтижелерді жинақтау*	–	*Технологияны бағалау көрсеткіштері*



Сурет 2. Интеллектуалды қолдаудың үш деңгейі

Ұйымдастырушылық-әрекеттік компонент студенттің технологиямен өзара әрекеттесу ретін белгілейді. Бұл цикл оқу тапсырмасын алудан басталып, рефлексиямен аяқталады. Циклдің басты ерекшелігі мынада: студент қиындық туындаған жағдайда ЖИ-ге жүгінеді, бірақ мөлшерленген қолдау алғаннан кейін міндетті түрде өзіндік шешімге қайта оралады. Өзара әрекеттесу циклі 3-суретте көрсетілген.



Сурет 3. Студенттің технологиямен өзара әрекеттесу циклі

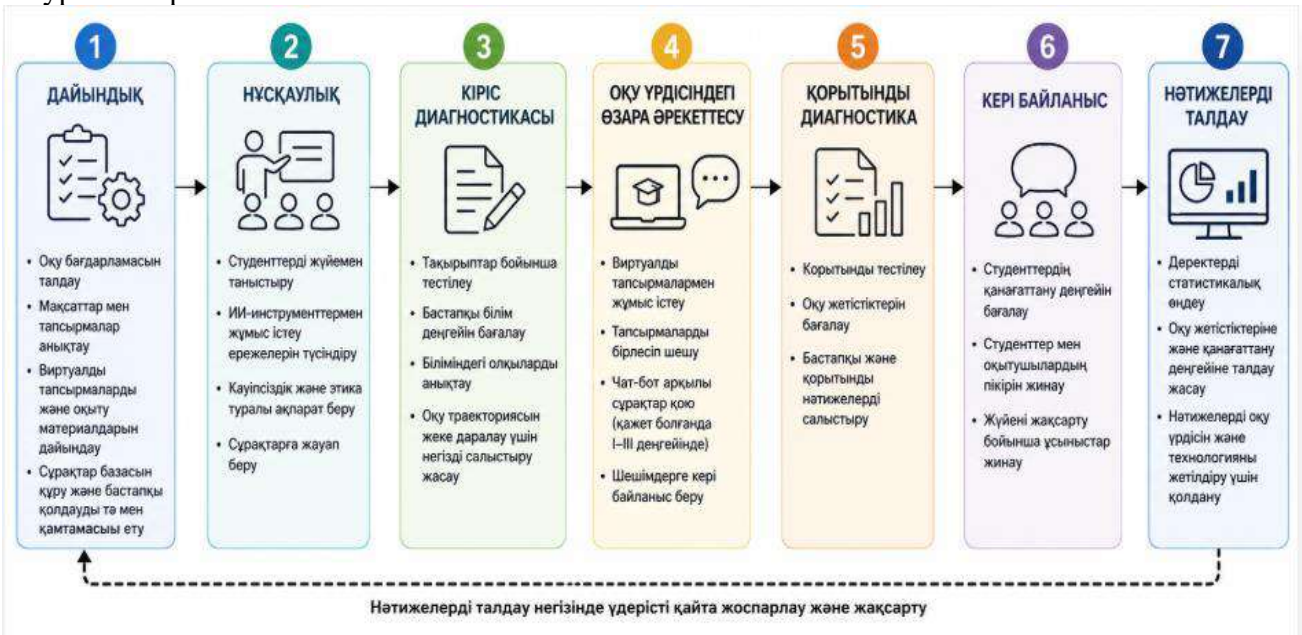
Ескерту: 1 – оқу тапсырмасы; 2 – өзіндік әрекет; 3 – қиындықты анықтау; 4 – ЖИ-ге жүгіну; 5 – мөлшерленген қолдау (1, 2 немесе 3-деңгей); 6 – шешуді жалғастыру; 7 – нәтижені тексеру; 8 – рефлексия.

Бағалау-нәтижелік компонент технологияны қолдану нәтижелерін кешенді бағалауға мүмкіндік береді. Бағалау екі бағытта жүргізіледі: біріншісі – студенттердің оқу жетістіктері, екіншісі – олардың қанағаттану деңгейі. Әрбір бағыт өзіне тән көрсеткіштермен және бағалау құралдарымен сипатталады. Бағалау жүйесінің құрылымы 2-кестеде келтірілген.

Кесте 2. Технологияны енгізу нәтижелерін бағалау жүйесі

Бағыт	*Көрсеткіштер*	*Құрал*
Оқу жетістіктері	*Жалпы нәтиже; концептуалды түсіну; есеп шығару; өзіндік дербестік*	*Диагностикалық тапсырмалар*
Қанағаттану	*Пайдалылық; өзара әрекеттесудің ыңғайлылығы; оқу әрекетіне әсері; жалпы қанағаттану*	*Сауалнама*

Технологияны қолдану нақты алгоритм негізінде жүзеге асырылады. Алгоритм дайындық кезеңінен басталып, нәтижелерді талдаумен аяқталады, ал әрбір кезең белгілі бір мақсатқа бағытталып, өзара логикалық байланыста тұрады. Технологияны қолданудың толық алгоритмі 4-суретте берілген.



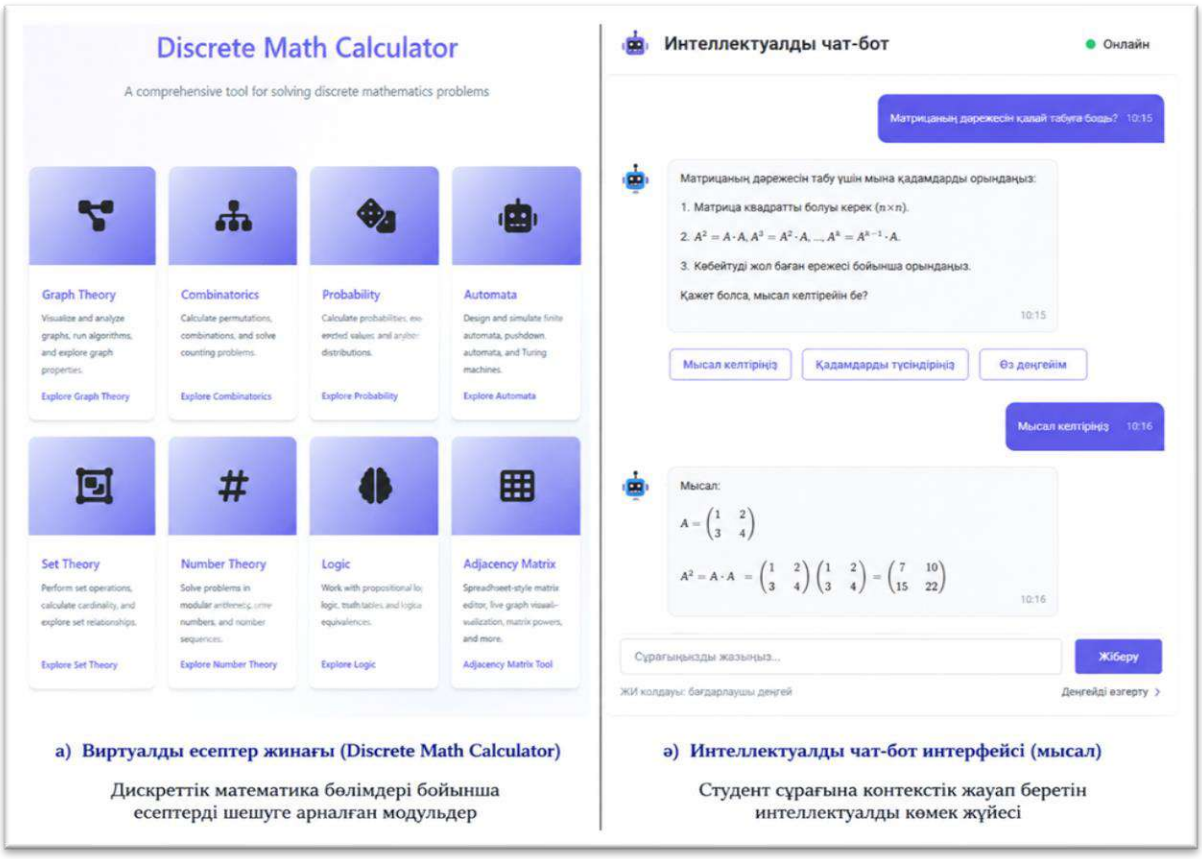
Сурет 4. Жасанды интеллект негізінде жоғары математиканы оқыту технологиясын қолдану алгоритмі

Ескерту: Алгоритм технологияны қолданудың барлық кезеңдерін қамтиды.

Виртуалды есептер жинағы мен интеллектуалды чат-боттың жұмыс істеу принципі 5-суретте көрсетілген. Виртуалды есептер жинағында студент тапсырманы таңдап, шешу барысын жүргізеді, ал чат-бот студенттің сұрағына жауап бере отырып, контекстік қолдау көрсетеді.

Студенттің технологияны қолдануға қанағаттануын анықтау мақсатында Лайкерттің бес балдық шкаласына негізделген сауалнама құрастырылды. Ол пайдалылық, ыңғайлылық, оқу әрекетіне әсері және жалпы қанағаттану деген төрт негізгі шкаланы қамтиды. Әрбір шкала бойынша ұсынылған тұжырымдар студенттің технология туралы пікірін анықтауға бағытталған. Сауалнама құрылымы 3-кестеде келтірілген.

Модельдік валидация барысында бастапқы және қорытынды диагностика нәтижелерінің сипаттамалық статистикалық көрсеткіштері есептелді. Әрбір диагностикалық параметр бойынша орташа мән (M), стандартты ауытқу (SD), медиана (Me) және квартильаралық ауқым (Q1–Q3) анықталды.



Сурет 5. Виртуалды есептер жинағы мен чат-боттың интерфейсі

Кесте 3. Студенттердің қанағаттануын бағалау құралының құрылымы

Шкала	*Бағаланатын аспект*	*Тұжырым мысалы*
Пайдалылық	*материалды түсінуге көмек*	*Технологияны қолдану оқу материалын жақсырақ түсінуге көмектеседі*
Ыңғайлылық	*өзара әрекеттесудің қарапайымдылығы*	*Виртуалды есептер жинағының құрылымы түсінікті*
Оқу әрекетіне әсері	*қателермен жұмыс және өзіндік дербестік*	*Технологияны қолдану жіберілген қателерді талдауға көмектеседі*
Жалпы қанағаттану	*технологияны қабылдау*	*Мен математиканың басқа бөлімдерін оқығанда осындай құралдарды қолданғым келеді*

Бұл көрсеткіштер модельдік деректер массивіндегі таралуды сипаттайды және аналитикалық хаттаманың осы типтегі деректерді өңдеуге қабілеттілігін тексеруге арналған. Сипаттамалық статистика 4-кестеде келтірілген.

Кесте 4.Модельдік деректер массивінің сипаттамалық статистикасы (балл)

Көрсеткіш	*Бастапқы M ± SD*	*Бастапқы Me [Q1–Q3]*	*Қорытынды M ± SD*	*Қорытынды Me [Q1–Q3]*
Жалпы нәтиже	*61,8 ± 14,5*	*63 [52–72]*	*74,2 ± 12,8*	*76 [66–84]*
Концептуалды түсіну	*60,3 ± 15,1*	*61 [50–71]*	*72,9 ± 13,4*	*74 [64–83]*
Шешу әдісін таңдау	*58,7 ± 16,0*	*59 [47–70]*	*71,5 ± 14,2*	*73 [62–82]*
Түрлендірулердің дұрыстығы	*62,4 ± 14,8*	*64 [53–73]*	*75,0 ± 13,1*	*77 [67–85]*

Келтірілген мәндер диагностикалық параметрлердің модельдік деректердегі таралуын көрсетеді және аналитикалық хаттаманың осы типтегі деректерді өңдеуге қабілетті екенін растайды. Аналитикалық хаттаманың жұмысқа қабілеттілігін тексеру үшін бірнеше статистикалық процедуралар қолданылды. Оларға жұптасқан t-критерий, Уилкоксон критерийі, әсер көлемін бағалау және Кронбах α-коэффициентін есептеу жатады. Әрбір процедураның қолданылу шарттары мен интерпретациясы 5-кестеде көрсетілген.

Кесте 5. Аналитикалық хаттаманы модельдік деректерде тексеру нәтижелері

Процедура	*Интерпретация*
Стьюденттің жұптастырылған t-критерийі	*Бастапқы және қорытынды диагностика нәтижелерін салыстыру (қалыпты үлестірім шарты орындалғанда)*
Уилкоксон критерийі	*Бастапқы және қорытынды диагностика нәтижелерін салыстыру (қалыпты үлестірім шарты орындалмағанда)*
Коэннің d коэффициенті	*Бақыланатын өзгерістердің әсер көлемін бағалау*
Кронбахтың α коэффициенті	*Сауалнама шкалаларының ішкі үйлесімділігін бағалау*

Модельдік валидация нәтижесі ұсынылған статистикалық талдау рәсімдерінің техникалық тұрғыдан қолдануға жарамдылығын растады. Сонымен қатар модельдік деректер негізінде алынған нәтижелер әзірленген әдістемелік модельдің педагогикалық тиімділігін бағалауға арналмаған және аналитикалық хаттаманың жұмысқа қабілеттілігін тексеру нәтижесі ретінде қарастырылады. Ұсынылған модельдің практикалық тиімділігін эмпирикалық тұрғыдан бағалау келесі зерттеу кезеңінде жоғары оқу орнының нақты білім беру үдерісінде пилоттық апробация жүргізу арқылы жүзеге асырылатын болады.

Сауалнаманың төрт шкаласы студенттің технологияны қалай қабылдағанын әр қырынан көрсетеді. Осы мақсатта әрбір шкала бойынша орташа мән, стандартты ауытқу, медиана және оң бағалардың үлесі анықталды (6-кесте). Төмендегі көрсеткіштер студенттердің нақты жауаптарын емес, әзірленген өлшеу құралының имитациялық деректердегі әрекетін сипаттайды.

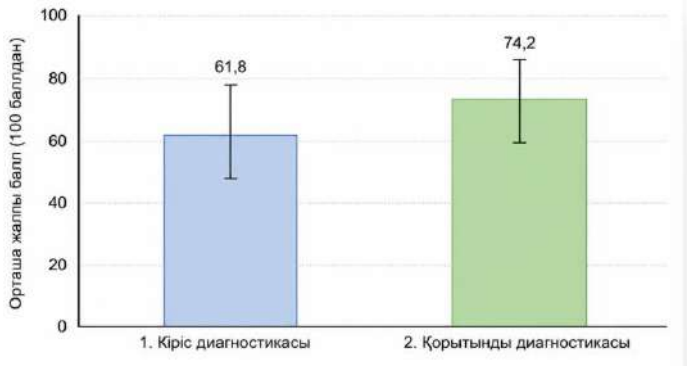
Кесте 6. Модельдік деректер массивіндегі қанағаттану көрсеткіштері

Көрсеткіш	*M ± SD*	*Me [Q1–Q3]*	*Оң бағалар үлесі, %*
Технологияның пайдалылығы	*4,24 ± 0,62*	*4,3 [4,0–4,8]*	*84,7*
Өзара әрекеттесудің ыңғайлылығы	*4,08 ± 0,68*	*4,0 [3,8–4,6]*	*79,2*
Оқу әрекетіне әсері	*4,17 ± 0,64*	*4,2 [3,9–4,7]*	*82,0*
Жалпы қанағаттану	*4,31 ± 0,59*	*4,4 [4,0–4,8]*	*87,5*

Модельдік (имитациялық) деректерді талдау барысында төрт шкала бойынша есептеу рәсімдерінің қолданылу мүмкіндігі тексерілді. Қалыптастырылған деректер массивінде жалпы қанағаттану шкаласының орташа мәні M = 4,31 (SD = 0,59), технологияның пайдалылығы – M = 4,24 (SD = 0,62), оқу әрекетіне әсері – M = 4,17 (SD = 0,64), ал өзара әрекеттесудің ыңғайлылығы – M = 4,08 (SD = 0,68) болды. Оң бағалар үлесі 79,2%-дан 87,5%-ға дейінгі аралықта қалыптастырылды. Бұл көрсеткіштер нақты студенттердің қанағаттану деңгейін сипаттамайды және әзірленген сауалнама нәтижелерін жинақтау мен статистикалық өңдеу рәсімдерінің жұмысқа қабілеттілігін көрсету мақсатында ғана пайдаланылды.

Аналитикалық хаттаманың жұмысын көрнекі көрсету үшін бастапқы және қорытынды диагностиканың болжамды нәтижесін құрылымын имитациялайтын модельдік деректер массиві құрылды, ал арасындағы айырмашылықты көрсету үшін орташа мән мен стандартты

ауытқу графигі салынды(6-сурет). Осы графикте оқу жетістігінің жалпы өсу үрдісі де, нәтиженің таралу ауқымы да көрінеді.



6-сурет - Кіріс және қорытынды диагностиканың жалпы балының орташа мәні (модельдік деректер). Ескерту: 1-бастапқы диагностикасы (M = 61,8; SD = 14,5); 2 – қорытынды диагностика (M = 74,2; SD = 12,8).

Мұндағы көрініс аналитикалық контурдың мүмкіндігін екі жағынан аңғартады: ол тек орташа балдың өзгеруін ғана емес, білім алушылар нәтижесінің таралу ауқымын да тіркей алады. Дегенмен эмпирикалық деректі талдағанда орташа мәнді салыстырудың өзі жеткіліксіз. Оны статистикалық маңыздылықты тексерумен және әсер мөлшерін бағалаумен толықтыру қажет. Осыған байланысты графикалық көрініс дербес дәлел емес, талдаудың көмекші элементі ретінде, статистикалық рәсімнің нәтижесімен бірге қарастырылады.

Дискуссия

Ұсынылған әдістемелік модель әр құрамдасының функциясын дербес бөліп көрсету арқылы жоғары математиканы оқытуда жасанды интеллектті қолданудың нақты жолын көрсетеді. Білім алушы генеративті жүйеге реттелмеген түрде жүгінгенде пайда болатын жағдайдан айырмашылығы мұнда анық: модельдегі виртуалды есептер жинағы математикалық практиканың дәйектілігі мен құрылымдылығын қамтамасыз етсе, чат-бот өлшенген контекстік қолдау рөлін атқарады. Мұндай бөліну Zawacki-Richter және т.б. (2019) авторлардың пікірімен үндеседі — олар да зерттеу назарын ЖИ-дің технологиялық мүмкіндігінен оны қолданудың нақ педагогикалық тетігіне бұру керектігін айтқан болатын.

Әзірленген модельдің маңызды тұсы ретінде интеллектуалды қолдаудың үш деңгейге бөлінуін айтуға болады. Олар бағдарлаушы, бағыттаушы және түсіндіруші. Көмек көлемінің біртіндеп артуы студенттің өз бетінше шешім іздеуін тоқтатпай, туындаған қиындықтардың сипатын ескеруге мүмкіндік береді. Ал қолдау алғаннан кейін студенттің өз жұмысына қайта оралуы міндетті болғандықтан, бұл модель дайын жауап беретін генеративті ЖИ-ден түбегейлі бөлек тұрады. Қазіргі зерттеулерде жиі айтылатын — білім алушының генеративті жүйеге шектен тыс тәуелденуі, жауапты сын көзбен қарамауы, дербес оқу белсенділігінің бәсеңдеуі сияқты қауіп ескерілгенде, дәл осы айырмашылықтың салмағы арта түседі[4,14,15].

Модельдің тағы бір ерекшелігі – оқу жетістіктері мен студенттердің қанағаттануын қатар алатын екі компонентті бағалау жүйесі. Егер тек академиялық көрсеткішке ғана сүйенсе, студенттің технологияны қаншалық пайдалы, ыңғайлы санайтыны және оның өзіндік жұмысқа қалай ықпал ететіні көмескі қалар еді. Ал пайдалылық, өзара әрекеттесудің ыңғайлылығы, оқу әрекетіне әсер ету дәрежесі мен жалпы қанағаттану шкаласын қосқанда, объективті диагностикалық көрсеткіш студенттің өз пікірімен толығады. Мұндай көзқарас білім беру технологиясының тиімділігі жалғыз нәтижемен емес, білім алушының цифрлық ортамен қалай әрекеттесетінімен де өлшенеді деген қазіргі түсінікке сай келеді[16,17].

Аналитикалық хаттаманың модельдік валидациясы қарастырылған деректің әртүрлі түрін өңдеуге мүмкіндік беретінін көрсетті: сипаттамалық статистиканы есептеу, байланысты өлшемдерді салыстыру, әсер мөлшерін бағалау және сауалнама шкалаларының ішкі үйлесімділігін тексеру рәсімдері модельдік массивте сәтті қолданылды. Алайда бұл нәтижелер әзірленген әдістемелік модельдің педагогикалық тиімділігін емес, аналитикалық құралдың техникалық жарамдылығын растайды.

Қазақстан Республикасының жоғары білім беру жүйесі тұрғысынан алғанда, ұсынылған модель ЖИ құралдарын математикалық пәндерге кіріктірудің қайта жаңғыртылатын тетігі ретінде қызығушылық тудырады: ол мазмұнды жобалауда, тапсырмалар банкін құруда, ЖИ-қолдаудың рұқсат етілген түрін белгілеуде және нәтижені бағалауда оқытушының жетекші рөлін сақтап қалады. Осылайша білім беру процесінің цифрлық трансформациясы академиялық адалдық талабымен, дербес деректі қорғаумен және студенттің өзіндік оқу белсенділігін сақтаумен қатар жүре алады.

Зерттеудің шектеулері бірнеше жайтпен байланысты: аналитикалық хаттама тек модельдік деректерде тексерілді; эмпирикалық апробация жүргізілген жоқ; зерттеу бір пән – жоғары математика – аясында шектелген. Аталған шектеулер модельдің практикалық құндылығын жоққа шығармайды, алайда оның педагогикалық тиімділігін эмпирикалық тұрғыдан растау келесі зерттеу кезеңінің міндеті болып қалады.

Қорытынды

Жүргізілген зерттеу ЖИ-технологияны жоғары математиканы оқытуда қолдануға арналған төрт блокты әдістемелік модельдің құрылымдық негіздемесін берді. Модельдің құрылымдық тұтастығы оның компоненттерінің функционалдық өзара байланысымен қамтамасыз етіледі. Интеллектуалды қолдаудың бағдарлаушы, бағыттаушы және түсіндіруші деңгейлері студентке дозаланған көмек алуға мүмкіндік берсе, ал қос компонентті бағалау жүйесі (оқу жетістіктері мен қанағаттану) нәтижелерді кешенді түрде қарастыруға жағдай жасайды. Модельдік валидация статистикалық өңдеу рәсімдерінің техникалық қолданылу мүмкіндігін растады. Әзірленген әдістемелік модель математикалық пәндерді оқытуда ЖИ құралдарын педагогикалық негізде енгізуге бағытталған әдіснамалық база ретінде жоғары оқу орындарының тәжірибесінде қолданыла алады. Зерттеудің келесі кезеңі нақты жоғары оқу орнының білім беру үдерісінде модельдің пилоттық апробациясын жүргізуге бағытталады.

Авторлардың үлесі:

Нурбеков Б.Ж. – зерттеу концептуализациясы, әдістемелік модельді және зерттеудің жалпы әдіснамасын әзірлеу, ғылыми жетекшілік;

Байарыстан А.О. – оқыту технологиясының құрылымын, диагностикалық құралдарды және нәтижелерді бағалау жүйесін әзірлеу;

Майдисарова Д.С. – ғылыми әдебиеттер мен нормативтік-әдістемелік дереккөздерді талдау, модельдің ұйымдастырушылық-әрекеттік құрамдасын әзірлеу;

Сантеева С.Ә. – аналитикалық хаттаманы әзірлеу, нәтижелерді жүйелеу және түсіндіру, деректерді визуализациялау;

Толегенова Г.Б. – редакциямен хат-хабар алысу, мақаланы басылымға дайындау және ұсыну.

Барлық авторлар нәтижелерді талқылауға, қолжазбаны редакциялауға қатысты, оның түпкілікті нұсқасын оқып, мақұлдады.

Алғыс

Зерттеу Қазақстан Республикасы Ғылым және жоғары білім министрлігі Ғылым комитетінің 2024–2026 жылдарға арналған ғылыми және (немесе) ғылыми-техникалық жобаларды гранттық қаржыландыру аясында жүзеге асырылатын AP23489530 «Болашақ АТ-кадрларының мансаптық перспективаларын арттыру үшін жасанды интеллектті (AI) пайдалана отырып, жоғары математиканы оқытудың инновациялық технологиясын зерттеу және әзірлеу» атты ғылыми жобасы шеңберінде орындалды.

Пайланаылған дереккөздер тізімі

[1]Ouyang F., Zheng L., Jiao P. Artificial intelligence in online higher education: A systematic review of empirical research from 2011 to 2020 // Educ Inf Technol. 2022. Т. 27, № 6. С. 7893–7925. DOI: 10.1007/s10639-022-10925-9.

[2] Zawacki-Richter O. и др. Systematic review of research on artificial intelligence applications in higher education – where are the educators? // Int J Educ Technol High Educ. 2019. Т. 16, № 1. С. 39. DOI: 10.1186/s41239-019-0171-0.

[3]BaiDoo-Anu D., Owusu Ansah L. Education in the Era of Generative Artificial Intelligence (AI): Understanding the Potential Benefits of ChatGPT in Promoting Teaching and Learning // Journal of AI. 2023. Т. 7, № 1. С. 52–62. DOI: 10.61969/jai.1337500.

[4] Alasadi E. A., Baiz C. R. Generative AI in Education and Research: Opportunities, Concerns, and Solutions // J. Chem. Educ. 2023. Т. 100, № 8. С. 2965–2971. DOI: 10.1021/acs.jchemed.3c00323.

[5] Kasneci E. и др. ChatGPT for good? On opportunities and challenges of large language models for education // Learning and Individual Differences. 2023. Т. 103. С. 102274. DOI: 10.1016/j.lindif.2023.102274.

[6] VanLEHN K. The Relative Effectiveness of Human Tutoring, Intelligent Tutoring Systems, and Other Tutoring Systems // Educational Psychologist. 2011. Т. 46, № 4. С. 197–221. DOI: 10.1080/00461520.2011.611369.

[7] Steenbergen-Hu S., Cooper H. A meta-analysis of the effectiveness of intelligent tutoring systems on college students' academic learning. // Journal of Educational Psychology. 2014. Т. 106, № 2. С. 331–347. DOI: 10.1037/a0034752.

[8] Sun S. и др. The Effects of ALEKS on Mathematics Learning in K-12 and Higher Education: A Meta-Analysis // Investigations in Mathematics Learning. 2021. Т. 13, № 3. С. 182–196. DOI: 10.1080/19477503.2021.1926194.

[9] Wu R., Yu Z. Do AI chatbots improve students learning outcomes? Evidence from a meta-analysis // Brit J Educational Tech. 2024. Т. 55, № 1. С. 10–33. DOI: 10.1111/bjet.13334.

[10]Cheng L. и др. Facilitating Student Learning With a Chatbot in an Online Math Learning Platform // Journal of Educational Computing Research. 2024. Т. 62, № 4. С. 687–717. DOI: 10.1177/07356331241226592.

[11]Pardos Z. A., Bhandari S. ChatGPT-generated help produces learning gains equivalent to human tutor-authored help on mathematics skills // PLoS ONE / под ред. Carvalho P. F. 2024. Т. 19, № 5. С. e0304013. DOI: 10.1371/journal.pone.0304013.

[12]Okonkwo C. W., Ade-Ibijola A. Chatbots applications in education: A systematic review // Computers and Education: Artificial Intelligence. 2021. Т. 2. С. 100033. DOI: 10.1016/j.caeai.2021.100033.

[13]Қазақстан Республикасының Заңы. (2013). Дербес деректер және оларды қорғау туралы: Қазақстан Республикасының 2013 жылғы 21 мамырдағы № 94-V Заңы (өзгерістер мен толықтырулармен). Қазақстан Республикасының Заңы. (2013). https://old.adilet.zan.kz/kaz/docs/Z1300000094

14.Lo C. K. What Is the Impact of ChatGPT on Education? A Rapid Review of the Literature // Education Sciences. 2023. Т. 13, № 4. С. 410. DOI: 10.3390/educsci13040410.

15.Bastani H. и др. Generative AI without guardrails can harm learning: Evidence from high school mathematics // Proc. Natl. Acad. Sci. U.S.A. 2025. Т. 122, № 26. С. e2422633122. DOI: 10.1073/pnas.2422633122.

16.Chiu T. K. F. и др. Systematic literature review on opportunities, challenges, and future research recommendations of artificial intelligence in education // Computers and Education: Artificial Intelligence. 2023. Т. 4. С. 100118. DOI: 10.1016/j.caeai.2022.100118.

17.Tlili A. и др. What if the devil is my guardian angel: ChatGPT as a case study of using chatbots in education // Smart Learn. Environ. 2023. Т. 10, № 1. С. 15. DOI: 10.1186/s40561-023-00237-x.

ҒТАХР 14.35.09

https://doi.org/10.51889/2959-5894.2026.95.3.027

А.Е. Сағымбаева[1], Қ.Ж. Сабраев[1*]

[1]Абай атындағы Қазақ ұлттық педагогикалық университеті, Алматы қ., Қазақстан
[*]*e-mail: kanatt82@gmail.com*

ЖАСАНДЫ ИНТЕЛЛЕКТ ЖАҒДАЙЫНДА БОЛАШАҚ ИНФОРМАТИКА МҰҒАЛІМДЕРІНІҢ ЛОГИКАЛЫҚ ОЙЛАУ ДАҒДЫЛАРЫН ҚАЛЫПТАСТЫРУДАҒЫ ПЕДАГОГИКАЛЫҚ МӘСЕЛЕЛЕР

Аңдатпа

Мақалада жасанды интеллект жағдайында болашақ информатика мұғалімдерінің логикалық ойлау дағдыларын қалыптастыру барысында туындайтын педагогикалық мәселелер қарастырылады. Зерттеудің мақсаты – болашақ информатика мұғалімдерінің логикалық ойлау дағдыларының информатика мазмұнын меңгеру барысында көрінетін пәндік әрекеттерін жүйелеу, жасанды интеллект жағдайында аталған дағдыларды қалыптастыруда туындайтын педагогикалық мәселелерді анықтау және олардың пайда болу себептері мен шешу жолдарының өзара байланысын негіздеу. Зерттеу барысында ғылыми әдебиеттерді талдау, салыстырмалы талдау, теориялық жалпылау, жүйелеу және педагогикалық интерпретациялау әдістері қолданылды. Логикалық ойлау дағдыларының оқу әрекетіндегі көріністері ойлау операциялары мен логикалық пайымдау әрекеттерінің өзара байланысы негізінде жүйеленді. Информатика мазмұнын талдау нәтижесінде олардың алгоритмдеу және программалау, ақпараттық модельдеу, деректермен жұмыс, компьютерлік жүйелер, желілік технологиялар, ақпараттық қауіпсіздік және компьютерлік графика бағыттарындағы пәндік көріністері айқындалды. Жасанды интеллект құралдарының аталған логикалық әрекеттерді қолдау немесе толық не ішінара орындау мүмкіндіктерін білім алушының дербес логикалық әрекетінің сақталуымен өзара байланыста талдау нәтижесінде үш негізгі педагогикалық мәселе анықталды: білім алушының дербес логикалық әрекетін сақтау; жасанды интеллект ұсынған нәтижелерді дербес тексеру және олардың негізінде негізделген қорытынды жасау; болашақ информатика мұғалімдерін жасанды интеллект құралдарын педагогикалық мақсатқа сәйкес қолдануға даярлау. Анықталған мәселелердің пайда болу себептері мен оларды шешудің педагогикалық жолдары жүйеленді. Зерттеу нәтижелері жасанды интеллект құралдарын логикалық ойлау дағдыларын қалыптастыруда қолданудың педагогикалық мәнін олардың технологиялық мүмкіндіктерімен ғана емес, білім алушының дербес логикалық әрекетін қолдау немесе алмастыру сипаты тұрғысынан қарастыру қажеттігін көрсетеді.

Түйін сөздер: жасанды интеллект, болашақ информатика мұғалімдері, логикалық ойлау дағдылары, логикалық әрекет, логикалық пайымдау, информатиканы оқыту, педагогикалық мәселелер.

А.Е. Сагимбаева[1], К.Ж. Сабраев[1],
[1]Казахский национальный педагогический университет имени Абая, г. Алматы, Казахстан

ПЕДАГОГИЧЕСКИЕ ПРОБЛЕМЫ ФОРМИРОВАНИЯ НАВЫКОВ ЛОГИЧЕСКОГО МЫШЛЕНИЯ БУДУЩИХ УЧИТЕЛЕЙ ИНФОРМАТИКИ В УСЛОВИЯХ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

Аннотация

В статье рассматриваются педагогические проблемы, возникающие в процессе формирования навыков логического мышления будущих учителей информатики в условиях искусственного интеллекта. Цель исследования – систематизировать предметные действия, в которых проявляются навыки логического мышления будущих учителей информатики при освоении содержания информатики, выявить педагогические проблемы, возникающие в процессе формирования данных навыков в условиях искусственного интеллекта, и обосновать взаимосвязь причин их возникновения и путей решения. В исследовании использовались методы анализа научной литературы, сравнительного анализа, теоретического обобщения, систематизации и педагогической интерпретации. Проявления навыков логического мышления в учебной деятельности систематизированы на основе взаимосвязи мыслительных операций и действий логического рассуждения. В результате анализа содержания информатики определены их предметные проявления в таких направлениях, как алгоритмизация и программирование, информационное моделирование, работа с данными, компьютерные системы, сетевые технологии, информационная безопасность и компьютерная графика. В результате анализа возможностей средств искусственного интеллекта поддерживать, полностью или частично выполнять указанные логические действия во взаимосвязи с сохранением самостоятельной логической деятельности обучающегося выявлены три основные педагогические проблемы: сохранение самостоятельной логической деятельности обучающегося; самостоятельная проверка результатов, предложенных искусственным интеллектом, и формулирование на их основе обоснованных выводов; подготовка будущих учителей информатики к использованию средств

искусственного интеллекта в соответствии с педагогическими целями. Систематизированы причины возникновения выявленных проблем и педагогические пути их решения. Результаты исследования показывают необходимость рассматривать педагогическое значение использования средств искусственного интеллекта при формировании навыков логического мышления не только с точки зрения их технологических возможностей, но и с позиции характера поддержки или замещения самостоятельной логической деятельности обучающегося.

Ключевые слова: искусственный интеллект, будущие учителя информатики, навыки логического мышления, логическое действие, логическое рассуждение, обучение информатике, педагогические проблемы.

A.E. Sagymbayeva[1], K.Zh. Sabrayev[1]
[1]Abai Kazakh National Pedagogical University, Almaty, Kazakhstan

PEDAGOGICAL ISSUES IN DEVELOPING LOGICAL THINKING SKILLS OF FUTURE INFORMATICS TEACHERS IN THE CONTEXT OF ARTIFICIAL INTELLIGENCE

Abstract

The article examines pedagogical issues arising in the development of logical thinking skills among future informatics teachers in the context of artificial intelligence. The aim of the study is to systematize the subject-specific activities through which future informatics teachers' logical thinking skills are manifested when mastering informatics content, to identify pedagogical issues arising in the development of these skills in the context of artificial intelligence, and to substantiate the relationship between their underlying causes and possible solutions. The study employed methods of scientific literature analysis, comparative analysis, theoretical generalization, systematization, and pedagogical interpretation. The manifestations of logical thinking skills in learning activities were systematized based on the interrelationship between thinking operations and logical reasoning activities. The analysis of informatics content identified their subject-specific manifestations in such areas as algorithmization and programming, information modeling, data processing, computer systems, network technologies, information security, and computer graphics. An analysis of the capacity of artificial intelligence tools to support or fully or partially perform these logical activities, in relation to preserving learners' independent logical activity, revealed three key pedagogical issues: preserving learners' independent logical activity; ensuring that learners independently verify AI-generated results and formulate reasoned conclusions based on them; and preparing future informatics teachers to use artificial intelligence tools in accordance with pedagogical goals. The causes underlying the identified issues and pedagogical approaches to addressing them were systematized. The findings indicate that the pedagogical significance of using artificial intelligence tools in the development of logical thinking skills should be considered not only in terms of their technological capabilities but also in terms of whether they support or substitute learners' independent logical activity.

Keywords: artificial intelligence, future informatics teachers, logical thinking skills, logical activity, logical reasoning, informatics education, pedagogical issues

Кіріспе

Қазіргі кезеңде жасанды интеллект технологияларының қарқынды дамуы білім беру үдерісінің мазмұны мен ұйымдастырылуын, сондай-ақ болашақ мамандардың кәсіби дайындығына қойылатын талаптарды айтарлықтай өзгертуде. Жасанды интеллект құралдарының ақпаратты өңдеу, мәселені талдау және шешім ұсыну сияқты интеллектуалдық әрекеттерді орындау мүмкіндігінің кеңеюі білім алушылардың дайын нәтижені қабылдаумен шектелмей, оны тексеру, салыстыру және негіздеу сияқты логикалық әрекеттерді өздігінен орындауының маңыздылығын арттырады. Мұндай логикалық әрекеттер информатиканы оқытуда ерекше мәнге ие, өйткені пән мазмұнын меңгеру проблеманы талдау, логикалық байланыстарды анықтау, шешу жолдарын іздеу, оларды салыстыру және негізделген қорытынды жасау сияқты логикалық әрекеттерді талап етеді. Болашақ информатика мұғалімі аталған әрекеттерді өзі орындаумен қатар, оларды орта мектепте білім алушылардың оқу әрекетінде ұйымдастыра білуі қажет. Осы тұрғыдан болашақ информатика мұғалімдерінің логикалық ойлау дағдыларын қалыптастыру олардың кәсіби даярлығының маңызды құрамдас бөлігі болып табылады.

Болашақ информатика мұғалімдерінің логикалық ойлау дағдыларын қалыптастыру алгоритмдеу және программалау, ақпараттық модельдеу, деректермен жұмыс, компьютерлік жүйелер, желілік технологиялар, ақпараттық қауіпсіздік, компьютерлік графика сияқты информатиканың мазмұндық бағыттарын меңгерумен тығыз байланысты. Бұл бағыттарды меңгеру мәселені құрылымдау, ақпаратты талдау және жүйелеу, логикалық байланыстарды

анықтау, шешімдерді салыстыру және негіздеу сияқты логикалық әрекеттерді жүзеге асыруды талап етеді.

Жасанды интеллект құралдарының білім беру үдерісіне енуі аталған логикалық әрекеттерді жүзеге асырудың жаңа мүмкіндіктерін қалыптастыруда. Жасанды интеллект құралдары мәселені талдауға, шешу жолдарын ұсынуға, алгоритмдерді салыстыруға, күрделі тапсырмаларды кезең-кезеңімен түсіндіруге және жедел кері байланыс алуға мүмкіндік береді [1, 2]. Мұндай мүмкіндіктер білім алушылардың баламалы шешімдерді қарастыруына, оларды салыстыруына және өз шешімдерін талдауына жағдай жасайды.

Қазіргі зерттеулерде жасанды интеллект құралдарының білім алушылардың логикалық ойлауы мен мәселе шешу дағдыларын дамытудағы мүмкіндіктері көрсетілген [3]. Сонымен қатар оларды білім беру үдерісінде қолдануға байланысты бірқатар педагогикалық тәуекелдер де қарастырылады [1, 2, 4, 5]. Жасанды интеллект ұсынған дайын жауаптар мен шешімдерге шамадан тыс сүйену білім алушылардың тапсырманы өздігінен талдау, шешу жолдарын іздеу, ақпаратты тексеру және өз шешімін негіздеу белсенділігінің төмендеуіне әкелуі мүмкін [1, 2, 4, 5]. Сонымен қатар болашақ мұғалімдердің жасанды интеллект құралдарын оқу мақсаттарына сәйкес педагогикалық тұрғыдан тиімді қолдануға даярлығы да маңызды мәселе ретінде қарастырылады.

Осылайша, жасанды интеллект құралдарын қолдану бір жағынан білім алушылардың логикалық әрекеттерін қолдауға мүмкіндік берсе, екінші жағынан дайын шешімдерге шамадан тыс сүйену олардың дербес ойлау белсенділігінің әлсіреу қаупін туындатады. Бұл қарама-қайшылық жасанды интеллект мүмкіндіктерін пайдалану мен білім алушының дербес логикалық әрекетін сақтаудың педагогикалық теңгерімін қамтамасыз ету қажеттілігін көрсетеді. Осы тұрғыдан негізгі мәселе жасанды интеллект құралдарын қолданудың өзінде емес, оларды білім алушының логикалық әрекетін алмастырмай, оны белсендіруге бағыттап педагогикалық тұрғыдан ұйымдастыруда болып табылады.

Ғылыми еңбектерде жасанды интеллекттің білім алушылардың ойлау әрекеттеріне ықпалы, оны қолданудың педагогикалық мүмкіндіктері мен тәуекелдері әртүрлі қырынан қарастырылғанымен, болашақ информатика мұғалімдерінің логикалық ойлау дағдыларын қалыптастыру тұрғысынан информатиканың мазмұндық бағыттары мен оларға тән логикалық әрекеттердің өзара байланысы жеткілікті деңгейде жүйеленбеген. Сонымен қатар жасанды интеллект жағдайында аталған дағдыларды қалыптастыру барысында туындайтын педагогикалық мәселелер, олардың пайда болу себептері мен шешу жолдары біртұтас жүйе ретінде жеткілікті деңгейде қарастырылмаған.

Зерттеудің мақсаты – болашақ информатика мұғалімдерінің логикалық ойлау дағдыларының информатика мазмұнын меңгеру барысында көрінетін пәндік әрекеттерін жүйелеу, жасанды интеллект жағдайында аталған дағдыларды қалыптастыру барысында туындайтын педагогикалық мәселелерді анықтау және олардың пайда болу себептері мен шешу жолдарының өзара байланысын негіздеу.

Зерттеу міндеттері:

-информатика мазмұнын меңгеру барысында болашақ информатика мұғалімдерінің логикалық ойлау дағдыларының пәндік көріністерін жүйелеу;

-жасанды интеллект жағдайында болашақ информатика мұғалімдерінің логикалық ойлау дағдыларын қалыптастыруда туындайтын педагогикалық мәселелер мен олардың себептерін анықтау;

-анықталған педагогикалық мәселелердің себептері мен оларды шешу жолдарының өзара байланысын негіздеу.

Әдебиеттерге шолу

Логикалық ойлау дағдыларын қалыптастыру мәселесін қарастыру олардың негізінде жатқан ойлау әрекеттері мен пайымдау үдерісінің ерекшеліктерін анықтауды талап етеді. С.Л. Рубинштейн ойлаудың негізгі операцияларына салыстыруды, талдау мен жинақтауды,

абстракциялау мен жалпылауды жатқызып, оларды заттар мен құбылыстардың мәнді байланыстары мен қатынастарын ашуға бағытталған өзара байланысты ойлау әрекеттері ретінде қарастырады [6]. Johnson-Laird, P. N. пайымдауды қорытынды жасауға алып келетін жүйелі ойлау үдерісі ретінде сипаттайды және дедуктивті пайымдауда қорытындының алғышарттардан логикалық тұрғыдан туындауына назар аударады [7].

Логикалық ойлау мәселесі қазіргі зерттеулерде пайымдау үдерісінің формалды және күнделікті міндеттерді шешудегі көріністерімен байланысты қарастырылады. Bronkhorst және т.б. логикалық пайымдауды білім алушылардың формалды және күнделікті міндеттерді шешу барысында қолданатын пайымдау стратегиялары тұрғысынан зерттеп, оның сыни ойлауды дамытудағы маңызын көрсетеді [8]. Информатиканы оқыту контексінде логикалық ойлау дағдылары нақты оқу міндеттерін орындау барысында көрінеді. Мәселен, Saygıner және Tüzün программалауды оқыту барысында білім алушылардың логикалық ойлау дағдыларының дамуын арнайы зерттеген [9].

Осы зерттеулер логикалық ойлау дағдыларын информатиканы меңгеру барысында жүзеге асырылатын оқу әрекеттерімен өзара байланыста қарастырудың маңызын көрсетеді. Информатика мазмұнын меңгеру барысында мәселені талдау, ақпарат пен объектілер арасындағы байланыстарды анықтау, шешу жолдарын салыстыру, тексеру және негіздеу сияқты логикалық әрекеттер жүзеге асырылады. Болашақ информатика мұғалімдерін даярлау барысында аталған логикалық әрекеттерді олардың өз оқу әрекетінде жүзеге асыруымен қатар, білім алушылардың оқу әрекетін ұйымдастыру тұрғысынан да қарастыру қажет. Жасанды интеллект құралдарының оқу үдерісіне енуі осы логикалық әрекеттерді жүзеге асырудың жаңа мүмкіндіктерімен қатар, оларды білім алушының дербес орындауын қамтамасыз ету мәселесін де өзекті етеді.

Жасанды интеллект құралдарын білім беру үдерісінде қолданудың педагогикалық мүмкіндіктері мен тәуекелдері бірқатар зерттеулерде қарастырылған. Kasneci және т.б. [1] үлкен тілдік модельдердің күрделі оқу материалдарын түсіндіру, жекелендірілген қолдау және жедел кері байланыс беру мүмкіндіктерін атап көрсетумен қатар, олардың жауаптарын сыни бағалау және тексеру қажеттілігіне назар аударады. Tlili және т.б. [2] ChatGPT-ті білім беруде қолдануда ақпараттың сенімділігі, академиялық адалдық және жауапты пайдалану мәселелерін көрсетеді. UNESCO материалдарында да жасанды интеллекттің білім алушыларға қолдау көрсету мүмкіндіктерімен қатар, оны педагогикалық тұрғыдан негізсіз қолданудың білім алушылардың дербестігі мен сыни рефлексиясына кері әсер ету қаупі көрсетіліп, мұғалімнің жетекші рөлін сақтау қажеттілігі атап өтіледі [4]. Осы тұрғыдан жасанды интеллект құралдарының педагогикалық әлеуеті олардың технологиялық мүмкіндіктерімен ғана емес, оқу үдерісінде мақсатты әрі жауапты қолданылуымен және білім алушының дербес ойлау әрекетінің сақталуымен де байланысты.

Жасанды интеллект құралдарының болашақ информатика мұғалімдерінің логикалық ойлау дағдыларына ықпалын анықтауға бағытталған зерттеулер де бар. Bokan және т.б. [3] жасанды интеллект құралдарын қолданудың болашақ информатика мұғалімдерінің логикалық ойлау дағдыларына ықпалын эксперименттік тұрғыдан зерттеген. Қазақстан жоғары оқу орнының 48 студенті қатысқан зерттеуде эксперименттік және бақылау топтарының нәтижелері салыстырылып, жасанды интеллект құралдарын қолданған эксперименттік топта логикалық ойлау және мәселе шешу көрсеткіштерінің бақылау тобымен салыстырғанда жоғарылағаны анықталған. Авторлар алынған нәтижелер негізінде жасанды интеллект құралдарын болашақ информатика мұғалімдерін даярлау үдерісіне мақсатты енгізудің олардың логикалық ойлау және мәселе шешу дағдыларын дамытудағы әлеуетін көрсетеді [3].

Жасанды интеллект құралдарының аталған педагогикалық әлеуетін іске асыру болашақ мұғалімдердің оларды оқу мақсаттарына сәйкес қолдануға даярлығымен де байланысты. Holmes, Bialik және Fadel [10] жасанды интеллектті білім беруге енгізу оқыту құралдарын өзгертумен ғана шектелмей, білім берудің мақсаттары мен педагогтің рөлін қайта қарастыруды қажет ететініне назар аударады. Luckin және т.б. [11] жасанды интеллект

құралдарын тиімді қолдану үшін мұғалімдердің олардың мүмкіндіктері мен шектеулерін түсінуі және технологияларды нақты педагогикалық мақсаттарға сәйкес қолдана білуі қажеттігін көрсетеді. Осы тұрғыдан болашақ информатика мұғалімдерін даярлауда жасанды интеллект құралдарын техникалық тұрғыдан меңгертумен қатар, оларды пән мазмұны мен оқу мақсаттарына сәйкес педагогикалық тұрғыдан қолдануға даярлау маңызды.

Қарастырылған зерттеулер жасанды интеллект құралдарының логикалық ойлау дағдыларын дамытудағы мүмкіндіктерімен қатар, оларды қолдану барысында білім алушының дербес логикалық әрекетін сақтау және болашақ мұғалімдерді осы құралдарды педагогикалық мақсатқа сәйкес қолдануға даярлау қажеттілігін көрсетеді. Бұл информатика мазмұнын меңгеру барысында жүзеге асырылатын логикалық әрекеттер мен жасанды интеллект жағдайында туындайтын педагогикалық мәселелерді, олардың себептері мен шешу жолдарын өзара байланыста жүйелеу қажеттілігін негіздейді.

Зерттеу әдіснамасы

Зерттеу жұмысы жасанды интеллект жағдайында болашақ информатика мұғалімдерінің логикалық ойлау дағдыларын қалыптастыру барысында туындайтын педагогикалық мәселелерді теориялық-аналитикалық тұрғыдан анықтауға және олардың шешу жолдарын негіздеуге бағытталды. Зерттеудің әдіснамалық негізі информатиканы оқыту барысында жүзеге асырылатын логикалық әрекеттерді, жасанды интеллект құралдарының оларды қолдаудағы педагогикалық мүмкіндіктері мен тәуекелдерін және мұғалімнің осы үдерісті педагогикалық тұрғыдан ұйымдастырудағы рөлін өзара байланыста қарастыруға негізделді.

Зерттеу барысында ғылыми әдебиеттерді талдау, салыстырмалы талдау, теориялық жалпылау, жүйелеу және педагогикалық интерпретациялау әдістері қолданылды. Ғылыми әдебиеттерді талдау негізінде логикалық ойлау дағдыларын қалыптастыру, жасанды интеллект құралдарының білім алушылардың ойлау әрекеттеріне ықпалы, оларды білім беру үдерісінде қолданудың педагогикалық мүмкіндіктері мен тәуекелдері, сондай-ақ мұғалімдердің жасанды интеллект құралдарын педагогикалық мақсатта қолдануға даярлығы мәселелеріне арналған зерттеулер қарастырылды.

Информатика мазмұнын талдау және жүйелеу арқылы оны меңгеру барысында жүзеге асырылатын ойлау операциялары мен логикалық пайымдау әрекеттері қарастырылып, олардың пәндік көріністері айқындалды. Жасанды интеллект жағдайында туындайтын педагогикалық мәселелерді айқындау үш өзара байланысты өлшем негізінде жүзеге асырылды: 1) логикалық ойлау дағдыларының информатика мазмұнын меңгеру барысында көрінетін пәндік әрекеттері; 2) осы әрекеттерді орындауда жасанды интеллект құралдарының білім алушыға қолдау көрсету немесе оларды толық не ішінара орындау мүмкіндіктері; 3) жасанды интеллектті қолдану барысында білім алушының дербес логикалық әрекетінің сақталу сипаты. Осы өлшемдерді өзара байланысты талдау жасанды интеллект құралдарын қолдану барысында білім алушының логикалық әрекеттерінің қай жағдайда қолдау табатынын және қай жағдайда олардың орындалуы жасанды интеллектке берілуі мүмкін екенін қарастыруға мүмкіндік берді. Бұл талдау болашақ информатика мұғалімдерінің логикалық ойлау дағдыларын қалыптастыруда туындайтын педагогикалық мәселелер мен олардың пайда болу себептерін айқындауға негіз болды. Анықталған педагогикалық мәселелер мен олардың себептерін педагогикалық интерпретациялау және теориялық жалпылау негізінде білім алушының дербес логикалық әрекетін сақтауға және белсендіруге бағытталған шешу жолдары жүйеленді.

Зерттеудің дереккөздік базасын логикалық ойлау дағдыларын қалыптастыру, информатиканы оқыту, жасанды интеллектті білім беруде қолдану және педагогтердің жасанды интеллект құралдарын педагогикалық мақсатта қолдануға даярлығы мәселелеріне арналған ғылыми еңбектер, сондай-ақ UNESCO материалдары құрады. Зерттеу нәтижелерінің негізділігі ғылыми дереккөздердегі тұжырымдарды салыстыру, теориялық талдау және алынған нәтижелерді өзара байланыста жүйелеу арқылы қамтамасыз етілді.

Зерттеу нәтижелері

Зерттеу аясында болашақ информатика мұғалімдерінің логикалық ойлау дағдылары оқу міндеттерін орындау барысында жүзеге асырылатын ойлау операциялары мен логикалық пайымдау әрекеттері арқылы қарастырылды. Мұндай жүйелеуге С.Л. Рубинштейннің ойлаудың негізгі операциялары туралы тұжырымы және H. Bronkhorst және т.б. ұсынған логикалық пайымдау үдерісінің сипаттамасы теориялық негіз болды. С.Л. Рубинштейн ойлаудың негізгі операцияларына талдау мен жинақтауды, салыстыруды, абстракциялау мен жалпылауды жатқызса [6], H. Bronkhorst және т.б. логикалық пайымдауды берілген контекстегі ақпаратты іріктеу және интерпретациялау, байланыстарды анықтау, тексеру және қорытынды жасау үдерісі ретінде қарастырады [8]. Осы тұрғыдан логикалық пайымдау ақпаратты жай қабылдаумен шектелмей, оның элементтері арасындағы байланыстарды анықтау, алынған тұжырымдарды тексеру және негізделген қорытындыға келуге бағытталған ойлау үдерісін сипаттайды. Осы теориялық тұжырымдарды негізге ала отырып, логикалық ойлау дағдыларының оқу әрекетіндегі көріністері екі өзара байланысты топқа жүйеленді: бірінші топты ойлау операциялары – талдау мен жинақтау, салыстыру, абстракциялау және жалпылау; екінші топты логикалық пайымдау әрекеттері – ақпаратты интерпретациялау, байланыстарды анықтау, тексеру және қорытынды жасау құрады (1-сурет).



Сурет 1. Логикалық ойлау дағдыларының оқу әрекетіндегі көріністері

1-суретте логикалық ойлау дағдыларының көріністері ретінде жүйеленген ойлау операциялары мен логикалық пайымдау әрекеттері информатиканың мазмұнын меңгеру барысында нақты пәндік сипатта көрінеді. Әрбір мазмұндық бағытқа тән оқу міндеттерін орындау барысында аталған ойлау операциялары мен логикалық пайымдау әрекеттері өзара байланыста жүзеге асырылады. Мысалы, алгоритмдеу және программалауда мәселені талдау, шешу тәсілдерін салыстыру, алгоритмнің орындалу логикасындағы байланыстарды анықтау және нәтижені тексеру жүзеге асырылса, ақпараттық модельдеуде объектінің мәнді қасиеттерін бөліп алу, абстракциялау, модельдер арасындағы байланыстарды анықтау және салыстыру көрініс табады. Деректермен жұмыс барысында ақпаратты салыстыру, жүйелеу, заңдылықтарды анықтау және олардың негізінде қорытынды жасау әрекеттері жүзеге асырылады.

Информатиканың мазмұндық бағыттарын айқындауда Қазақстан Республикасының негізгі орта білім беру деңгейіне арналған «Информатика» оқу пәнінің үлгілік оқу бағдарламасының мазмұны ескерілді [12]. Бағдарлама мазмұнын зерттеу мақсаты тұрғысынан талдау негізінде логикалық ойлау дағдыларының пәндік көріністерін қарастыру үшін алгоритмдеу және программалау, ақпараттық модельдеу, деректермен жұмыс, компьютерлік жүйелер, желілік технологиялар, ақпараттық қауіпсіздік және компьютерлік графика бағыттары бөліп алынды. Осы бағыттар мен логикалық ойлау дағдыларының пәндік көріністері арасындағы байланыс 1-кестеде көрсетілді.

Кесте 1. Информатика мазмұнының бағыттары мен логикалық ойлау дағдыларының пәндік көріністері арасындағы байланыс

Информатика мазмұнының бағыттары	*Логикалық ойлау дағдыларының пәндік көріністері*
Алгоритмдеу және программалау	*Қойылған есепті талдау; есептің бөліктері мен олардың арасындағы байланыстарды анықтау; шешу тәсілдерін салыстыру; алгоритмнің орындалуын тексеру; алынған нәтиже негізінде қорытынды жасау*
Ақпараттық модельдеу	*Объектінің мәнді қасиеттерін талдау және бөліп алу; абстракциялау; объектілер мен олардың қасиеттері арасындағы байланыстарды анықтау; модельдерді салыстыру; модельдің сәйкестігін тексеру*
Деректермен жұмыс	*Деректерді талдау және салыстыру; оларды белгілері бойынша топтау және жүйелеу; деректер арасындағы байланыстар мен заңдылықтарды анықтау; нәтижелерді интерпретациялау және қорытынды жасау*
Компьютерлік жүйелер	*Жүйе компоненттерінің қызметін талдау; компоненттер арасындағы функционалдық байланыстарды анықтау; олардың өзара әрекеттесуін интерпретациялау; жүйенің жұмыс нәтижесі бойынша қорытынды жасау.*
Желілік технологиялар	*Желінің құрылымы мен компоненттерін талдау; олардың арасындағы байланыстарды анықтау; деректер алмасу үдерісін интерпретациялау; желілік шешімдерді салыстыру және олардың жұмысын тексеру*
Ақпараттық қауіпсіздік	*Қауіптер мен олардың себептерін талдау; себеп-салдарлық байланыстарды анықтау; қауіпсіздік шешімдерін салыстыру; ықтимал салдарды бағалау; негізделген қорытынды жасау*
Компьютерлік графика	*Графикалық объектілердің қасиеттерін талдау және салыстыру; олардың құрылымдық байланыстарын анықтау; объектіні модельдеу кезінде мәнді қасиеттерді бөліп алу және абстракциялау; нәтижені бастапқы талаптармен салыстырып тексеру*

Болашақ информатика мұғалімдерінің логикалық ойлау дағдыларының оқу әрекетіндегі көріністерін жүйелеу және олардың информатика мазмұнындағы пәндік көріністерін анықтау жасанды интеллект құралдарын қолдану жағдайында аталған әрекеттердің орындалу сипатын қарастыруға негіз болады. Жасанды интеллект құралдары есепті талдау, шешу тәсілдерін ұсыну, алгоритм құру, программалық кодты генерациялау, қатені анықтау және нәтижені түсіндіру сияқты әрекеттерді орындауда білім алушыға қолдау көрсете алады. Сонымен қатар бұл мүмкіндіктер білім алушы орындауы тиіс жекелеген логикалық әрекеттердің толық немесе ішінара жасанды интеллект құралдарына берілуіне де жағдай жасайды. Сондықтан логикалық ойлау дағдыларын қалыптастыру тұрғысынан жасанды интеллект құралдарының мүмкіндіктерін білім алушының дербес логикалық әрекетінің сақталуымен өзара байланыста қарастыру маңызды деп есептейміз.

Ғылыми әдебиеттерді талдау жасанды интеллект құралдарын оқу үдерісінде қолданудың педагогикалық мүмкіндіктерімен қатар, дайын жауаптарға сүйену, ұсынылған ақпаратты жеткілікті тексермеу және оқу әрекеттерінің бір бөлігін жасанды интеллектке беру тәуекелдерінің бар екенін көрсетті [1, 2, 4]. Бұл тұжырымдар жасанды интеллект құралдарын қолдану білім алушының логикалық әрекетін қолдаумен қатар, оқу әрекеті тиісті түрде ұйымдастырылмаған жағдайда оның дербес орындалуын шектеуі мүмкін екенін көрсетеді. Болашақ информатика мұғалімдерін даярлау контексінде бұл мәселе ерекше мәнге ие, өйткені олар жасанды интеллект құралдарын өз оқу әрекетінде қолданумен қатар, болашақ кәсіби қызметінде білім алушылардың оқу әрекетін педагогикалық тұрғыдан ұйымдастыруы тиіс.

Осыған байланысты 1-суретте жүйеленген логикалық ойлау дағдыларының оқу әрекетіндегі көріністері мен 1-кестеде анықталған олардың информатика мазмұнындағы пәндік көріністері жасанды интеллект құралдарының аталған әрекеттерді қолдау немесе толық не ішінара орындау мүмкіндіктері және білім алушының дербес логикалық әрекетінің сақталуы тұрғысынан өзара байланыста талданды.

Талдау нәтижесінде жасанды интеллект жағдайында болашақ информатика мұғалімдерінің логикалық ойлау дағдыларын қалыптастыруда үш негізгі педагогикалық мәселе айқындалды: 1) жасанды интеллект құралдарын қолдану барысында білім алушының дербес логикалық әрекетін сақтауды қамтамасыз ету; 2) жасанды интеллект ұсынған нәтижелерді білім алушының дербес тексеруін және негізделген қорытынды жасауын қамтамасыз ету; 3) болашақ информатика мұғалімдерін жасанды интеллект құралдарын білім алушылардың дербес логикалық әрекетін сақтай отырып, педагогикалық мақсатқа сәйкес қолдануға даярлау. Аталған мәселелер жасанды интеллект құралдарының өзінен емес, білім алушының өзі орындауы тиіс логикалық әрекеттер мен жасанды интеллект қолдайтын немесе орындайтын әрекеттердің арақатынасын педагогикалық тұрғыдан ұйымдастыру ерекшеліктерімен байланысты. Олардың пайда болу себептері мен шешу жолдары 2-кестеде берілді.

Кесте 2. Жасанды интеллект жағдайында болашақ информатика мұғалімдерінің логикалық ойлау дағдыларын қалыптастырудағы педагогикалық мәселелер, олардың себептері мен шешу жолдары

Педагогикалық мәселе	*Пайда болу себептері*	*Шешу жолдары*
Білім алушының дербес логикалық әрекетін сақтауды қамтамасыз ету	*ЖИ құралдарының есепті талдау, байланыстарды анықтау, шешу тәсілдерін ұсыну, алгоритм мен программалық код құру, нәтижені түсіндіру сияқты білім алушының логикалық әрекеттерімен байланысты міндеттерді толық немесе ішінара орындау мүмкіндігі; оқу тапсырмаларының білім алушының өзіндік талдауын, салыстыруын және шешім құруын міндетті түрде талап етпеуі*	*Оқу тапсырмаларын білім алушының дербес логикалық әрекеті сақталатындай ұйымдастыру; ЖИ-ді дайын шешім алу үшін емес, бастапқы шешімді толықтыру, балама нұсқаларды салыстыру және жетілдіру үшін қолдану; «өз шешімі → ЖИ ұсынған шешім → салыстыру → жетілдіру» тәрізді тапсырмалар құру*
ЖИ ұсынған нәтижелерді дербес тексеру және негізделген қорытынды жасауды қамтамасыз ету	*ЖИ ұсынған жауаптың, алгоритмнің, кодтың, модельдің немесе түсіндірменің дайын нәтиже ретінде қабылдануы; ұсынылған нәтижені интерпретациялау, салыстыру, байланыстарды анықтау және*	*ЖИ нәтижелерін тексеруге, қатесін анықтауға, балама шешімдермен салыстыруға және бағалауға бағытталған тапсырмалар енгізу; білім алушыдан ЖИ нәтижесін қабылдау немесе қабылдамау себебін*

	тексеруді талап ететін оқу тапсырмаларының жеткіліксіздігі	*негіздеуді және өзіндік қорытынды жасауды талап ету*
Болашақ информатика мұғалімдерін ЖИ құралдарын білім алушылардың дербес логикалық әрекетін сақтай отырып педагогикалық мақсатқа сәйкес қолдануға даярлау	*ЖИ құралдарын қолданудың көбіне техникалық мүмкіндіктеріне бағдарлану; ЖИ орындайтын әрекет пен білім алушының өзі орындауы қажет логикалық әрекетті педагогикалық тұрғыдан ажыратудың жеткіліксіздігі; оқу мақсаты, логикалық әрекет және ЖИ мүмкіндігі арасындағы байланысты ескермеу*	*Болашақ мұғалімдерді оқу мақсатына сәйкес ЖИ қолдану сценарийлерін жобалауға үйрету; тапсырмада білім алушы өзі орындайтын және ЖИ қолдайтын әрекеттерді алдын ала анықтау; ЖИ-дің логикалық әрекетті алмастыру немесе белсендіру ықтималдығын педагогикалық тұрғыдан бағалауға бағытталған тапсырмалар мен әдістемелік жағдаяттарды енгізу*

Дискуссия

Зерттеу нәтижелері жасанды интеллект жағдайында болашақ информатика мұғалімдерінің логикалық ойлау дағдыларын қалыптастыруда білім алушының дербес логикалық әрекетін сақтау маңызды педагогикалық мәселе екенін көрсетті. Бұл нәтиже жасанды интеллект құралдарын білім беруде қолдануда білім алушының дербестігін сақтау, ұсынылған ақпаратты сыни бағалау және тексеру қажеттігін көрсететін зерттеулермен үндеседі [1, 2, 4]. Біздің зерттеуде аталған мәселе информатика мазмұнын меңгеру барысында жүзеге асырылатын логикалық әрекеттермен байланыста нақтыланды. Осы тұрғыдан жасанды интеллект құралдарын қолданудың педагогикалық мәнін олардың функционалдық мүмкіндіктерімен ғана емес, білім алушының дербес логикалық әрекетінің сақталуы тұрғысынан қарастыру маңызды.

Сонымен қатар алынған нәтижелер жасанды интеллект құралдарын логикалық ойлау дағдыларының қалыптасуына біржақты кері әсер ететін фактор ретінде қарастыруға болмайтынын көрсетеді. Bokan және т.б. [3] зерттеуінде жасанды интеллект құралдарын қолданған болашақ информатика мұғалімдерінің логикалық ойлау және мәселе шешу көрсеткіштерінің жоғарылағаны анықталған. Біздің зерттеу бұл тұжырымды толықтыра отырып, жасанды интеллекттің педагогикалық рөлін оның білім алушының логикалық әрекетін қолдау немесе алмастыру сипатымен байланысты қарастыру қажеттігін негіздейді. Демек, негізгі мәселе жасанды интеллекттің дайын шешім ұсынуында емес, оны қолдану барысында білім алушының қандай логикалық әрекеттерді дербес жүзеге асыратынында.

Үшінші педагогикалық мәселе болашақ информатика мұғалімдерінің жасанды интеллект құралдарын педагогикалық мақсатқа сәйкес қолдануға даярлығымен байланысты. Бұл нәтиже жасанды интеллекттің білім беруге енуі жағдайында педагогтің рөлі мен кәсіби міндеттерінің өзгеруін, сондай-ақ мұғалімнің жасанды интеллект мүмкіндіктері мен шектеулерін түсіну қажеттігін көрсететін зерттеулермен сәйкес келеді [10, 11]. Біздің зерттеуде бұл мәселе логикалық ойлау дағдыларын қалыптастыру тұрғысынан нақтыланып, болашақ мұғалімнің білім алушы өзі орындайтын және жасанды интеллект қолдайтын логикалық әрекеттердің арақатынасын педагогикалық тұрғыдан анықтай білу қажеттілігімен байланыстырылды.

Жалпы алғанда, алынған нәтижелер жасанды интеллект құралдары мен логикалық ойлау дағдыларының қалыптасуы арасындағы байланысты «оң әсер – теріс әсер» тұрғысынан біржақты бағалаудың жеткіліксіз екенін көрсетеді. Жасанды интеллекттің педагогикалық рөлі оның қандай оқу мақсатына және білім алушының қандай логикалық әрекеттерімен үйлестіріліп қолданылатынына байланысты. Сондықтан негізгі педагогикалық бағыт технологияны шектеу емес, білім алушының дербес логикалық әрекеті мен жасанды интеллект қолдауының педагогикалық тұрғыдан негізделген арақатынасын қамтамасыз ету болып табылады.

Қорытынды

Жасанды интеллект құралдарының білім беру үдерісіне кеңінен енгізілуі болашақ информатика мұғалімдерінің логикалық ойлау дағдыларын қалыптастыру мәселесін жаңа педагогикалық тұрғыдан қарастыруды қажет етеді. Жүргізілген теориялық-аналитикалық зерттеу логикалық ойлау дағдыларының оқу әрекетіндегі көріністерін ойлау операциялары мен логикалық пайымдау әрекеттерінің өзара байланысы негізінде жүйелеуге мүмкіндік берді. Информатика мазмұнын талдау нәтижесінде аталған дағдылардың алгоритмдеу және программалау, ақпараттық модельдеу, деректермен жұмыс, компьютерлік жүйелер, желілік технологиялар, ақпараттық қауіпсіздік және компьютерлік графика бағыттарындағы пәндік көріністері айқындалды.

Зерттеу нәтижесінде жасанды интеллект жағдайында логикалық ойлау дағдыларын қалыптастыруға байланысты үш негізгі педагогикалық мәселе анықталды: білім алушының дербес логикалық әрекетін сақтау; жасанды интеллект ұсынған нәтижелерді дербес тексеру және олардың негізінде негізделген қорытынды жасау; болашақ информатика мұғалімдерін жасанды интеллект құралдарын педагогикалық мақсатқа сәйкес қолдануға даярлау. Аталған мәселелерді талдау жасанды интеллекттің логикалық ойлау дағдыларын қалыптастырудағы рөлін оның технологиялық мүмкіндіктерімен ғана емес, білім алушының логикалық әрекетін қолдау немесе оны толық не ішінара алмастыру сипаты арқылы бағалау қажеттігін көрсетті.

Осыған байланысты жасанды интеллект құралдарын информатиканы оқытуға кіріктіруде білім алушының өзі орындауы қажет логикалық әрекеттер мен жасанды интеллект қолдайтын әрекеттердің арақатынасын педагогикалық тұрғыдан анықтау маңызды. Оқу тапсырмаларын білім алушының талдау, салыстыру, байланыстарды анықтау, тексеру және қорытынды жасау әрекеттерін сақтайтындай ұйымдастыру, сондай-ақ жасанды интеллект ұсынған нәтижелерді талдау, тексеру және негіздеуді талап ету оның дербес логикалық әрекетін қолдаудың маңызды бағыты болып табылады. Бұл болашақ информатика мұғалімдерін даярлауда жасанды интеллект құралдарын техникалық меңгертумен қатар, оларды оқу мақсаты мен қалыптастырылатын логикалық әрекеттерге сәйкес педагогикалық тұрғыдан қолдануға даярлау қажеттігін негіздейді.

Келесі зерттеулерде анықталған педагогикалық мәселелер мен ұсынылған оны шешу жолдарын болашақ информатика мұғалімдерін даярлау үдерісінде эксперименттік апробациялау, сондай-ақ жасанды интеллект құралдарын қолдану жағдайында білім алушылардың дербес логикалық әрекетінің сақталу ерекшеліктерін зерттеу көзделеді.

Пайдаланылған дереккөздердің тізімі

[1] Kasneci E., Sessler K., Küchemann S., Bannert M., Dementieva D., Fischer F. et al. ChatGPT for good? On opportunities and challenges of large language models for education // Learning and Individual Differences. – 2023. – Vol. 103. – 102274. – DOI: https://doi.org/10.1016/j.lindif.2023.102274

[2] Tlili A., Shehata B., Adarkwah M.A., Bozkurt A. et al. What if the devil is my guardian angel: ChatGPT as a case study of using chatbots in education // Smart Learning Environments. – 2023. – Vol. 10. – Article 15. – DOI: https://doi.org/10.1186/s40561-023-00237-x

[3] Bokan M., Abdimalikkyzy J., Babaeva A., Keldibekova A., Makpyr S., Smagulov Y. Enhancing Logical Thinking Skills of Future Informatics Teachers through Artificial Intelligence // Qubahan Academic Journal. – 2025. – Vol. 5, No. 1. – P. 543–551. – DOI: https://doi.org/10.48161/qaj.v5n1a1329

[4] UNESCO. AI and the Future of Education: Disruptions, Dilemmas and Directions. – Paris: UNESCO, 2025. AI and the future of education. Disruptions, dilemmas and directions

[5] Chan C.K.Y., Hu W. Students' voices on generative AI: perceptions, benefits, and challenges in higher education // International Journal of Educational Technology in Higher Education. – 2023. – Vol. 20. – Article 43. – DOI: https://doi.org/10.1186/s41239-023-00411-8

[6] Рубинштейн С.Л. Основы общей психологии. – 2-е изд. – М.: Учпедгиз, 1946. – 704 с.

[7] Johnson-Laird P.N. Deductive reasoning // Annual Review of Psychology. – 1999. – Vol. 50. – P. 109–135. – DOI: https://doi.org/10.1146/annurev.psych.50.1.109

[8] Bronkhorst H., Roorda G., Suhre C., Goedhart M. Logical Reasoning in Formal and Everyday Reasoning Tasks // International Journal of Science and Mathematics Education. – 2020. – Vol. 18. – P. 1673–1694. – DOI: https://doi.org/10.1007/s10763-019-10039-8

[9] Saygıner Ş., Tüzün H. The effects of block-based visual and text-based programming training on students' achievement, logical thinking skills, and motivation // Journal of Computer Assisted Learning. – 2023. – Vol. 39, No. 2. – P. 644–658. – DOI: https://doi.org/10.1111/jcal.12771

[10] Holmes W., Bialik M., Fadel C. Artificial Intelligence in Education: Promises and Implications for Teaching and Learning. – Boston: Center for Curriculum Redesign, 2019.

[11] Luckin R., George K., Cukurova M. AI for School Teachers. – Boca Raton: CRC Press, 2022. – 132 p. DOI: https://doi.org/10.1201/9781003193173

[12] Қазақстан Республикасы Оқу-ағарту министрінің 2022 жылғы 16 қыркүйектегі № 399 бұйрығы «Жалпы білім беру ұйымдарына арналған жалпы білім беретін пәндердің, бастауыш, негізгі орта және жалпы орта білім деңгейлерінің таңдау курстарының үлгілік оқу бағдарламаларын бекіту туралы» (өзгерістер мен толықтырулармен) [Электрондық ресурс]. – «Әділет» ақпараттық-құқықтық жүйесі. – URL: https://adilet.zan.kz/kaz/docs/V2200029767 (қаралған күні: 04.09.2026).

References

[1] Kasneci E., Sessler K., Küchemann S., Bannert M., Dementieva D., Fischer F. et al. (2023) ChatGPT for good? On opportunities and challenges of large language models for education. Learning and Individual Differences, 103, 102274. https://doi.org/10.1016/j.lindif.2023.102274

[2] Tlili A., Shehata B., Adarkwah M.A., Bozkurt A. et al. (2023) What if the devil is my guardian angel: ChatGPT as a case study of using chatbots in education. Smart Learning Environments, 10, Article 15. https://doi.org/10.1186/s40561-023-00237-x

[3] Bokan M., Abdimalikkyzy J., Babaeva A., Keldibekova A., Makpyr S., Smagulov Y. (2025) Enhancing Logical Thinking Skills of Future Informatics Teachers through Artificial Intelligence. Qubahan Academic Journal, 5(1), 543–551. https://doi.org/10.48161/qaj.v5n1a1329

[4] UNESCO. (2025) AI and the Future of Education: Disruptions, Dilemmas and Directions. Paris: UNESCO.

[5] Chan C.K.Y., Hu W. (2023) Students' voices on generative AI: perceptions, benefits, and challenges in higher education. International Journal of Educational Technology in Higher Education, 20, Article 43. https://doi.org/10.1186/s41239-023-00411-8

[6] Rubinshtein S.L. (1946) Osnovy obshhei psihologii [Fundamentals of General Psychology]. 2nd ed. Moscow: Uchpedgiz, 704 p. (In Russian)

[7] Johnson-Laird P.N. (1999) Deductive reasoning. Annual Review of Psychology, 50, 109–135. https://doi.org/10.1146/annurev.psych.50.1.109

[8] Bronkhorst H., Roorda G., Suhre C., Goedhart M. (2020) Logical Reasoning in Formal and Everyday Reasoning Tasks. International Journal of Science and Mathematics Education, 18, 1673–1694. https://doi.org/10.1007/s10763-019-10039-8

[9] Saygıner Ş., Tüzün H. (2023) The effects of block-based visual and text-based programming training on students' achievement, logical thinking skills, and motivation. Journal of Computer Assisted Learning, 39(2), 644–658. https://doi.org/10.1111/jcal.12771

[10] Holmes W., Bialik M., Fadel C. (2019) Artificial Intelligence in Education: Promises and Implications for Teaching and Learning. Boston: Center for Curriculum Redesign.

[11] Luckin R., George K., Cukurova M. (2022) AI for School Teachers. Boca Raton: CRC Press, 132 p. https://doi.org/10.1201/9781003193173

[12] Kazakstan Respublikasy Oku-agartu ministrinin 2022 zhylgy 16 kyrkuyektegi № 399 buirygy «Zhalpy bilim beru uiymdaryna arnalgan zhalpy bilim beretin panderdin, bastauysh, negizgi orta zhane zhalpy orta bilim dengeilerinin tandau kurstarynyn ulgilik oku bagdarlamalaryn bekitu turaly» (ozgerister men tolyktyrularmen) [On Approval of Model Curricula for General Education Subjects and Elective Courses for Primary, Basic Secondary and General Secondary Education Levels] [Electronic resource]. «Adilet» akparattyk-kukyktyk zhuyesi. https://adilet.zan.kz/kaz/docs/V2200029767 (accessed September 4, 2026). (In Kazakh)

ҒТАХР 29.01.45

https://doi.org/10.51889/2959-5894.2026.95.3.028

Ж.К. Сыдыкова[1], Б. Ерженбек[1], Г.И. Жанбекова[1], Ж.М. Нұрмухамедова[1], Д.М. Нурбаева[1]

[1]Абай атындағы Қазақ ұлттық педагогикалық университеті, Алматы қ., Қазақстан
[*]*e-mail:bulbul.83@mail.ru*

ОҚУШЫЛАРДЫҢ ФУНКЦИОНАЛДЫҚ САУАТТЫЛЫҒЫН ҚАЛЫПТАСТЫРУДА ЖАСАНДЫ ИНТТЕЛЕКТІНІ ҚОЛДАНУ ӘДІСТЕМЕСІ

Аңдатпа

Бұл мақалада физика пәнін оқыту барысында оқушылардың функционалдық сауаттылығын қалыптастыруда жасанды интеллектіні қолдану мүмкіндіктері қарастырылады. Зерттеудің мақсаты – оқушылардың функционалдық сауаттылығын қалыптастыруда жасанды интеллектіні қолдану жолдарын көрсету. Яғни,физика сабақтарында цифрлық білім беру ресурстарын, виртуалды эксперименттерді, 3D модельдерді және жағдаяттық тапсырмаларды қолдану арқылы оқушылардың функционалдық сауаттылығын қалыптастыру тиімділігін анықтау. Мақалада физиканы оқытуда foxar.education, efizika.ru және Corinth платформаларындағы виртуалды эксперименттерді қолдану қарастырылды. Оқушыларға Атмосфералық қысым. Атмосфералық қысымды өлшеу, Кішкентай денелердің өлшемін анықтау, Тұрақты электр тогы тақырыптары бойынша практикалық жағдаяттық тапсырмалар ұсынылды. Физиканы оқытуда жасанды интеллектінің мүмкіндіктерін тиімділігін бағалау мақсатында мұғалімдер арасында сауалнама жүргізілді. Мұғалімдердің басым бөлігі жасанды интеллекткөмегімен жүргізілгенвиртуалды эксперименттердің оқушылардың пәнге қызығушылығын арттыруға және физикалық құбылыстарды терең түсінуіне, функционалдық сауаттылығын қалыптастыруға оң әсер ететінін атап өтті.Зерттеу қорытындылары мектепте физиканы оқытуда жасанды интеллектіні жүйелі қолдану оқушылардың функционалдық сауаттылығын, зерттеушілік және практикалық дағдыларын қалыптастыратындығын көрсетті.

Түйін сөздер: функционалдық сауаттылық, физика, жасанды интеллект, оқыту әдістемесі, виртуалды эксперименттер.

Ж.К. Сыдыкова[1], Б. Ерженбек[1], Г.И. Жанбекова[1], Ж.М.Нұрмухамедова[1], Д.М.Нурбаева[1]
[1]Казахский Национальный Педагогический Университет имени Абая, г. Алматы, Казахстан

МЕТОДИКА ИСПОЛЬЗОВАНИЯ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА ДЛЯ ФОРМИРОВАНИЯ ФУНКЦИОНАЛЬНОЙ ГРАМОТНОСТИ ОБУЧАЮЩИХСЯ

Аннотация

В данной статье рассматриваются возможности применения искусственного интеллекта в процессе формирования функциональной грамотности учащихся при обучении физике. Цель исследования – показать пути использования искусственного интеллекта для формирования функциональной грамотности учащихся, а именно определить эффективность формирования функциональной грамотности через применение цифровых образовательных ресурсов, виртуальных экспериментов, 3D-моделей и ситуационных задач на уроках физики. В статье рассмотрено использование виртуальных экспериментов на платформах foxar.education, efizika.ru и Corinth в процессе обучения физике. Учащимся были предложены практические ситуационные задания по темам «Атмосферное давление. Измерение атмосферного давления», «Определение размеров малых тел», «Постоянный электрический ток». С целью оценки эффективности возможностей искусственного интеллекта в обучении физике был проведён опрос среди учителей. Большинство учителей отметили, что виртуальные эксперименты, проводимые с использованием искусственного интеллекта, положительно влияют на повышение интереса учащихся к предмету, способствуют более глубокому пониманию физических явлений и формированию функциональной грамотности. Результаты исследования показали, что систематическое применение искусственного интеллекта в школьном обучении физике способствует формированию у учащихся функциональной грамотности, исследовательских и практических навыков.

Ключевые слова: функциональная грамотность, физика, искусственный интеллект, методика обучения, виртуальные эксперименты.

Zh.K. Sydykova[1], B. Yerzhenbek[1], G.I. Zhanbekova[1], Zh.M.Nurmukhamedova[1], D.M.Nurbayeva[1]
[1]AbaiKazakh National Pedagogical University, Republic of Kazakhstan, Almaty

METHODOLOGY FOR THE USE OF ARTIFICIAL INTELLIGENCE IN DEVELOPING STUDENTS' FUNCTIONAL LITERACY

Abstract

This article examines the possibilities of using artificial intelligence in the development of students' functional literacy in the process of teaching physics. The aim of the study is to demonstrate ways of applying artificial intelligence in the formation of students' functional literacy, specifically to determine the effectiveness of developing functional literacy through the use of digital educational resources, virtual experiments, 3D models, and situational tasks in physics lessons. The article considers the use of virtual experiments on the platforms foxar.education, efizika.ru, and Corinth in physics instruction. Students were given practical situational tasks on the topics «Atmospheric Pressure. Measurement of Atmospheric Pressure», «Determination of the Size of Small Objects», and «Direct Current». To evaluate the effectiveness of artificial intelligence in physics education, a survey was conducted among teachers. The majority of teachers noted that virtual experiments supported by artificial intelligence have a positive impact on increasing students' interest in the subject, contribute to a deeper understanding of physical phenomena, and support the development of functional literacy. The results of the study show that the systematic use of artificial intelligence in school physics education contributes to the development of students' functional literacy, as well as their research and practical skills.

Keywords: functional literacy, physics, artificial intelligence, teaching methodology, virtual experiments.

Кіріспе

Қазақстан Республикасындағы білім беру жүйесі соңғы жылдары бірқатар жаңарту үдерістерін бастан өткеріп, мазмұндық және құрылымдық тұрғыдан өзгерістерге ұшырауда. Қазіргі қоғамдағы әлеуметтік-экономикалық даму үрдістері мектеп түлектерінен тек пәндік білімді ғана емес, сонымен қатар мобильділік, коммуникативтік қабілет және когнитивтік дағдылар сияқты кең ауқымды құзыреттіліктерді талап етеді. Осыған байланысты білім беру үдерісіне қойылатын талаптар да едәуір жаңарып, оқыту мазмұны мен нәтижелеріне жаңа көзқарас қалыптасты. Бұл өзгерістер мұғалім мен оқушының рөлін қайта қарауды, сондай-ақ білім, білік және дағдыларды қалыптастыру жүйесін жетілдіруді қажет етеді. Оқыту үдерісінің тиімділігі тек игерілген білімнің көлемімен немесе сапасымен ғана емес, сонымен бірге сол білімді әртүрлі жағдайларда қолдана алу деңгейімен де өлшенеді [1].

Қазіргі заманғы білім беру жүйесі тұлғаның жан-жақты дамуына бағытталған, оқыту мен тәрбиені біртұтас процесс ретінде қарастыратын жаңа педагогикалық модельді қажет етеді. Мұндай жүйеде мұғалім білімді дайын күйінде беруші емес, оқушылардың танымдық іс-әрекетін ұйымдастырушы және бағыттаушы қызметін атқарады [2].

PISA халықаралық функционалдық сауаттылықты бағалау зерттеуі бойынша функционалдық сауаттылықтың құрамдас бөліктері ретінде математикалық сауаттылық, оқу сауаттылығы, жаратылыстану-ғылыми сауаттылығы, ақпараттық сауаттылық, жаһандық құзыреттіліктер мен креативті ойлауды жатқызады [3]. Бұл оқушылардың аталған салалардағы алған білімін нақты өмірлік жағдайларда қолдана алу қабілетіне ие болуы қажет екендігін көрсетеді. Біз мақалада оқушылардың жаратылыстану-ғылыми сауаттылығына қысқаша тоқталамыз.

Жаратылыстану-ғылыми сауаттылықты қалыптастыру негізінен эксперименттік және зерттеушілік сипаттағы тапсырмалар арқылы жүзеге асырылады. Мұндай тапсырмалар оқушылардың жаратылыстану ғылымдары бойынша алған білімдерін күнделікті өмірде кездесетін физикалық құбылыстар мен үдерістерді түсіндіруге қолдану дағдыларын дамытуға мүмкіндік береді. Сонымен қатар, олар қоршаған ортаны ғылыми тұрғыдан танып-білуге, бақылау жүргізуге және алынған нәтижелерді талдауға үйретеді [4].

Оқушылардың жаратылыстану-ғылыми сауаттылығын қалыптастыруда олар негізгі үш құзыреттілікке ие болады:

1) Жаратылыстану-ғылыми таным әдісінің негізгі ерекшеліктерін түсіну. Бұл әртүрлі ғылыми таным әдістерін қолдана білуді қамтиды.

2) Жаратылыстану бағытындағы құбылыстарды ғылыми білім негізінде сипаттай алуы және олардың өзгерістерін болжай білу.

3) Ғылыми дәлелдер мен қолда бар деректерді пайдаланып қорытынды жасау, алынған нәтижені талдау және сенімділігін бағалай алу.

Осы тұрғыдан алғанда, мұғалімнің негізгі мақсаты – оқушылардың алған білімдері, дағдылары мен тәжірибелік іс-әрекеттерін өмірлік жағдаяттарда тиімді пайдалана алу құзыреттілігін дамыту. Осы орайда мұғалім оқу үдерісінің тиімділігін қамтамасыз ету мақсатында әртүрлі оқыту әдістері мен жасан интелектіні пайдаланып, сабақ мазмұнын түрлендіріп отырғаны маңызды.

Қазіргі уақытта білім беру жүйесін цифрландыру үдерісі қарқынды дамып, жасанды интеллектке негізделген технологияларды оқу-тәрбие жұмысына енгізу маңызды бағыт ретінде қарастырылуда. Зерттеушілердің еңбектерінде цифрлық құралдардың тек ақпаратты өңдеу мен ұсыну қызметін атқарып қана қоймай, білім беру үдерісіне қатысушылар арасындағы өзара байланыс пен ынтымақтастықты күшейтетін тиімді механизм екендігі атап көрсетіледі. Мұндай технологияларды пайдалану оқыту тәсілдерін жетілдіруге, оқушылардың оқу әрекетіне белсенді қатысуына мүмкіндік береді[5].

Жасанды интеллект дегеніміз адамның интеллектуалдық қабілеттерін имитациялайтын және ақпаратты өңдеу, білімді меңгеру, талдау, пайымдау, тілдік өзара әрекеттесу мен шешім қабылдау функцияларын орындай алатын интеллектуалды компьютерлік жүйелердің жиынтығын түсінеміз.

Білім беруде қолданылатын жасанды интеллекттің түрлері: 1) Adaptive learning systems- білім беруге бейімделген интеллектуалды жүйелер; 2) AI tutors - виртуалды ассистенттер мен чат-боттар; 3) AI-based simulations- симуляциялық және модельдеу жүйелері; 4) Learning analytics - деректерді талдау мен бағалау жүйелері; 5) Generative AI- білім беру контентін генерациялайтын ЖИ жүйелері.

Физика пәнін оқыту барысында білім алушылардың функционалдық сауаттылығын қалыптастыруда оқу эксперименттерінің маңызы ерекше. Демонстрациялық тәжірибелер, фронтальды зертханалық жұмыстар, эксперименттік тапсырмалар, қолдан жасалатын құралдар және үй жағдайында орындалатын эксперименттер оқушылардың жаратылыстану-ғылыми сауаттылығын қалыптастыруға ықпал етеді. Сонымен қатар, мұндай тәжірибелік жұмыстар оқушылардың ғылыми-зерттеу дағдыларын қалыптастырудың тиімді құралы болып табылады. Физикалық эксперименттерді орындау кезінде жасанды интелект қолдану бақылау жүргізу, құбылыстарды бейнелі түрде қабылдау және алынған нәтижелерді талдау мен жалпылау қабілеттерін дамытуға мүмкіндік береді [6].

Физиканы оқыту барысында теориялық білімді тәжірибемен ұштастыру оқушылардың оқу материалын терең меңгеруіне ықпал етеді. Осы тұрғыда жасанды интеллект технологияларын қолдану арқылы физикалық құбылыстарды модельдеу, бақылау және қайта жаңғырту оқушылардың білімін бекітудің тиімді тәсілі болып табылады. Оқушылар құбылыстарды өз бетінше зерттеп, олардың жүру заңдылықтарын тәжірибе жүзінде көрген кезде, алынған ақпарат ұзақ уақыт есте сақталады. Сонымен қатар, физикалық процестердің көрнекі түрде ұсынылуы мен практикалық қолданылуы пәнге деген қызығушылықты арттырады. Оқушылар физикалық заңдардың техникада, өндірісте және күнделікті өмірде қолданылу аясын түсіну арқылы функционалдық сауаттылығын дамытып қана қоймай, физика пәнінің маңыздылығын тереңірек сезінеді [7].

Зерттеу әдіснамасы

Оқушылардың физика пәніне деген қызығушылығын арттырумен қатар, олардың функционалдық сауаттылығын қалыптастыру мақсатында эксперименттік тапсырмаларды жасанды интеллект технологияларын пайдалана отырып ұйымдастыруға болады.

Функционалдық сауаттылықты қалыптастыруға бағытталған төмендегідей эксперименттік тапсырмалар ұсынылады.

7-сыныпта «Атмосфералық қысым. Атмосфералық қысымды өлшеу» тақырыбын оқытуда жасанды интелект арқылы атмосфералық қысымды анықтау экспериментін https://foxar.education/ платформасы арқылы көрсетуге болады (1-сурет). Бұл эксперимент арқылы оқушыларға қысым күшінің ауданға қалай тәуелді болатыны көрсетіледі.



Сурет1. Атмосфералық қысымды анықтау эксперименті

Тақырыпты бекіту мақсатында оқушылардың жаратылыстану-ғылыми сауаттылығын қалыптастыруға арналғанкелесі жағдаяттық тапсырмаларды ұсынуға болады:

1- тапсырма. Мыжылған бөтелке. Оқушы ыстық сумен шайылған пластик бөтелкенің қақпағын мықтап жауып қояды. Біраз уақыттан кейін бөтелкенің қабырғалары ішке қарай майысып қалғаныбайқалады.

Сұрақтар:

1. Бөтелкенің пішіні неліктен өзгерді?

2. келесі келтірілген тұжырымдардың қайсысы дұрыс: а) бөтелке ішіндегі ауа қысымы артты; ә) бөтелке ішіндегі ауа қысымы азайды; б) атмосфералық қысым азайды; в) бөтелке массасы өзгерді.

3. Егер бөтелкенің қақпағы жабылмаса, дәл осындай құбылыс байқалар ма еді? Жауабыңызды түсіндіріңіз.

2-тапсырма. Жасанды өкпе. Оқушылар пластик бөтелке мен шардан өкпенің жұмысын модельдеді. Бөтелке түбіндегі резеңке қабық төмен тартылғанда ішкі шар үрленді, ал жоғары басылғанда шар кішірейді (2-сурет).



Сурет 2. «Жасанды өкпе» тәжірибесі

Сұрақтар:

1. Бұл модельде бөтелке, шар және резеңке қабық адам ағзасының қандай бөліктерін бейнелейді?

2. Шардың үрлену себебін қысым ұғымын пайдаланып түсіндіріңіз.

3. Адам терең дем алған кезде көкет (диафрагма) қандай бағытта қозғалады?

4. Бұл модельдің артықшылықтары мен шектеулерін атаңыз.

3-тапсырма. Ауа шары: мерекелік шарада әртүрлі пішіндегі ауа шарлары пайдаланылды. Кейбіреулері шар тәрізді болса, кейбіреулері ұзынша болды.

Сұрақтар:

1. Үрленген шардың ішіндегі газ қысымы қалай таралады?

2. Неліктен ауа бір бағыттан үрленсе де, шар барлық бағытта ұлғайды?

3. Газдардың қандай қасиеттері осы тәжірибеде байқалады?

4-тапсырма. Сабын көпіршігі. Оқушы сабынды ерітіндіні түтікше арқылы үрлеп көпіршік жасап, оның әрқашан шар тәрізді болатынын байқады.

Сұрақтар:

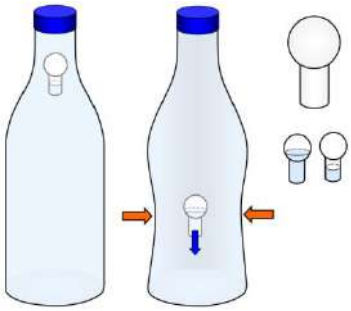
1. Сабын көпіршігінің шар пішінді болуының негізгі себебі неде?

2. Неліктен көпіршік куб немесе цилиндр тәрізді болмайды?

3. Сабын қабығындағы күштердің әсерін түсіндіріңіз.

4. Осы құбылыстың табиғаттағы немесе техникадағы қолданылуына мысал келтіріңіз.

5-тапсырма. Картезиандық сүңгуір. Оқушылар суға толтырылған бөтелкенің ішіне қалқыма орналастырды. Бөтелкені қысқанда қалқыма төмен түсіп, қысымды азайтқанда қайта жоғары көтерілді (3-сурет).



Сурет 3. «Картезиандық сүңгуір» тәжірибесі

Сұрақтар:

1. Қалқыманың төмен түсу себебін түсіндіріңіз.

2. Бөтелкені қысқанда қалқыманың ішіндегі ауа көлемі қалай өзгереді?

3. Неліктен қалқыманың тығыздығы артады?

4. Бұл тәжірибе қандай физикалық заңдарды түсіндіруге мүмкіндік береді?

6-тапсырма. Шам және ауа шары. Оқушы жанған шамды төңкерілген стақанның ішінде ұстап, содан кейін стақанды үрленген ауа шарының үстіне орналастырады. Біраз уақыттан кейін шардың пішіні өзгергені байқалады.

Сұрақтар:

1. От сөнгеннен кейін стақан ішіндегі қысым неліктен төмендейді?

2. Атмосфералық қысым бұл тәжірибеде қандай рөл атқарады?

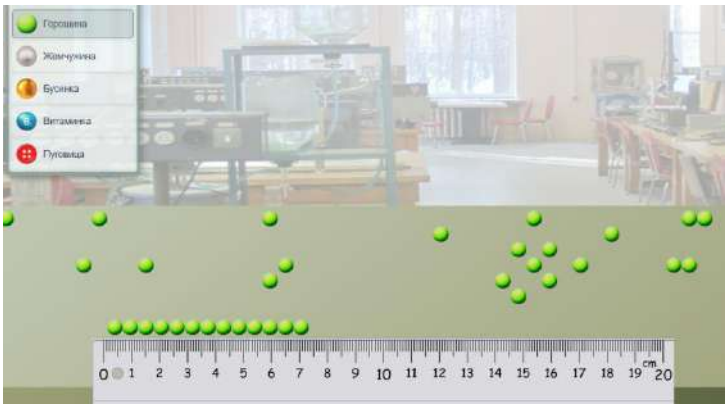
3. Егер тәжірибе үлкенірек стақанмен орындалса, нәтиже қалай өзгеруі мүмкін? Болжам жасаңыз және түсіндіріңіз.

Физика пәнін оқыту барысында оқушылардың функционалдық сауаттылығын қалыптастыруға бағытталған жағдаяттықтапсырмалардыжүйелі түрде қолданса, бұл оқушылардың жаратылыстану-ғылыми сауаттылығын арттырып, олардың жан-жақты дамуына оң әсер етеді. Мұндай жұмыстар оқушылардың шығармашылық қабілеттерін жетілдіруге, стандартты емес ойлау дағдыларын қалыптастыруға және интеллектуалдық

белсенділігін дамытуға, бастысы оқушылардың функционалдық сауаттылығын қалыптастыруда мүмкіндік береді [8].

Оқушыларға 7-сыныптың «Кішкентай денелердің өлшемін анықтау» тақырыбындағы №1 зертханалық жұмысын орындатуда жасанды интелектінің мүмкіндігіне сүйенуге болады. Ол үшін https://efizika.ru/html5/266/index.html платформасын қолдануға болады. Жұмыста өлшеу құралы ретінде сызғыш алынады. Алайда сызғыштың көмегімен өте кішкентай дененің, мысалы, бір тары дәнінің өлшемін тікелей және дәл анықтау мүмкін емес. Егер сызғышты бір ғана дәнге жақындатып өлшейтін болсақ, оның диаметрі 1 мм-ден үлкен, бірақ 2 мм-ден кіші екенін ғана шамамен айтуға болады. Мұндай өлшеу әдісі жоғары дәлдікті қамтамасыз етпейді.

Өлшеу дәлдігін арттыру үшін бірнеше бірдей дәндерді бір қатарға, араларында бос орын қалдырмай орналастыру қажет. Осылайша дәндер қатарының жалпы ұзындығы (L) (мм) өлшенеді. Дәндердің өлшемдері шамамен бірдей деп есептелетіндіктен, бір дәннің диаметрін анықтау үшін қатардың жалпы ұзындығын (L) оны құрайтын дәндер санына (N) бөлу жеткілікті (4-сурет).



Сурет 4. «Кішкентай денелердің өлшемін анықтау» эксперименті

Енді осы «Кішкентай денелердің өлшемін анықтау» тақырыбында оқушылардың функционалдық сауаттылығын қалыптастыруға бағытталған келесі практикалық жағдаяттық тапсырмаларды ұсынуға болады https://seninvg07.narod.ru/004_fiz_lab.htm .

1-тапсырма. Ұсақ денелердің көлемін анықтау. Оқушыларға көлемдері бірдей 200 дана ұсақ түйіршік берілді. Бір түйіршіктің көлемі өте кішкентай болғандықтан, оны тікелей өлшеу мүмкін болмады. Сондықтан оқушылар мензуркаға 50 мл су құйып, барлық 200 түйіршікті суға салды. Нәтижесінде су деңгейі 60 мл-ге дейін көтерілді.

Оқушылар осы практикалық жағдаяттық тапсырманы орындап келесі сұрақтарға жауап береді.

Сұрақтар:

1. Түйіршіктерді суға салғаннан кейін судың деңгейі неге көтерілді?
2. Барлық түйіршіктердің жалпы көлемін қалай анықтауға болады?
3. Бір түйіршіктің көлемін есептеңіз.
4. Неліктен тәжірибені орындауда бір түйіршіктің орнына 200 түйіршік қолданылды?
5. Егер түйіршіктер саны 400 болса және су деңгейі 70 мл-ге дейін көтерілсе, бір түйіршіктің көлемі өзгерер ме еді? Жауабын түсіндіріңіз.
6. Әртүрлі ұсақ денелердің көлемін анықтау үшін қандай физикалық әдіс қолданылады?
7. Бұл әдісті күнделікті өмірде немесе өндірісте қандай заттардың көлемін анықтау үшін пайдалануға болады? Мысал келтіріңіз.

2-тапсырма. CD-дискінің көлемін анықтау. Оқушыларға CD-дискінің көлемін анықтау тапсырылды. Зерттеу жүргізу үшін оларға миллиметрлік қағаз және 10 дана CD-диск берілді. Дискінің көлемін анықтау үшін алдымен оның тесігін есептемегендегі ауданын және

қалыңдығын табу қажет болды. Оқушылар дискіні миллиметрлік қағазға орналастырып, сыртқы шекарасын белгілеп, диаметрін өлшеді. Нәтижесінде дискінің диаметрі 12 см екені анықталды. Сондай-ақ ортасындағы тесіктің диаметрі 1,4 см болды. Кейін олар 10 дискіні қатар қойып, жалпы қалыңдығының 1,3 см екенін анықтады https://seninvg07.narod.ru/004_fiz_lab.htm.

Оқушыларға осы практикалық жағдаяттық тапсырманы орындаған соң келесі сұрақтарға жауап берулері ұсынылады.

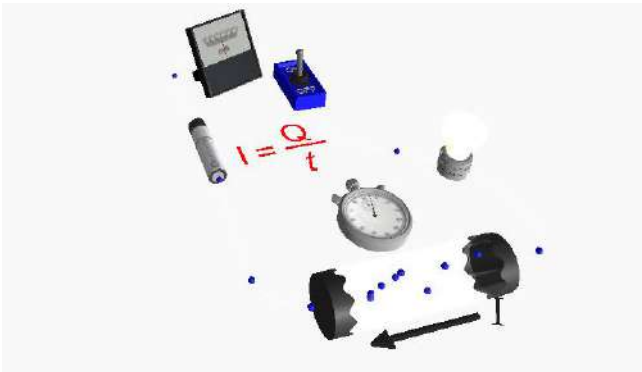
Сұрақтар:

1. Неліктен дискінің ауданын есептеу кезінде үлкен шеңбердің ауданынан ортасындағы тесіктің ауданын азайту қажет?
2. Дискінің сыртқы радиусын және тесіктің радиусын анықтаңыз.
3. Дискінің тесігін есептемегендегі ауданын есептеңіз.
4. Бір CD-дискінің қалыңдығын және көлемін анықтаңыз.
5. Егер дискінің ортасында тесік болмаса, оның көлемі қалай өзгерер еді?
6. Неліктен бір дискінің қалыңдығын емес, 10 дискінің жалпы қалыңдығын өлшеген тиімді?
7. Осындай әдіспен көлемін анықтауға болатын күнделікті тұрмыста кездесетін тағы қандай заттарды атауға болады? Мысалдар келтіріңіз.

Оқушылар күрделі пішінді денелердің көлемін анықтау үшін геометриялық білімдерін қолданады, тәжірибелік өлшеулер жүргізеді, алынған мәліметтерді талдайды және қорытынды жасайды [9].

Оқушыларға «Тұрақты электр тогы» тарауын меңгерту барысында электр тізбегінің құрылысын және тұрақты токтың пайда болу шарттарын көрнекі түрде түсіндіру маңызды. Осы мақсатта үшөлшемді (3D) интерактивті модельдерді пайдалану оқыту үдерісінің тиімділігін арттырады. Мұндай модельдер қысқа уақыт ішінде іске қосылады, пайдалану кезінде қауіпсіздікті қамтамасыз етеді және оқушыларға электр тізбегінің элементтерімен өздігінен жұмыс жасауға мүмкіндік береді. Атап айтқанда, оқушылар қажетті құрылғыларды таңдап, оларды тізбекке қосу немесе басқа элементтермен алмастыру арқылы әртүрлі тәжірибелік жағдайларды модельдей алады.

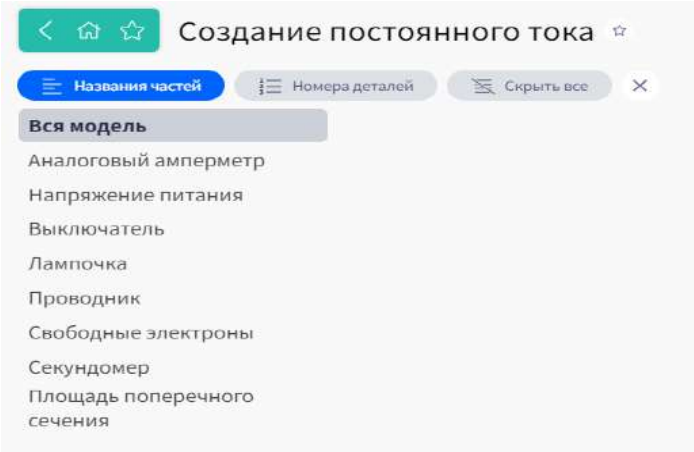
Электр тогының пайда болуын және электр тізбегінің жұмыс істеу принципін түсіндіру үшін интерактивті оқыту ресурстарының бірі ретінде Corinth 3D платформасындағы модельді қолдануға болады (5-сурет) [10]. Бұл модель электр тізбегінің негізгі құрамдас бөліктерін зерттеуге, олардың өзара байланысын бақылауға және тұрақты токтың пайда болу механизмін көрнекі түрде түсінуге мүмкіндік береді.



Сурет 5. «Тұрақты электр тогы» тарауын оқытуда 3D модельдерді қолдану

3D модельдерді қолдану арқылы оқушылар электр тізбегінің элементтерін өзара байланыстырып, тұрақты электр тогын алу үдерісін тәжірибелік тұрғыдан бақылай алады. Бұл оқушылардың жаратылыстану-ғылыми сауаттылығын қалыптастырады. Модель электр тізбегін кезең-кезеңімен құрастыруға мүмкіндік беретіндіктен, оқушылар оның жұмыс істеу принципін көрнекі түрде меңгереді.

Бағдарламаның ерекшеліктерінің бірі – экранның арнайы бөлігінде тізбек құрамына кіретін құрылғылардың атауларының көрсетілуі. Кез келген атауды таңдаған кезде сәйкес элемент модельде ерекшеленіп белгіленеді және оның тізбектегі орналасу орны көрсетіледі (6-сурет). Мұндай интерактивті мүмкіндік оқушылардың электр тізбегінің құрамдас бөліктерін тануына, олардың атауларын дұрыс меңгеруіне және ұзақ уақыт есте сақтауына ықпал етеді [11].



Сурет 6. Тұрақты электр тоғын алу 3D моделі

1-тапсырма. Электр тізбегінің кедергісін анықтау. Оқушы виртуалды 3D моделі көмегімен электр тізбегін құрастырды. Тәжірибені орындау барысында амперметр 0,5 А, ал вольтметр 6 В көрсетті. Оқушы $R = \frac{U}{I}$ Ом заңын қолданып тізбек бөлігінің кедергісін анықтауы керек [12].

Нәтижесінде оқушылар осы практикалық жағдаяттық тапсырманы орындап келесі сұрақтарға жауап берулері керек.

Сұрақтар:

1. Амперметр мен вольтметрдің көрсеткіштерін қолданып, тізбек бөлігінің кедергісін есептеңіз.

2. Ом заңы бойынша ток күші, кедергі және кернеу арасында қандай байланыс бар? Егер тізбектегі ток күшін екі есе арттырып, кедергі өзгеріссіз қалдырса, онда кернеу қалай өзгереді?

3. Электр тізбегінің кедергісін анықтау үшін неліктен амперметр мен вольтметр қолданылады?

4. Электр құрылғыларын пайдалануда кедергінің жоғары немесе төмен болуы қандай жағдайларда маңызды рөл атқарады? Мысал келтіріңіз.

Оқушылар электр тізбегінің кедергісін Ом заңын қолдана отырып анықтайды, өлшеу құралдарының қызметін түсіндіреді және электр тізбектерінің жұмысын талдайды[13].

Зерттеу нәтижелері

Зерттеу жұмысын орындау барысында мектепте физиканы оқытудажасанды инттелектінің мүмкіндіктеріне сүйену оқушылармен мұғалімдер үшін қаншалықты тиімді болғанын анықтау мақсатындамұғалімдер арасында сауалнама жүргізілді. Мұғалімдер арасында сауалнама жүргізу кезінде төмендегі сұрақтар қойылды:

1. Виртуалды эксперименттердің орындалуы сіздерге қаншалықты қиындық туғызды?

2. Эксперименттік тапсырмаларды орындауда жасанды инттелект қолдану оқушылардың назарын аудара алды ма?

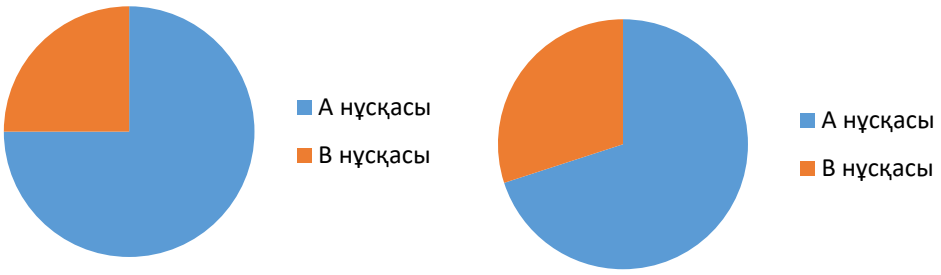
3. Жасанды инттелектқолдану оқушыларды физикалық құбылысты түсінуіне септігін тигізе алды ма?

4. Виртуалды эксперименттерді қолдану оқушыларды физикалық құбылысты түсінуіне септігін тигізе алды ма?

Мұғалімдер арасында жүргізілген сауалнама нұсқалары келесі түрге жіктелді (7-8 суреттер):

1-сауалнама бойынша: А нұсқасы – қиындықтар туындаған жоқ; В нұсқасы – қиындықтар болды, бірақ ол оқу сапасына кері әсерін тигізген жоқ; С нұсқасы – әдістемені қолдану тиімділігі төмен немесе күмәнді деп бағаланды (7-сурет, а).

2-сауалнама бойынша: А нұсқасы –оқушылардың назарын толық аударды; В нұсқасы – оқушылардың назарын аударды, кейбір қиындықтар болды; С нұсқасы –әдістемені қолдану тиімділігі төмен (7-сурет, ә).

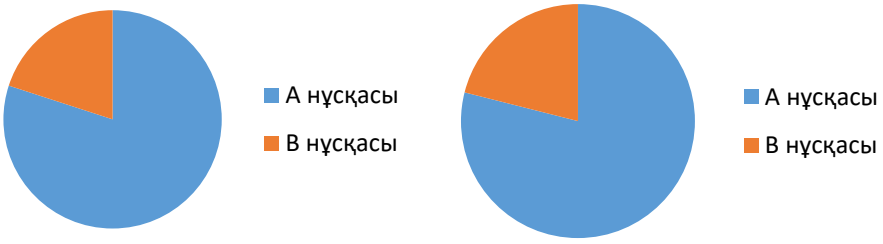


а) *ә)*

Сурет 7. Сауалнама нәтижелері

3-сауалнама бойынша: А нұсқасы – Физикалық құбылыстарды түсінуде қиындықтар туындаған жоқ; В нұсқасы – Кейбір қиындықтар болды, бірақ олар материалды меңгеруге кедергі келтірген жоқ; С нұсқасы – АКТ-ны қолданудың физикалық құбылыстарды түсінуге әсері төмен (8-сурет, а).

4-сауалнама бойынша: А нұсқасы – Физикалық құбылыстарды түсінуде қиындықтар туындаған жоқ; В нұсқасы – Кейбір қиындықтар болды, бірақ олар оқу сапасына кері әсер еткен жоқ; С нұсқасы – Жасанды интеллектінің мүмкіндіктеріне сүйену тиімділігі төмен немесе күмәнді деп бағаланды (8-сурет, ә).



а) *ә)*

Сурет 8. Сауалнама нәтижелері

Сауалнаманың нәтижелерін 7-8 суреттердегі диаграммалардан қарай отырып, С нұсқасы жауабының мүлде болмағанын, яғни жасанды интеллектінің мүмкіндіктеріне сүйену мұғалім үшін тиімді болғанын көруге болады.

Дискуссия

Зерттеу нәтижелері физика сабақтарында жасанды интеллектіні қолданып, виртуалды эксперименттерді және 3D модельдерді қолдану оқушылардың функционалдық сауаттылығын қалыптастыруға оң ықпал ететінін көрсетті. Мұғалімдер арасында жүргізілген сауалнамада әдістеменің тиімділігін төмен немесе күмәнді деп бағалаған жауаптардың болмауы ұсынылған тәсілдердің білім беру үдерісінде қолдануға қолайлы екенін аңғартады. Сауалнамада

мұғалімдердің басым бөлігі виртуалды эксперименттерді орындау барысында елеулі қиындықтардың туындамағанын және аталған құралдардың оқушылардың назарын сабақ мазмұнына аударуға мүмкіндік бергенін атап өтті.

Алынған нәтижелерді оқушылардың жаратылыстану-ғылыми сауаттылығына қойылатын талаптар тұрғысынан қарастырғанда, виртуалды тәжірибелер мен жағдаяттық тапсырмалар бірнеше маңызды құзыреттілікті қатар дамытуға жағдай жасайтыны байқалады. Оқушылар физикалық құбылысты бақылап қана қоймай, оның пайда болу себептерін түсіндіреді, қажетті формулаларды қолданады, өлшеу нәтижелерін талдайды және ғылыми дәлелдер негізінде қорытынды жасайды.

Зерттеу нәтижелерін бағалау кезінде бірқатар шектеулерді де назарға алған жөн. Зерттеу қорытындылары негізінен мұғалімдер арасында жүргізілген сауалнамаға сүйенеді және олардың кәсіби пікірлерін сипаттайды. Сондықтан ұсынылған әдістеменің оқушылардың оқу жетістіктеріне тікелей әсерін нақтылау үшін алдағы зерттеулерде эксперименттік және бақылау топтарының нәтижелерін салыстыру, оқушылардың бастапқы және қорытынды білім деңгейін анықтау, сондай-ақ функционалдық сауаттылық көрсеткіштерінің өзгерісін сандық тұрғыдан бағалау қажет.

Осылайша, зерттеу барысында ұсынылған виртуалды эксперименттер, 3D модельдер және практикалық жағдаяттық тапсырмалар теориялық білімді өмірлік жағдайлармен байланыстыруға мүмкіндік береді. Оларды дәстүрлі оқыту әдістерімен үйлестіріп, жүйелі әрі мақсатты қолдану оқушылардың физикалық құбылыстарды ғылыми тұрғыдан түсіндіру, ақпаратты талдау, есептеу жүргізу, дәлелді қорытынды жасау және алған білімін практикада қолдану қабілеттерін дамытуға ықпал етеді.

Қорытынды

Физикадан оқушылардың функционалдық сауаттылығын қалыптастыруда ЖИ-ді қолдану – білім беру үдерісін жетілдірудің маңызды құралдарының бірі болып табылады. Физиканы оқытуда жасанды интеллектінің мүмкіндігіне сүйену оқушыларға алған білімдерін нақты өмірде қолдана алу қабілетін қалыптастыруға бағытталған, сондай-ақ олардың аналитикалық ойлау, мәселелерді шешу дағдыларын дамытуды, олардың функционалдық сауаттылығын қалыптастыруды көздейді. ЖИ-і соның ішінде виртуалды эксперименттерді, интерактивті платформалардың және мультимедиа ресурстарының көмегімен білім беру оқушылардың пәнге деген қызығушылығын арттырып қана қоймай, сонымен қатар олардың функционалдық сауаттылығын да қалыптастыруға мүмкіндік береді. Сондай-ақ, виртуалды эксперименттер оқушылардың білімдерін тереңдетуге септігін тигізеді.

Жалпы, физикадан оқушылардың функционалдық сауаттылығын қалыптастыруда ЖИқолдану білім беру сапасын арттыруға және оқушылардың кәсіби өмірінде табысты болуына ықпал етеді. Қорытындылай келе, физиканы оқытуда ЖИ-нің мүмкіндіктерін кеңінен қолдану оқушылардың білім сапасын арттыруға және олардың функционалдық сауаттылығын қалыптастыруға ықпал етеді деген тұжырым жасауға болады.

Пайдаланылған әдебиеттер тізімі

[1] Постановление Правительства Республики Казахстан от 27 декабря 2019 года № 988 «Об утверждении Государственной программы развития образования и науки Республики Казахстан на 2020–2025 годы» [Электронный ресурс]. – URL: https://adilet.zan.kz/rus/docs/P1900000988 (дата обращения: 10.05.2026).

[2] Джадрина М. Д. Содержание школьного образования в Казахстане: новые перспективы // Вестник КазНУ. Серия «Педагогические науки». – 2014. – № 3 (43). – С. 14–23. – URL: https://bulletin-pedagogic-sc.kaznu.kz/index.php/1-ped/ru/article/view/72 (дата обращения: 08.04.2026).

[3] Сейсова А. Б., Хамзина Ш. Ш., Муканова Р. Ж. Применение проблемного подхода при формировании функциональной грамотности в специализированных школах // Вестник КазНПУ имени

Абая. Серия «Педагогические науки». – 2024. – № 3 (83). – С. 381–390. – DOI: https://doi.org/10.51889/2959-5762.2024.83.3.034.

[4] Бондарева С. А., Петросова Е. В., Веремейцева Т. И. Формирование функциональной грамотности на уроках физики // Вестник науки. – 2024. – № 6 (75). – С. 725–730. – URL: https://cyberleninka.ru/article/n/formirovanie-funktsionalnoy-gramotnosti-na-urokah-fiziki (дата обращения: 08.04.2026).

[5] Crompton H., Burke D. Artificial intelligence in higher education: the state of the field // International Journal of Educational Technology in Higher Education. – 2023. – Vol. 20. – Article 22. – DOI: https://doi.org/10.1186/s41239-023-00392-8.

[6] Абдуллаева Г. Ф. Формирование функциональной грамотности учащихся на уроках физики // Вестник Казанского государственного педагогического университета. – 2018. – С. 70–74. – URL: https://www.elibrary.ru/item.asp?edn=yrhppb (дата обращения: 11.03.2026).

[7] Deta U. A., Ayun S. K., Laila L., Prahani B. K., Suprapto N. PISA science framework 2018 vs 2025 and its impact in physics education: Literature review // Momentum: Physics Education Journal. – 2024. – Vol. 8, no. 1. – P. 95–107. – DOI: https://doi.org/10.21067/mpej.v8i1.9215.

[8] Li Y., Guo M. Scientific literacy in communicating science and socio-scientific issues: Prospects and challenges // Frontiers in Psychology. – 2021. – Vol. 12. – Article 758000. – DOI: https://doi.org/10.3389/fpsyg.2021.758000.

[9] Agussuryani Q., Sudarmin S., Sumarni W., Cahyono E., Ellianawati E. STEM literacy in growing vocational school student HOTS in science learning: A meta-analysis // International Journal of Evaluation and Research in Education. – 2022. – Vol. 11, no. 1. – P. 51–60. – DOI: https://doi.org/10.11591/ijere.v11i1.21647.

[10] Khoiri A. Meta-analysis study: Effect of STEM (Science, Technology, Engineering and Mathematics) towards achievement // Formatif: Jurnal Ilmiah Pendidikan MIPA. – 2019. – Vol. 9, no. 1. – P. 71–82. – DOI: https://doi.org/10.30998/formatif.v9i1.2937.

[11] Tunggyshbay M., Balta N., Admiraal W. Flipped classroom strategies and innovative teaching approaches in physics education: A systematic review // Eurasia Journal of Mathematics, Science and Technology Education. – 2023. – Vol. 19. – DOI: https://doi.org/10.29333/ejmste/13258.

[12] Thornhill-Miller B., Camarda A., Mercier M., Burkhardt J.-M., Morisseau T., Bourgeois-Bougrine S., Vinchon F., El Hayek S., Augereau-Landais M., Mourey F., Feybesse C., Sundquist D., Lubart T. Creativity, critical thinking, communication, and collaboration: Assessment, certification, and promotion of 21st century skills for the future of work and education // Journal of Intelligence. – 2023. – Vol. 11, no. 3. – Article 54. – DOI: https://doi.org/10.3390/jintelligence11030054.

[13] Courtney M., Karakus M., Ersozlu Z., Nurumov K. The influence of ICT use and related attitudes on students' math and science performance: Multilevel analyses of the last decade's PISA surveys // Large-scale Assessments in Education. – 2022. – Vol. 10. – Article 8. – DOI: https://doi.org/10.1186/s40536-022-00128-6.

References

[1] Postanovlenie Pravitelstva Respubliki Kazakhstan ot 27 dekabrya 2019 goda № 988 «Ob utverzhdenii Gosudarstvennoi programmy razvitiya obrazovaniya i nauki Respubliki Kazakhstan na 2020–2025 gody» [Resolution of the Government of the Republic of Kazakhstan dated December 27, 2019, No. 988 "On Approval of the State Program for the Development of Education and Science of the Republic of Kazakhstan for 2020–2025"] [Electronic resource]. Available at: https://adilet.zan.kz/rus/docs/P1900000988 (accessed: 10.05.2026). (In Russian)

[2] Dzhadrina M. D. Soderzhanie shkolnogo obrazovaniya v Kazakhstane: novye perspektivy [The content of school education in Kazakhstan: New perspectives]. Vestnik KazNU. Seriya «Pedagogicheskie nauki» [Journal of Educational Sciences], 2014, no. 3 (43), pp. 14–23. Available at: https://bulletin-pedagogic-sc.kaznu.kz/index.php/1-ped/ru/article/view/72 (accessed: 08.04.2026). (In Russian)

[3] Seisova A. B., Khamzina Sh. Sh., Mukanova R. Zh. Primenenie problemnogo podkhoda pri formirovanii funktsionalnoi gramotnosti v spetsializirovannykh shkolakh [Application of a problem-based approach in developing functional literacy in specialized schools]. Abai atyndagy KazUPU-nin Khabarshysy. «Pedagogika gylymdary» seriyasy [Bulletin of Abai KazNPU. Series of Pedagogical Sciences], 2024, no. 3 (83), pp. 381–390. DOI: https://doi.org/10.51889/2959-5762.2024.83.3.034. (In Russian)

[4] Bondareva S. A., Petrosova E. V., Veremeitseva T. I. Formirovanie funktsionalnoi gramotnosti na urokakh fiziki [Development of functional literacy in physics lessons]. Mezhdunarodnyi nauchnyi zhurnal «Vestnik nauki» [International Scientific Journal "Bulletin of Science"], 2024, no. 6 (75), pp. 725–730. Available at: https://cyberleninka.ru/article/n/formirovanie-funktsionalnoy-gramotnosti-na-urokah-fiziki (accessed: 08.04.2026). (In Russian)

[5] Crompton H., Burke D. Artificial intelligence in higher education: The state of the field. International Journal of Educational Technology in Higher Education, 2023, vol. 20, article 22. DOI: https://doi.org/10.1186/s41239-023-00392-8.

[6] Abdullaeva G. F. Formirovanie funktsionalnoi gramotnosti uchashchikhsya na urokakh fiziki [Development of students' functional literacy in physics lessons]. Vestnik Kazanskogo gosudarstvennogo pedagogicheskogo universiteta [Bulletin of Kazan State Pedagogical University], 2018, pp. 70–74. Available at: https://www.elibrary.ru/item.asp?edn=yrhppb (accessed: 11.03.2026). (In Russian)

[7] Deta U. A., Ayun S. K., Laila L., Prahani B. K., Suprapto N. PISA science framework 2018 vs 2025 and its impact in physics education: Literature review. Momentum: Physics Education Journal, 2024, vol. 8, no. 1, pp. 95–107. DOI: https://doi.org/10.21067/mpej.v8i1.9215.

[8] Li Y., Guo M. Scientific literacy in communicating science and socio-scientific issues: Prospects and challenges. Frontiers in Psychology, 2021, vol. 12, article 758000. DOI: https://doi.org/10.3389/fpsyg.2021.758000.

[9] Agussuryani Q., Sudarmin S., Sumarni W., Cahyono E., Ellianawati E. STEM literacy in growing vocational school student HOTS in science learning: A meta-analysis. International Journal of Evaluation and Research in Education, 2022, vol. 11, no. 1, pp. 51–60. DOI: https://doi.org/10.11591/ijere.v11i1.21647.

[10] Khoiri A. Meta-analysis study: Effect of STEM (Science, Technology, Engineering and Mathematics) towards achievement. Formatif: Jurnal Ilmiah Pendidikan MIPA, 2019, vol. 9, no. 1, pp. 71–82. DOI: https://doi.org/10.30998/formatif.v9i1.2937.

[11] Tunggyshbay M., Balta N., Admiraal W. Flipped classroom strategies and innovative teaching approaches in physics education: A systematic review. Eurasia Journal of Mathematics, Science and Technology Education, 2023, vol. 19. DOI: https://doi.org/10.29333/ejmste/13258.

[12] Thornhill-Miller B., Camarda A., Mercier M., Burkhardt J.-M., Morisseau T., Bourgeois-Bougrine S., Vinchon F., El Hayek S., Augereau-Landais M., Mourey F., Feybesse C., Sundquist D., Lubart T. Creativity, critical thinking, communication, and collaboration: Assessment, certification, and promotion of 21st century skills for the future of work and education. Journal of Intelligence, 2023, vol. 11, no. 3, article 54. DOI: https://doi.org/10.3390/jintelligence11030054.

[13] Courtney M., Karakus M., Ersozlu Z., Nurumov K. The influence of ICT use and related attitudes on students' math and science performance: Multilevel analyses of the last decade's PISA surveys. Large-scale Assessments in Education, 2022, vol. 10, article 8. DOI: https://doi.org/10.1186/s40536-022-00128-6.

ҒТАХР 14.35.07

https://doi.org/10.51889/2959-5894.2026.95.3.029

Ш.Т. Шекербекова[1], Ү.Қ. Базарбаева[1*], Б.Б. Қалданов[1]

[1]Абай атындағы Қазақ ұлттық педагогикалық университеті, Алматы қ., Қазақстан
[*]*e-mail: umit.kuan@bk.ru*

БОЛАШАҚ ИНФОРМАТИКА МҰҒАЛІМДЕРІН АҚПАРАТТЫҚ ҮДЕРІСТЕРДІ ОҚЫТУҒА ДАЙЫНДАУДА ЖАСАНДЫ ИНТЕЛЛЕКТ НЕГІЗІНДЕГІ ИНФОГРАФИКАНЫ ҚОЛДАНУ

Аңдатпа

Мақалада болашақ информатика мұғалімдерін ақпараттық үдерістерді оқытуға жасанды интеллект негізіндегі инфографиканы қолдану арқылы дайындау әдістемесі ұсынылады. «Көрнекілік», «визуализация», «инфографика» және жасанды интеллект негізіндегі инфографика ұғымдарының арақатынасы нақтыланып, инфографиканың дидактикалық функциялары айқындалды. Негізгі жасанды интеллект негізіндегі инфографика платформалары (ChatGPT, Gemini, Copilot, Canva Magic Studio, Piktochart AI, Venngage, Napkin AI, код негізіндегі визуализация), инфографиканың мүмкіндіктері, артықшылықтары мен кемшіліктері және «Ақпараттық үдерістер» тақырыбына сәйкестігі бойынша салыстырылды. Зерттеу барысында төрт компонентті дайындау әдістемесі, тапсырмалар жүйесі және жасанды интеллект міндетті мазмұндық тексерумен үйлестіретін инфографика жасаудың алты кезеңді үлгісі әзірленді. Экспериментке 6В01507 Информатика білім беру бағдарламасының 64 студенті (болашақ информатика мұғалімдері) және 18 мектеп мұғалімі қатысты. Нәтижелер платформаларды оқу мақсатына сай гибридті үйлестіру және мұғалімнің сараптамалық бақылауы жағдайында жасанды интеллект негізіндегі инфографиканы қолданудың орындылығын көрсетеді.

Түйін сөздер: оқу ақпаратын визуализациялау, көрнекілік, инфографика, ақпараттық үдерістер, болашақ информатика мұғалімі, жасанды интеллект, жасанды интеллект негізіндегі инфографика, жасанды интеллект негізіндегі платформалар.

Ш.Т. Шекербекова[1], У.К. Базарбаева[1], Б.Б. Калданов[1]
[1]Казахский национальный педагогический университет имени Абая, г. Алматы, Казахстан

ИСПОЛЬЗОВАНИЕ ИНФОГРАФИКИ НА ОСНОВЕ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА ПРИ ПОДГОТОВКЕ БУДУЩИХ УЧИТЕЛЕЙ ИНФОРМАТИКИ К ОБУЧЕНИЮ ИНФОРМАЦИОННЫМ ПРОЦЕССАМ

Аннотация

В статье предлагается методика подготовки будущих учителей информатики к обучению информационным процессам с использованием инфографики на основе искусственного интеллекта (ИИ-инфографики). Уточнено соотношение понятий «наглядность», «визуализация», «инфографика» и «ИИ-инфографика», определены дидактические функции инфографики. Основные платформы ИИ-инфографики (ChatGPT, Gemini, Copilot, Canva Magic Studio, Piktochart AI, Venngage, Napkin AI, визуализация на основе кода) сопоставлены по инфографическим возможностям, преимуществам и недостаткам, а также соответствию теме «Информационные процессы». Разработаны четырёхкомпонентная методика подготовки, система заданий и шестиэтапная модель создания инфографики, сочетающая использование ИИ с обязательной содержательной проверкой. В эксперименте приняли участие 64 студента (будущие учителя информатики) образовательной программы 6В01507 Информатика и 18 школьных учителей. Результаты подтверждают целесообразность применения ИИ-инфографики при гибридном сочетании платформ в соответствии с учебной целью и экспертном контроле учителя.

Ключевые слова: визуализация учебной информации, наглядность, инфографика, информационные процессы, будущий учитель информатики, искусственный интеллект, инфографика на основе искусственного интеллекта, платформы ИИ-инфографики.

Sh.T. Shekerbekova[1], U.K. Bazarbayeva[1], B.B. Kaldanov[1]
[1]Abai Kazakh National Pedagogical University, Almaty, Kazakhstan

USING AI-BASED INFOGRAPHICS TO PREPARE FUTURE INFORMATICS TEACHERS TO TEACH INFORMATION PROCESSES

Abstract

The article proposes a methodology for preparing future computer science teachers to teach information processes using AI-based infographics. The relationship between the concepts of "visual aids", "visualization", "infographics" and "AI-based infographics" is clarified, and the didactic functions of infographics are identified. The main AI infographic platforms (ChatGPT, Gemini, Copilot, Canva Magic Studio, Piktochart AI, Venngage, Napkin AI and code-based visualization) are compared in terms of their infographic capabilities, advantages, limitations and suitability for the topic "Information Processes". A four-component training methodology, a system of tasks and a six-stage model for creating infographics that combines AI use with mandatory content verification were developed. The experiment involved 64 students (future computer science teachers) of the 6B01507 Informatics programme and 18 schoolteachers. The results confirm the feasibility of AI-based infographics when platforms are combined in a hybrid way according to the learning goal and the teacher exercises expert control.

Keywords: visualization of educational information, visual aids, infographics, information processes, pre-service computer science teacher, artificial intelligence, AI-based infographics, AI infographic platforms.

Кіріспе

Қазіргі қоғамның трансформациялануы жағдайында білім беру жүйесін заманауи талаптарға сай бейімдеу мемлекеттік саясаттың басым бағыттарының біріне айналды. Осы тұрғыда, Қазақстан Республикасында 2023–2029 жылдарға арналған мектепке дейінгі, орта, техникалық және кәсіптік білім беруді дамыту тұжырымдамасында білім беру үдерісін цифрландыру, сапалы цифрлық ресурстарды әзірлеу және кәсіби құзыреттілігін арттыру басым бағыт ретінде белгіленген [1]. Аталған міндеттерді іске асыру білім беру ортасын технологиялық тұрғыдан жаңартумен қатар, болашақ информатика мұғалімдерін дайындау мазмұны мен әдістерін жетілдіруді талап етеді.

Қазіргі цифрлық білім беру кеңістігінде білім алушылар ақпараттың үлкен көлемімен жұмыс істейді, бұл қажетті мәліметтерді анықтау, сұрыптау және жүйелеу үдерістерін күрделендіреді. Әсіресе, ақпараттың көлемді мәтін түрінде ұсынылуы оны қабылдау мен түсінуге кедергі келтіруі мүмкін. Осыған орай, білім беру үдерісінде «*қандай ақпаратты беру керек»* деген сұрақпен қатар *«оны қалай ұсыну керек»* деген мәселе өзекті болып отыр [2]. Бұл мәселені шешу ақпаратты визуалды модельдеу, құрылымдау және көрнекі түрде ұсыну құралдарын мақсатқа сай қолдануды қажет етеді.

Я.А. Коменский, И.Г. Песталоцци және К.Д. Ушинскийдің еңбектерінде негізделген көрнекілік принципі цифрлық білім беру жағдайында жаңа мазмұнмен толығып, оқу ақпаратын құрылымдау, мазмұндық байланыстарын айқындау және оны визуалды құралдар арқылы түсінікті ұсыну тәсілі ретінде дамып келеді [1]. Ақпараттық үдерістер мектеп информатикасы курсының негізгі мазмұндық бағыттарының бірі болып саналады және оларды оқытудың әдістемелік ерекшеліктері ғылыми еңбектерде қарастырылған [3]. Атап айтқанда, ақпаратты беру үдерісінің К. Шеннон ұсынған байланыс моделіне негізделуі оның құрылымын сызбалар мен визуалды құралдар арқылы түсіндіруге қолайлы жағдай жасайды. Болашақ информатика мұғалімдерін цифрлық технологияларды қолдануға дайындаудың теориялық негіздерінің бірі – пәндік білімді, педагогикалық тәсілдерді және технологиялық құралдарды өзара ұштастыратын TPACK моделі. Генеративті жасанды интеллекттің дамуымен аталған модель болашақ информатика мұғалімінің ЖИ технологияларының әлеуеті мен шектеулерін сыни тұрғыдан талдау қабілетін қамтитын құзыреттіліктермен толықтырыла түседі [4, 5]. Жоғары білім беру саласындағы ЖИ-ге қатысты зерттеулердің басым бөлігі оның технологиялық мүмкіндіктерін қарастыруға бағытталғанымен, оны оқу үдерісінде қолданудың педагогикалық негіздері жеткілікті деңгейде ашылмаған [6]. Сонымен қатар,

үлкен тілдік модельдер оқу материалдарын әзірлеуде кең мүмкіндіктер ұсынғанымен, мазмұн жасауда тәуекелдер де кездеседі [7]. Осылайша, жасанды интеллектті (ЖИ) болашақ информатика мұғалімдерін дайындау үдерісіне енгізу оның нақты пәндік мазмұнмен сабақтастығы қамтамасыз етілген жағдайда ғана нәтижелі болады. Дегенмен, болашақ информатика мұғалімдерін ақпараттық үдерістерді оқытуға дайындау барысында ЖИ негізіндегі инфографиканы қолдану мәселесі әлі де арнайы зерттеуді қажет етеді.

Информатиканың негізгі ұғымдары – ақпарат, деректер, ақпараттық үдерістер, кодтау, ақпаратты өлшеу – абстрактілі сипатқа ие: ақпаратты жинау, өңдеу, сақтау және беру үдерістері өзара тығыз байланысты, бірақ тек мәтіндік анықтамалар арқылы олардың логикасын көру қиын. Бұған дейінгі зерттеуімізде болашақ информатика мұғалімдерін ақпараттық үдерістерді оқытуға арнайы дайындау қажеттілігі негізделген болатын [8]. Бұл зерттеуде инфографика оқу мазмұнындағы негізгі ұғымдар мен олардың арасындағы байланыстарды мәтіндік және графикалық элементтерді, сондай-ақ визуалды метафораларды үйлестіру арқылы ықшам әрі көрнекі түрде ұсынатын құрал ретінде қарастырылады. Ал ЖИ негізіндегі инфографика дегеніміз – мазмұнды құрылымдау, макет әзірлеу және визуалды элементтерді жасау кезеңдерінің біреуі немесе бірнешеуі генеративті жасанды интеллект құралдарының көмегімен мәтіндік сұраныс (промпт) негізінде жүзеге асырылатын, ал ақпаратты іріктеу, оның дұрыстығын тексеру және оқу мақсаттарына сәйкестігін бағалау болашақ информатика мұғалімнің бақылауында қалатын визуалды оқу материалы. ЖИ негізіндегі инфографика дәстүрлі инфографикадан оны әзірлеу үдерісінің ерекшеліктерімен, адам мен ЖИ арасындағы өзара әрекеттесу сипатымен және нәтижені бақылау тәсілдерімен ерекшеленеді. Атап айтқанда, мәтіндік сұраныстарды пайдалану арқылы қысқа уақыт ішінде инфографиканың бірнеше нұсқасын жасауға болады. Сонымен қатар, алынған материалдың мазмұндық дұрыстығын міндетті түрде тексеру қажет. Бұл жағдайда болашақ информатика мұғалімнің қызметі графикалық дизайн жасаумен шектелмей, оқу мазмұнын жобалау, ЖИ ұсынған нәтижелерді сыни бағалау және оларды дидактикалық мақсаттарға сәйкестендіру міндеттерін қамтиды.

Инфографиканың дидактикалық мүмкіндіктері бірқатар ғылыми зерттеулерде зерделенген: инфографиканың негізгі түрлері мен оны әзірлеу қағидалары айқындалған [1], болашақ информатика мұғалімдерін объектіге бағытталған бағдарламалауға оқытуда қолдану тәжірибесі қарастырылған [9], сондай-ақ мектеп информатикасы курсының жекелеген тақырыптарына арналған инфографикалық материалдар әзірленген [10]. Бүгінде ChatGPT сияқты жалпы мақсаттағы тілдік модельдерді болашақ информатика мұғалімдері мен мектеп мұғалімдерінің көпшілігі қолданады, алайда инфографика жасауға арналған басқа ЖИ негізіндегі инфографика платформаларының (Canva Magic Studio, Napkin AI, Piktochart AI, Venngage және т.б.) мүмкіндіктері ақпараттық үдерістерді оқыту тұрғысынан салыстырмалы түрде талданбаған, ал мұндай талдау қай платформаның қандай оқу тапсырмасына тиімді екенін алдын ала анықтауға мүмкіндік береді.

Зерттеудің мақсаты – болашақ информатика мұғалімдерін ақпараттық үдерістерді оқытуға ЖИ негізіндегі инфографиканы қолдануға дайындау әдістемесін жасау және оның тиімділігін эксперимент арқылы бағалау. Зерттеу мақсатына сәйкес келесі зерттеу сұрақтары айқындалды: инфографика мен ЖИ негізіндегі инфографика ақпараттық үдерістерді оқытуда қандай дидактикалық қызметтер атқарады? (*ЗС1*), қолжетімді ЖИ негізіндегі инфографика платформаларының қайсысы ақпараттық үдерістерді визуализациялауға неғұрлым тиімді? (*ЗС2*), болашақ информатика мұғалімдерін ЖИ негізіндегі инфографика арқылы ақпараттық үдерістерді оқытуға дайындау әдістемесі қандай компоненттерден тұруы тиіс? (*ЗС3*), болашақ және информатика мұғалімдер ЖИ негізіндегі инфографиканы қолданудың тиімділігін қалай бағалайды? (*ЗС4)*.

Зерттеу әдіснамасы

Зерттеу екі кезеңде жүргізілді. *Теориялық кезеңде* ғылыми-әдістемелік әдебиетті талдау, салыстыру және жүйелеу арқылы инфографиканың функциялары мен қолдану тәсілдері жинақталып, «Ақпараттық үдерістер» тақырыбы бойынша тапсырмалар жүйесі әзірленді, сондай-ақ ЖИ негізіндегі инфографика платформаларына алты критерий бойынша салыстырмалы талдау жүргізілді: инфографикалық мүмкіндіктері, мазмұндық дәлдік пен құрылымды бақылау, нәтижені өңдеу мүмкіндігі, қазақ тілін қолдау, қолжетімділігі және «Ақпараттық үдерістер» тақырыбына сәйкестігі. Талдау платформалардың ресми сипаттамалары мен оларды сабақта қолдану тәжірибесіне негізделді. Эмпирикалық кезеңде әзірленген тәсілдердің нәтижесі тәжірибелік эксперимент арқылы бағаланды.

Эксперименттің ұйымдастыру Абай атындағы ҚазҰПУ-дың математика, физика және информатика факультетінде үш кезеңде өтті. *Дайындық кезеңінде* инфографика үлгілері (1-сурет), тапсырмалар (2-кесте) және 10 тұжырымнан тұратын сауалнама әзірленді. *Қолдану кезеңінде* 6В01507 Информатика ББББ-ның 3-4 курс студенттері (n = 64) «Информатиканы оқыту әдістемесі» сабақтарында семестр бойы инфографиканы талдап, өз бетінше (оның ішінде ЖИ арқылы) жасады және сабақ жоспарына енгізді; Алматы қаласы мектептерінің информатика мұғалімдері (n = 18) әдістемелік семинарда осы тәсілдермен танысып, өз сабақтарында қолданып көрді. *Бағалау кезеңінде* екі топқа бірдей сауалнама жүргізілді, яғни ол инфографикамен нақты жұмыс істеген адамдардың пікірін өлшейді.

Сауалнама және деректерді өңдеу. Бірінші бөлімде 10 тұжырым (С1–С10) бес балдық Ликерт шкаласы бойынша бағаланды (1 – «мүлдем келіспеймін», 5 – «толық келісемін»). Тұжырымдар бес топқа біріктірілді: түсіну және есте сақтау (С1–С3), қызығушылық (С4), өз бетінше жасау арқылы дағдыларды дамыту (С5–С6), ЖИ құралдарын пайдалану (С7–С8) және кәсіби дайындық (С9–С10). Екінші бөлімде инфографиканы қолдану тиімді сабақ кезеңдері таңдалды, үшінші бөлімде қиындықтар мен ұсыныстар туралы ашық сұрақтар қойылды. Әр тұжырым бойынша орташа мән (*M*), стандартты ауытқу (*SD*) және 4 немесе 5 балл қойғандардың үлесі (% келісемін) есептелді. Шкаланың сенімділігі Кронбахтың α коэффициентімен ($\alpha \geq 0{,}70$), топтар арасындағы айырмашылық реттік деректерге және көлемі әртүрлі топтарға жарамды Манн–Уитнидің U-критерийімен ($p < 0{,}05$) тексерілді; ашық жауаптар мазмұндық талдау арқылы санаттарға топтастырылды.

Зерттеу нәтижелері

Инфографика ұғымы: көрнекіліктен визуализацияға (ЗС1). «Көрнекілік» пен «визуализация» жиі синоним ретінде қолданылады, алайда көрнекілік демонстрациялық сипатта болып, онда иллюстративтік функция басым, ал визуализацияда когнитивтік функция басым: А.А. Вербицкийдің анықтамасы бойынша бұл ой мазмұнын көрнекі бейнеге «жинақтау» үдерісі, ал бейне кейін қайта «ашылып», ойлау және практикалық әрекеттерге тірек болады [11]. Информатика ұғымдары қатынастар мен үдерістерге негізделгендіктен, оларды оқытуда дайын бейнені көрсету жеткіліксіз, модельдеу қажет.

Инфографика – визуализацияның нақты формасы. В.В. Лаптев оны ақпаратты, байланыстарды, сандық деректер мен білімді графикалық ұсынуға негізделген коммуникативтік дизайн саласы ретінде анықтайды [12]. Инфографиканың графиктер мен диаграммалардан басты айырмашылығы – деректермен қатар өмірлік ұқсастықтар мен визуалды бейнелерді біріктіретін метафоралық сипаты. Осыған сүйеніп, *білім беру инфографикасын* оқу мақсатына сәйкес іріктелген мазмұнды, оның негізгі ұғымдары мен байланыстарын мәтін, графикалық элементтер және визуалды метафоралар бірлігі арқылы ықшам ұсынатын, білім алушының ақпаратты түсіну, талдау және қайта жаңғырту әрекетін қолдайтын визуализация құралы ретінде анықтаймыз.

Инфографиканың оқытудағы қажеттілігі бірнеше негізге сүйенеді. R.E. Mayer мультимедиялық оқыту теориясына сәйкес сөз бен бейне бірге берілгенде, әсіресе мәтін мен бейне қатар орналасып (кеңістіктік жақындық принципі), артық элементтер алынып

тасталғанда (келісімділік принципі), оқыту нәтижесі жоғары болады [13]. Қазіргі білім алушылар визуалды мәдениет ортасында өсуде, ал инфографика мәтінді алмастырмай, ақпаратқа «кіру нүктесін» жасайды. Инфографиканы өз бетінше жасау мазмұнды сыни талдауға және негізгі ақпаратты қосымшасынан ажыратуға үйретеді, ал мета-талдау оның оқу жетістіктеріне оң әсерін көрсетеді [14].

Инфографиканың дидактикалық функциялары 1-кестеде жүйеленді.

Кесте 1. Инфографиканың оқу үдерісіндегі дидактикалық функциялары

Функция	*Мазмұны*	*Информатика сабағындағы мысал*
Когнитивтік	*Абстрактілі ұғымдарды, байланыстар мен қатынастарды көрнекі модель арқылы түсінуге көмектеседі*	*Ақпараттық үдерістердің (жинау, өңдеу, сақтау, беру) өзара байланысын көрсету*
Құрылымдаушы (жүйелеуші)	*Үлкен көлемдегі ақпаратты ықшамдап, негізгі элементтерді бөліп, логикалық құрылымға келтіреді*	*Ақпараттың өлшем бірліктері мен олардың арақатынасын бір кестелі-сызбалық бейнеге біріктіру*
Мотивациялық	*Тақырыпқа назар аударады, қызығушылық пен оқуға деген ынтаны арттырады*	*Сабақ басында «Ақпарат және деректер» инфографикасы арқылы проблемалық сұрақ қою*
Коммуникативтік	*Ақпаратты қысқа әрі түсінікті жеткізуге, оны талқылауға және өз ойын дәлелдеуге мүмкіндік береді*	*Топтық жұмыста инфографика бойынша ақпаратты беру схемасын түсіндіру*
Бақылау-диагностикалық	*Білім алушылардың тақырыпты қаншалықты түсінгенін анықтауға мүмкіндік береді*	*Оқушылардың өзі жасаған инфографикасындағы қателер мен олқылықтарды талдау*
Дамытушылық	*Визуалды, сыни және жүйелі ойлауды, функционалдық сауаттылық пен цифрлық дағдыларды дамытады*	*Инфографикадағы ақпаратты кестеге немесе диаграммаға қайта кодтау, өз инфографикасын жасау*

Инфографиканың түрлері және онымен жұмыс тәсілдері. Ұсыну формасы бойынша статикалық, интерактивті және бейнеинфографика ажыратылады; мектепте ең қолжетімдісі – басып шығаруға және тірек-конспект ретінде пайдалануға ыңғайлы статикалық инфографика. Мазмұндық құрылымы бойынша үдеріс (ақпаратты беру схемасы), жіктеуші (ақпарат түрлері мен қасиеттері), салыстырмалы (аналогтық және дискретті ақпарат), статистикалық, хронологиялық және алгоритмдік инфографика бөлінеді. Білім беру инфографикасына қарапайымдылық, дәлдік, бейнелілік, эстетикалық тартымдылық, ғылыми дұрыстық және жас ерекшелігіне сәйкестік талаптары қойылады.

Инфографикамен жұмыс тәсілдерін оқу әрекетінің күрделену реттілігі ретінде жүйеледік: 1) болашақ информатика мұғалімдерінің дайын инфографиканы түсіндіруі; 2) білім алушылардың инфографиканы «оқуы» және талдауы; 3) ақпаратты қайта кодтау (кестеге, диаграммаға айналдыру); 4) сыни талдау (қателерді анықтап, түзету ұсыну); 5) өз инфографикасын жасау – бұл жұмыстың ең жоғары деңгейі [1].

Болашақ информатика мұғалімдерін ЖИ негізіндегі инфографика *арқылы ақпараттық үдерістерді оқытуға дайындау әдістемесі (ЗС3).* Зерттеудің өзегі – «Информатиканы оқыту әдістемесі» пәні аясында әзірленген әдістеме. Оның теориялық негізі – TPACK моделінің ЖИ жағдайына бейімделген нұсқалары [4, 5]: болашақ информатика мұғалімі пәндік мазмұнды (ақпараттық үдерістер), педагогикалық тәсілдерді (инфографикамен жұмыс) және технологияны (ЖИ негізіндегі инфографика) біртұтас үйлестіре білуі тиіс. *Мақсаттық компонент* – ақпараттық үдерістерді ЖИ негізіндегі инфографика арқылы визуализациялау

және сабақта қолдану дайындығын қалыптастыру. *Мазмұндық компонент* үш модульден тұрады: М1 – ақпараттық үдерістердің ғылыми негіздері (жинау, өңдеу, сақтау, беру, іздеу, қорғау; Шеннонның байланыс моделі) және визуализация принциптері; М2 – ЖИ негізіндегі инфографика платформаларын салыстыру және таңдау, промпт құрастыру, ЖИ нәтижесіндегі қателерді анықтау; М3 – ЖИ негізіндегі инфографиканы сабақ жоспары мен тапсырмалар жүйесіне енгізу, микросабақ. *Процессуалдық компонент* жұмыс тәсілдерінің реттілігіне, тапсырмалар жүйесіне (2-кесте) және алты кезеңді үлгіге (2-сурет) сүйенеді. *Нәтижелік-бағалау компоненті* студенттердің инфографикасын ғылыми дұрыстық, логика, түсініктілік және дизайн критерийлері бойынша бағалауды және сауалнаманы қамтиды.

«Ақпараттық үдерістер» тақырыбы бойынша инфографика мысалы. Негізгі мектептің информатика курсындағы «Ақпарат және деректер», «Ақпараттық үдерістер», «Ақпаратты ұсыну», «Ақпаратты өлшеу» тақырыптарын инфографика арқылы жүйелеуге болады; инфографика оқулықты алмастырмайды, бірақ түсіндіру кезінде көрнекі құрал, ал өзіндік жұмыста тірек-конспект бола алады (1-сурет).



Сурет 1. «Ақпарат және деректер» (а) және «Ақпараттық үдерістер» (б) инфографикалары (авторлық мазмұн негізінде ЖИ құралдарымен визуализацияланған)

«Ақпарат және деректер» инфографикасы (1-сурет, а) ақпарат көздерін, қасиеттерін және ақпарат пен деректердің айырмашылығын өмірлік мысалдармен көрсетеді. «Ақпараттық үдерістер» инфографикасында (1-сурет, б) ақпаратты жинау, өңдеу, сақтау және беру жеке блоктарға бөлініп, әрқайсысы танымал белгімен бейнеленген, бұл кеңістіктік жақындық принципіне сәйкес келеді [13]. Сонымен бірге үдерістердің реттілігі толық айқын емес, «іздеу» мен «қорғау» үдерістері көрсетілмеген; мұндай олқылықтар оқу ресурсына айналады, ал инфографиканың қазақша нұсқасын жасау терминологиялық дәлдікті тексеретін тапсырма бола алады. Осы инфографика негізінде тапсырмалар әзірленді (2-кесте).

Тапсырмалар танудан талдауға, одан шығармашылыққа қарай күрделенеді, бұл Блум таксономиясының деңгейлеріне сәйкес келеді, ал білім алушылар жасаған инфографика тақырыптың қаншалықты меңгерілгенін бағалауға мүмкіндік береді [10].

Кесте 2.«Ақпараттық үдерістер» инфографикасымен жұмыс істеуге арналған тапсырмалар

Сабақ кезеңі	*Тапсырма түрі*	*Тапсырма мысалы*
Білімді өзектендіру, мотивация	*Болжау*	*Инфографиканы 1–2 минут қарап шығыңыз. Суреттегі адамдар мен құрылғылар ақпаратпен қандай әрекеттер жасайды деп ойлайсыз?*

Жаңа материалды түсіндіру	*Ақпаратты іздеу және шығарып алу*	*Инфографикада қандай ақпараттық үдерістер көрсетілген? Әрқайсысына сәйкес белгіні атаңыз.*
Жаңа материалды түсіндіру	*Негізгі және қосымша ақпаратты ажырату*	*Қай элементтер ақпараттық үдерістің мәнін ашады, ал қайсысы тек безендіру үшін берілген?*
Бекіту	*«Дұрыс/бұрыс» Өмірмен байланыстыру*	*Тұжырымдарды бағалаңыз: «Мәтінді кестеге айналдыру – ақпаратты сақтау»; «Дискета – ақпарат тасымалдаушының мысалы». Смартфонмен суретке түсіріп, досыңызға жіберген кезде қандай ақпараттық үдерістер жүзеге асады? Оларды инфографикадағы блоктармен сәйкестендіріңіз.*
Қайталау және жалпылау	*Сыни талдау және толықтыру*	*Инфографикада қандай үдерістер жетіспейді (мысалы, іздеу, қорғау)? Үдерістердің реттілігін көрсеткілер арқылы нақтылаңыз.*
Жоба, үй тапсырмасы	*Шығармашылық (өз инфографикасын жасау)*	*Инфографиканың қазақ тіліндегі нұсқасын жасаңыз немесе «Ақпаратты беру» тақырыбына өз инфографикаңызды құрастырыңыз.*
Бақылау	*Рефлексия, бағалау*	*Өзіңіз немесе сыныптасыңыз жасаған инфографиканы критерийлер бойынша бағалаңыз: ғылыми дұрыстық, логика, түсініктілік, дизайн.*

Инфографиканы жасау кезеңдері. ЖИ құралдары оқу материалын алдын ала құрылымдауға, қысқа тұжырымдар дайындауға және дизайн нұсқаларын жылдам жасауға көмектеседі, алайда олардың нәтижесінде терминологиялық қателер, логикалық реттіліктің бұзылуы, мағынасыз безендіру кездесуі мүмкін. Сондықтан инфографиканы жасау арнайы ұйымдастырылған педагогикалық дизайн әрекеті ретінде қарастырылуы тиіс [15]. Осыны ескеріп, алты кезеңді үлгі ұсынамыз (2-сурет): ЖИ тек төртінші кезеңде, мазмұн мен визуализация тәсілін болашақ информатика мұғалімдерінің (студенттің) өзі анықтағаннан кейін қолданылады, одан кейін міндетті мазмұндық тексеру жүргізіледі.



Сурет 2. Білім беру инфографикасын жасаудың алты кезеңді үлгісі (авторлар әзірлеген)

Мысалы, «Ақпаратты беру» тақырыбы үшін ЖИ-ге ақпарат көзін, хабарламаны, байланыс арнасын, кедергілерді және қабылдаушыны көрсететін, қазақ тіліндегі 7-сыныпқа арналған құрылым сұралады; алынған нұсқада терминология, элементтер реттілігі және бейнелердің мазмұнға сәйкестігі тексеріліп, қате табылса, жұмыс алдыңғы кезеңдерге қайтарылады. (2-сурет).

ЖИ негізіндегі инфографика *платформаларын салыстырмалы талдау (ЗС2).* Тәжірибеде ЖИ қолдану көбіне ChatGPT сияқты тілдік модельдермен шектеледі, алайда инфографика жасауға мүмкіндіктері әртүрлі үш платформалар тобы бар: 1) кескін генерациясы бар тілдік

модельдер; 2) ЖИ модульдері бар дизайн-платформалар мен инфографика генераторлары; 3) мәтінді сызбаға айналдыратын сервистер және код негізіндегі визуализация (3-кесте).

Кесте 3. Негізгі ЖИ негізіндегі инфографика платформаларын салыстыру

Платформа	*Инфографикалық мүмкіндіктері*	*Артықшылықтары*	*Кемшіліктері*	*«Ақпараттық үдерістер» тақырыбына сәйкестігі*
ChatGPT, Gemini (тілдік модель + кескін генерациясы)	*Мазмұнды құрылымдау, инфографика сценарийі, кескін генерациясы, диаграмма коды*	*Диалог арқылы мазмұнды нақтылау; қазақша мәтін құрау; тегін қолжетімділік*	*Кескіндегі қазақша жазуда қателер; «галлюцинация»; растрлық кескінді өңдеу қиын*	*Жоғары – мазмұнды құрылымдауда ; орташа – визуализацияда*
Microsoft Copilot (Designer)	*Мәтіннен кескін жасау; PowerPoint-те дизайн ұсыну*	*Тегін; Microsoft 365 ортасымен интеграция*	*Құрылымды бақылау әлсіз; жазулардағы қателер*	*Төмен–орташа – мотивациялық бейнелер*
Canva Magic Studio (дизайн-платформа)	*Шаблондар, мәтіннен макет құру, кескін генерациясы, қабаттарды өңдеу*	*Шаблон қоры үлкен; әр элемент өңделеді; тегін білім беру нұсқасы*	*Қазақша қаріптер мен шаблондар аз; безендірудің мазмұннан басым болу қаупі*	*Жоғары – безендіру, жіктеуші және салыстырмалы инфографика*
Piktochart AI, Venngage (инфографика генераторлары)	*Сұраныс бойынша инфографика макеті (процесс, уақыт шкаласы, салыстыру), диаграммалар*	*Инфографика жанрына арналған дайын құрылымдар*	*Негізгі функциялары ақылы; шаблондар ағылшын тілді; мазмұны жалпылама*	*Орташа – статистикалық, хронологиялық инфографика*
Napkin AI (мәтіннен сызбаға)	*Мәтін абзацын блок-схемаға, цикл, иерархия диаграммасына айналдыру; SVG/PNG экспорты*	*Мәтіннің логикасын сақтайды; сызбаны жылдам жасайды; өңделеді*	*Стильдік мүмкіндіктері шектеулі; интерфейсі ағылшынша*	*Өте жоғары – процестік инфографика (жинау → өңдеу → сақтау → беру)*
ЖИ + код (Mermaid, draw.io, Python)	*Тілдік модель сызба кодын жазады, ол арнайы ортада визуализацияланады*	*Кириллица мен терминология дәл; толық өңделеді; алгоритмдік ойлауды дамытады*	*Бағдарламалау дағдысын талап етеді; метафоралық бейнелері шектеулі*	*Өте жоғары – процестік және алгоритмдік инфографика*

Талдау нәтижесінде бірде-бір платформаның инфографика жасаудың барлық кезеңін бірдей сапалы қамтамасыз етпейтінін көрсетеді: тілдік модельдер мазмұнды құрылымдауда күшті, бірақ визуализацияда (әсіресе қазақша жазуларда) әлсіз, ал дизайн-платформалар керісінше. Сондықтан платформаларды *гибридті үйлестіру* ұсынылады: тілдік модель арқылы мазмұн құрылымын жобалау (2-суреттегі 4-кезең) → Napkin AI немесе Mermaid көмегімен процестік сызба қаңқасын құру → Canva-да безендіру және қазақша мәтінді түзету → міндетті мазмұндық тексеру. Оқу үдерісінде 3-кесте ЖИ негізіндегі инфографика құралдарын саналы түрде таңдауға бағыт беретін әдістемелік бағдар ретінде қолданылады. Мысалы, болашақ

информатика мұғалімдері ақпаратты беру тақырыбы бойынша бір сызбаны екі түрлі платформада әзірлеп, алынған нәтижелерді белгіленген критерийлер бойынша салыстырады.

*Эксперимент нәтижелері бойынша ш*каланың сенімділігі жоғары болды: Кронбахтың α коэффициенті студент (болашақ информатика мұғалімі) үшін 0,86, мектептегі информатика мұғалімдері үшін 0,84. Сауалнаманың бірінші бөлімінің нәтижелері 4-кестеде берілген.

Кесте 4. Сауалнаманың бірінші бөлімі бойынша нәтижелер (M – орташа балл, SD – стандартты ауытқу, % – 4 немесе 5 балл қойғандар үлесі)

№	Тұжырым	Студенттер (n = 64), M (SD)	Студенттер, % келісемін	Мұғалімдер (n = 18), M (SD)	Мұғалімдер, % келісемін
C1	Инфографика оқу материалын тезірек түсінуге көмектеседі	4,52 (0,64)	89,1	4,61 (0,61)	94,4
C2	Инфографика ұғымдар арасындағы байланысты (мысалы, ақпараттық үдерістердің реттілігін) көруге көмектеседі	4,47 (0,69)	87,5	4,56 (0,62)	88,9
C3	Инфографикамен ұсынылған материал есте жақсырақ сақталады	4,31 (0,75)	82,8	4,33 (0,77)	83,3
C4	Инфографика сабаққа деген қызығушылықты арттырады	4,44 (0,71)	85,9	4,50 (0,71)	88,9
C5	Инфографиканы өз бетінше жасау тақырыпты тереңірек талдауды талап етеді	4,25 (0,80)	79,7	4,39 (0,78)	83,3
C6	Инфографика жасау барысында ақпаратты іріктеу және құрылымдау дағдылары дамиды	4,39 (0,72)	84,4	4,44 (0,70)	88,9
C7	ЖИ құралдары инфографика жасауға кететін уақытты қысқартады	4,08 (0,88)	75,0	3,94 (0,94)	66,7
C8	ЖИ арқылы жасалған инфографика мазмұндық тексеруді қажет етеді	4,56 (0,61)	90,6	4,72 (0,57)	94,4
C9	Инфографиканы өз сабақтарымда қолдануға дайынмын	4,20 (0,79)	78,1	4,28 (0,83)	77,8
C10	Инфографиканы информатика сабағында жүйелі қолдану қажет деп санаймын	4,36 (0,74)	84,4	4,50 (0,71)	88,9
	Шкала бойынша жалпы орташа мән	4,36	83,8	4,43	85,6

Ең жоғары бағалар инфографиканың материалды тез түсінуге (C1), ұғымдар арасындағы байланысты көруге (C2) және есте сақтауға (C3) көмектесуіне берілді, бұл когнитивтік және құрылымдаушы функцияларды растайды. Респонденттердің басым бөлігі инфографиканың

қызығушылықты арттыратынын (С4), ал оны өз бетінше жасау тақырыпты тереңірек талдауды талап етіп, ақпаратты іріктеу мен құрылымдау дағдыларын дамытатынын (С5–С6) атап өтті. ЖИ-ге қатысты тұжырымдарда топтар арасындағы ең айқын айырмашылық байқалды: ЖИ уақытты үнемдейді деген тұжырым (С7) мектептегі мұғалімдерде 4 балдан төмен бағаланған жалғыз тұжырым болды, ал ЖИ нәтижесін мазмұндық тексеру қажет деген тұжырым (С8) екі топта да ең жоғары бағаны алды, бұл 2-суреттегі міндетті тексеру кезеңінің негізділігін растайды. Респонденттердің шамамен төрттен үші инфографиканы өз сабағында қолдануға дайын (С9), ал одан да көбі оны жүйелі қолдануды қажет деп санайды (С10); бұл инфографика жасау дағдыларын әдістемелік пәндерде арнайы қалыптастыру керектігін көрсетеді.

Манн–Уитнидің U-критерийі студенттер мен мектептегі информатика мұғалімдерінің бағалары арасында статистикалық маңызды айырмашылықты анықтаған жоқ ($U = 521,5$; $p = 0,42$), яғни екі топтың пікірі негізінен сәйкес келеді.

Инфографиканы қолдануға ең тиімді кезеңдер ретінде студенттер жаңа материалды түсіндіруді (71,9%), бекітуді (65,6%) және қайталау мен жалпылауды (59,4%) атады; мектептегі информатика мұғалімдерінде де осы үш кезең алдыңғы орында болды (77,8%, 61,1% және 72,2%). Негізгі қиындықтар: мазмұнға сәйкес визуалды бейнені таңдау (34,4%), ЖИ мәтініндегі терминологиялық және логикалық қателер (28,1%), жасауға кететін уақыт (18,8%), сервистерде қазақ тіліндегі шаблондар мен қаріптердің жеткіліксіздігі (15,6%). Респонденттер қазақ тіліндегі инфографикалар банкін құруды және бағалау критерийлерін әзірлеуді ұсынды.

Дискуссия

Алынған нәтижелер инфографиканың оқу нәтижелеріне оң әсерін көрсеткен халықаралық зерттеулермен сәйкес келеді [14]. Болашақ информатика мұғалімдері үшін инфографика тек оқу құралы емес, кәсіби әрекеттің объектісі: оны жасау барысында олар мазмұнды талдауды және ұсыну форматын таңдауды үйренеді. ЖИ-ге қатысты нәтижелер инфографиканы жасау арнайы педагогикалық дизайнды қажет етеді деген ұстанымды қолдайды [15]: ЖИ жобалауды қолдайтын көмекші құрал, ал мазмұнның дұрыстығы үшін жауапкершілік информатика мұғалімде қалады.

Ұсынылған тәсілдің ерекшелігі. ChatGPT сияқты құралдарды мектептегі информатика мұғалімдері мен студенттердің көпшілігі қолданады, сондықтан ЖИ-ді пайдаланудың өзі жаңалық емес. Біздің тәсіліміздің айырмашылығы: 1) ЖИНИ платформалары нақты пәндік мазмұн мен инфографика түрі тұрғысынан критерийлер негізінде таңдалады (3-кесте); 2) бір платформаның орнына гибридті тізбек ұсынылады, мазмұнның дұрыстығы мен оқу мақсаттарына сәйкестігін тексеру мұғалімнің жауапкершілігінде болады; 3) ЖИ жіберген қателер сыни талдауға арналған оқу ресурсы ретінде пайдаланылады (2-кесте); 4) болашақ мұғалім дайын өнімді тұтынушы емес, ЖИ негізіндегі инфографиканы сабақ мақсатына сай жобалайтын педагогикалық дизайнер ретінде дайындалады, сонымен қатар, кодқа негізделген визуализация ақпаратты көрнекі ұсынумен қатар, білім алушылардың логикалық ойлау дағдыларын дамытуға мүмкіндік береді. Бұл тұжырым ЖИ жағдайында кеңейтілген TPACK нұсқаларында қарастырылатын технологияларды сыни бағалау құзыреттілігінің [4, 5] және жасанды интеллект құралдарын педагогикалық мақсаттарға сәйкес негіздеп қолдану қажеттігінің [6] маңыздылығын айқындайды.

Қорытынды

Зерттеу сұрақтарына сәйкес төмендегідей қорытындылар тұжырымдалды. Инфографика ақпараттық үдерістерді оқытуда когнитивтік, құрылымдаушы, мотивациялық, коммуникативтік, бақылау-диагностикалық және дамытушылық қызметтерді атқарады. ЖИ негізіндегі инфографиканың дәстүрлі нұсқадан басты айырмашылығы – оны адам мен ЖИ-дің бірлесіп әзірлеуі, мәтіндік сұраныстар арқылы құрастыру мүмкіндігі және дайын материалдың мазмұндық дұрыстығын міндетті түрде тексеру қажеттілігі (ЗС1). Зерттеу нәтижелері ЖИ негізіндегі инфографика платформаларының ешқайсысы оны әзірлеудің барлық кезеңдерін

бірдей деңгейде қамтамасыз етпейтінін көрсетті. Тілдік модельдер оқу мазмұнын құрылымдау мен мәтіндік ақпаратты жүйелеуге, Napkin AI және кодқа негізделген визуализация құралдары үдерістер мен олардың арасындағы байланыстарды сызбалық түрде көрсетуге, ал Canva дайын материалды көркем безендіруге қолайлы. Осыған байланысты инфографика әзірлеуде аталған құралдардың мүмкіндіктерін өзара толықтыратын гибридті тәсілді қолдану ұсынылады (ЗС2). Болашақ информатика мұғалімдерін ақпараттық үдерістерді оқытуда ЖИ негізіндегі инфографиканы қолдануға дайындаудың төрт компоненттен тұратын әдістемесі, практикалық тапсырмалар жүйесі және өзара байланысты алты кезеңді жүзеге асыру үлгісі әзірленді (ЗС3). Екі топ қатысушыларының да инфографиканы қолдануға оң баға бергенін көрсетті: бағалаудың жалпы орташа көрсеткіштері тиісінше 4,36 және 4,43 болды (ЗС4). Сонымен қатар, қатысушылар ЖИ көмегімен әзірленген материалдардың мазмұндық дұрыстығын міндетті түрде тексеру қажеттігін атап өтті. Топтар арасындағы бағалау айырмашылығы статистикалық тұрғыдан маңызды болмады. Қорытындылай келе, ақпараттық үдерістердің мазмұны мен инфографика түріне сәйкес ЖИ негізіндегі инфографика құралдарын критерийлер бойынша іріктеу және оларды кезең-кезеңімен қолдану тәсілдерін болашақ информатика мұғалімдерін дайындау әдістемесіне кіріктірумен сипатталады. Зерттеу нәтижелерін жоғары оқу орындарында әдістемелік пәндерде және информатика мұғалімдерінің біліктілігін арттыру курстарында қолдануға болады.

Пайдаланылған дереккөздер тізімі

[1] Қазақстан Республикасында мектепке дейінгі, орта, техникалық және кәсіптік білім беруді дамытудың 2023–2029 жылдарға арналған тұжырымдамасы. ҚР Үкіметінің 2023 жылғы 28 наурыздағы № 249 қаулысы. https://adilet.zan.kz/kaz/docs/P2500000271

[2] Жакупова А., Жолшаева М. Инфографика педагогикалық процесте қолданылатын ақпаратты визуализациялау құралы ретінде // Педагогика и методы обучения. 2021. № 4(57). С. 5–9. https://doi.org/10.47344/sdu20bulletin.v57i4.603

[3] Лапчик М.П., Рагулина М.И., Семакин И.Г., Хеннер Е.К. Методика обучения информатике: учебное пособие. СПб.: Лань, 2016. 392 с.

[4] Mishra P., Warr M., Islam R. TPACK in the age of ChatGPT and Generative AI // Journal of Digital Learning in Teacher Education. 2023. Vol. 39(4). P. 235–251. https://doi.org/10.1080/21532974.2023.2247480

[5] Celik I. Towards Intelligent-TPACK: an empirical study on teachers' professional knowledge to ethically integrate artificial intelligence (AI)-based tools into education // Computers in Human Behavior. 2023. Vol. 138. Art. 107468. https://doi.org/10.1016/j.chb.2022.107468

[6]Zawacki-Richter O., Marín V.I., Bond M., Gouverneur F. Systematic review of research on artificial intelligence applications in higher education – where are the educators? // International Journal of Educational Technology in Higher Education. 2019. Vol. 16. Art. 39. https://doi.org/10.1186/s41239-019-0171-0

[7] Kasneci E., Sessler K., Küchemann S. et al. ChatGPT for good? On opportunities and challenges of large language models for education // Learning and Individual Differences. 2023. Vol. 103. Art. 102274. https://doi.org/10.1016/j.lindif.2023.102274

[8] Шекербекова Ш., Базарбаева Ү. Білімді цифрландыру жағдайында болашақ информатика мұғалімдерін ақпараттық үдерістерді оқытуға дайындау қажеттігі // Абай атындағы ҚазҰПУ Хабаршысы. Физика-математика ғылымдары сериясы. 2025. № 91(3). Б. 335–345. https://doi.org/10.51889/2959-5894.2025.91.3.030

[9] Bekezhanova A., Bidaibekov Y., Mametzhanova N. Using infographics to teach object-oriented programming to future computer science teachers // Journal of e-Learning and Knowledge Society. 2023. Vol. 19(2). P. 60–67. https://doi.org/10.20368/1971-8829/1135752

[10] Бондарева К.Д., Исаева М.А. Инфографика в обучении информатике: эффективный способ представления материала для VII класса // Информатика в школе. 2024. № 1. С. 28–34. https://doi.org/10.32517/2221-1993-2024-23-1-28-34

[11] Вербицкий А.А. Активное обучение в высшей школе: контекстный подход. М.: Высшая школа, 1991. 207 с.

[12] Лаптев В.В. Изобразительная статистика: введение в инфографику. СПб.:Эйдос, 2012. 180 с.

[13] Mayer R.E. Multimedia Learning. 2nd ed. New York: Cambridge University Press, 2009. 304 p.

[14] Elaldı Ş., Çifçi T. The effectiveness of using infographics on academic achievement: a meta-analysis and a meta-thematic analysis // Journal of Pedagogical Research. 2021. Vol. 5(4). P. 92–118. https://doi.org/10.33902/JPR.2021473498

[15] Ozdamli F., Ozdal J. Developing an instructional design for the design of infographics and the evaluation of infographic usage in teaching // EURASIA Journal of Mathematics, Science and Technology Education. 2018. Vol. 14(4). P. 1197–1219. https://doi.org/10.29333/ejmste/81868

References

[1] Қazaқstan Respublikasynda mektepke dejingi, orta, tehnikalyқ zhәne kәsiptik bilim berudi damytudyң 2023–2029 zhyldarғa arnalғan tұzhyrymdamasy [Concept for the Development of Preschool, Secondary, and Technical and Vocational Education in the Republic of Kazakhstan for 2023–2029]. ҚR Ykimetinің 2023 zhylғy 28 nauryzdaғy № 249 қaulysy. https://adilet.zan.kz/kaz/docs/P2500000271

[2] Zhakupova A., Zholshaeva M. (2021) Infografika pedagogikalyқ proceste қoldanylatyn aқparatty vizualizacijalau құraly retinde [Infographics as a tool for visualizing information used in the pedagogical process] // Pedagogika i metody obuchenija. № 4(57). 5–9. https://doi.org/10.47344/sdu20bulletin.v57i4.603

[3] Lapchik M.P., Ragulina M.I., Semakin I.G., Henner E.K. (2016) Metodika obuchenija informatike [Methodology of Teaching Computer Science]: uchebnoe posobie. SPb.: Lan', 392.

[4] Mishra P., Warr M., Islam R. (2023) TPACK in the age of ChatGPT and Generative AI. Journal of Digital Learning in Teacher Education. Vol. 39(4). 235–251. https://doi.org/10.1080/21532974.2023.2247480

[5] Celik I. (2023) Towards Intelligent-TPACK: an empirical study on teachers' professional knowledge to ethically integrate artificial intelligence (AI)-based tools into education // Computers in Human Behavior. Vol. 138. Art. 107468. https://doi.org/10.1016/j.chb.2022.107468

[6] Zawacki-Richter O., Marín V.I., Bond M., Gouverneur F. (2019) Systematic review of research on artificial intelligence applications in higher education – where are the educators? International Journal of Educational Technology in Higher Education. Vol. 16. Art. 39. https://doi.org/10.1186/s41239-019-0171-0

[7] Kasneci E., Sessler K., Küchemann S. et al. (2023) ChatGPT for good? On opportunities and challenges of large language models for education // Learning and Individual Differences. Vol. 103. Art. 102274. https://doi.org/10.1016/j.lindif.2023.102274

[8] Shekerbekova Sh., Bazarbaeva Y. (2025) Bilimdi cifrlandyru zhaғdajynda bolashaқ informatika mұғalimderin aқparattyқ yderisterdi oқytuға dajyndau қazhettigi [The need to prepare future computer science teachers to teach information processes in the context of the digitalization of education.] // Abaj atyndaғy ҚazҰPU Habarshysy. Fizika-matematika ғylymdary serijasy. №91(3). 335–345. https://doi.org/10.51889/2959-5894.2025.91.3.030

[9] Bekezhanova A., Bidaibekov Y., Mametzhanova N. (2023) Using infographics to teach object-oriented programming to future computer science teachers // Journal of e-Learning and Knowledge Society. 2023. Vol. 19(2). 60–67. https://doi.org/10.20368/1971-8829/1135752

[10] Bondareva K.D., Isaeva M.A. (2024) Infografika v obuchenii informatike: jeffektivnyj sposob predstavlenija materiala dlja VII klassa [Infographics in Computer Science Education: An Effective Way to Present Material for 7th Grade] // Informatika v shkole. № 1. 28–34. https://doi.org/10.32517/2221-1993-2024-23-1-28-34

[11] Verbickij A.A. (1991) Aktivnoe obuchenie v vysshej shkole: kontekstnyj podhod [Active Learning in Higher Education: A Contextual Approach. Visual]. M.: Vysshaja shkola, 207.

[12] Laptev V.V. (2012) Izobrazitel'naja statistika: vvedenie v infografiku [Statistics: An Introduction to Infographics.]. SPb.: Jejdos, 180.

[13] Mayer R.E. (2009) Multimedia Learning. 2nd ed. New York: Cambridge University Press, 304.

[14] Elaldı Ş., Çifçi T. (2021) The effectiveness of using infographics on academic achievement: a meta-analysis and a meta-thematic analysis // Journal of Pedagogical Research. Vol. 5(4). 92–118. https://doi.org/10.33902/JPR.2021473498

[15] Ozdamli F., Ozdal J. Developing an instructional design for the design of infographics and the evaluation of infographic usage in teaching // EURASIA Journal of Mathematics, Science and Technology Education. 2018. Vol. 14(4). P. 1197–1219. https://doi.org/10.29333/ejmste/81868

МРНТИ 29.03.85

https://doi.org/10.51889/2959-5894.2026.95.3.030

К.Ж. Шетиева[1*], А.Н. Салыкова[1*], А. Есейкызы[1], У.Э. Мамбетакунов[2]

[1]Жетысуский университет имени И.Жансугурова, г.Талдыкорган, Казахстан
[2]Бишкекский музыкально-педагогический институт имени Т. Эрматова,
г.Бишкек, Кыргызстан
[*]*e-mail: karlygash.shetiyeva@gmail.com*

СИСТЕМНЫЙ АНАЛИЗ СИМУЛЯТОРОВ ДОПОЛНЕННОЙ РЕАЛЬНОСТИ В ОБУЧЕНИИ ФИЗИКЕ

Аннотация

В статье представлен системный анализ симуляторов дополненной реальности, применяемых для визуализации физических законов и явлений в процессе обучения физике. Актуальность исследования обусловлена необходимостью повышения наглядности и доступности абстрактных физических понятий при изучении механики, оптики, молекулярной физики, электричества и магнетизма. Цель исследования заключается в сравнительном анализе образовательных возможностей современных AR-приложений и определении их педагогического потенциала. Методологическую основу исследования составили методы системного, функционального и сравнительного анализа семи AR-приложений: MARVLS AR Physics 1 Mechanics, Merge Explorer, ARLOOPA, ARPhysycs2, AR.X Optics, ObjectViewer и Merpi AR Lab. Оценка проводилась по критериям визуализации, интерактивности, физической корректности моделей и дидактической эффективности. Результаты исследования показали, что наибольший образовательный эффект достигается при сочетании интерактивного взаимодействия, параметрического управления моделями и физически корректной визуализации процессов. Наиболее высокие показатели педагогического потенциала продемонстрировали приложения AR.X Optics, Merge Explorer и ARPhysycs2. Научная новизна исследования заключается в применении единой системы критериев для комплексного сравнительного анализа AR-симуляторов, используемых в обучении физике. Практическая значимость исследования заключается в возможности использования полученных результатов при выборе и интеграции AR-технологий в образовательный процесс по физике.

Ключевые слова: дополненная реальность, визуализация физических процессов, обучение физике, цифровые образовательные технологии, интерактивное обучение, иммерсивные технологии.

Қ.Ж.Шетиева[1], А.Н.Салықова[1], А.Есейқызы[1], У.Э. Мамбетакунов[2]
[1]І.Жансүгіров атындағы Жетісу университеті, Талдықорған, Қазақстан
[2]Т. Ерматов атындағы Бішкек музыкалық-педагогикалық институты, Бішкек қ., Қырғызстан

ФИЗИКАНЫ ОҚЫТУДАҒЫ КЕҢЕЙТІЛГЕН ШЫНДЫҚ СИМУЛЯТОРЛАРЫНЫҢ ЖҮЙЕЛІК ТАЛДАУЫ

Аңдатпа

Мақалада физиканы оқыту процесінде физикалық заңдар мен құбылыстарды визуализациялауға арналған толықтырылған шындық симуляторларына жүйелі талдау ұсынылған. Зерттеудің өзектілігі механика, оптика, молекулалық физика, электр және магнетизм бөлімдерін оқыту барысында абстрактілі физикалық ұғымдардың көрнекілігі мен қолжетімділігін арттыру қажеттілігімен негізделеді. Зерттеудің мақсаты – заманауи AR-қосымшалардың білім беру мүмкіндіктеріне салыстырмалы талдау жүргізу және олардың педагогикалық әлеуетін анықтау. Зерттеудің әдіснамалық негізін MARVLS AR Physics 1 Mechanics, Merge Explorer, ARLOOPA, ARPhysycs2, AR.X Optics, ObjectViewer және Merpi AR Lab сияқты жеті AR-қосымшаны жүйелік, функционалдық және салыстырмалы талдау әдістері құрады. Бағалау визуализация деңгейі, интерактивтілік, модельдердің физикалық дұрыстығы және дидактикалық тиімділігі критерийлері бойынша жүргізілді. Зерттеу нәтижелері өзара интерактивті әрекеттесу, модель параметрлерін басқару және физикалық процестердің ғылыми тұрғыдан дұрыс визуализациясы үйлескен жағдайда білім беру тиімділігі

жоғары болатынын көрсетті. Педагогикалық әлеуеті ең жоғары қосымшалар ретінде AR.X Optics, Merge Explorer және ARPhysycs2 анықталды. Зерттеудің ғылыми жаңалығы физиканы оқытуда қолданылатын толықтырылған шындық (AR) симуляторларын кешенді салыстырмалы талдауға арналған бірыңғай бағалау критерийлері жүйесін қолдануымен сипатталады. Зерттеудің практикалық маңыздылығы алынған нәтижелерді физиканы оқыту процесінде AR-технологияларды таңдау және енгізу барысында қолдану мүмкіндігімен айқындалады.

Түйін сөздер: толықтырылған шындық, физикалық процестерді визуализациялау, физиканы оқыту, цифрлық білім беру технологиялары, интерактивті оқыту, иммерсивті технологиялар.

K.Shetiyeva[1], A.Salykova[1], A.Yesseikyzy[1], U.E. Mambetakunov[2]
[1]Zhetysu University named after I. Zhansugurov, Taldykorgan, Kazakhstan
[2]T. Ermatov Bishkek Music and Pedagogical Institute, Bishkek, Kyrgyzstan

SYSTEM ANALYSIS OF AUGMENTED REALITY SIMULATORS IN PHYSICS TEACHING

Abstract

This article presents a systems analysis of augmented reality simulators used to visualize physical laws and phenomena in physics education. The relevance of this study stems from the need to improve the clarity and accessibility of abstract physical concepts in the study of mechanics, optics, molecular physics, electricity, and magnetism. The aim of the research is to conduct a comparative analysis of the educational capabilities of modern AR applications and to determine their pedagogical potential. The methodological framework of the study is based on the methods of systematic, functional, and comparative analysis applied to seven AR applications: MARVLS AR Physics 1 Mechanics, Merge Explorer, ARLOOPA, ARPhysycs2, AR.X Optics, ObjectViewer, and Merpi AR Lab. The evaluation was carried out according to the criteria of visualization quality, interactivity, physical accuracy of models, and didactic effectiveness. The results of the study demonstrated that the greatest educational effect is achieved through the integration of interactive engagement, parametric control of models, and physically accurate visualization of processes. The applications AR.X Optics, Merge Explorer, and ARPhysycs2 showed the highest indicators of pedagogical potential. The scientific novelty of this study lies in the application of a unified system of evaluation criteria for the comprehensive comparative analysis of augmented reality (AR) simulators used in physics education. The practical significance of the research lies in the possibility of applying the obtained results in the selection and integration of AR technologies into the educational process of physics teaching.

Keywords: augmented reality, visualization of physical processes, physics education, digital educational technologies, interactive learning, immersive technologies.

Введение

Основные положения

В данном исследований был проведен сравнительный анализ семи симуляторов дополненной реальности, используемых в преподавании физики. Результаты показывают, что эффективность технологий дополненной реальности определяется не только качеством визуализации, но и уровнем интерактивности, физической корректностью моделей и возможностью параметрического управления. По набору критериев наибольший педагогический потенциал продемонстрировали приложения AR.X Optics, Merge Explorer и ARPhysycs2. Результаты исследования показывают, что использование технологий дополненной реальности снижает когнитивную нагрузку и способствует формированию причинно-следственного понимания физических процессов.

Кроме того, AR-решений имеют ограничения, такие как, фрагментированность контента, ограниченность количественного анализа, зависимость от технических характеристик устройств. Цифровые и иммерсивные технологии все чаще используются в научной и образовательной среде, меняя привычные подходы к передаче знаний. Для физики, значительная часть которой основана на абстрактных моделях, математических уравнениях и ненаблюдаемых процессах, такие изменения приобретают особое значение. Традиционные средства наглядности – печатные иллюстрации, двухмерные графики и лабораторные

демонстрации, в некоторой степени недостаточны для целостного и интуитивного восприятия процессов.

В последние годы внимание исследователей привлекают технологии дополненной реальности (AR) – разновидность иммерсивных технологий, которые в реальном времени накладывают виртуальные объекты и динамические модели на физическую среду. В отличие от виртуальной реальности, дополненная реальность сохраняет визуальный и пространственный контекст реального мира. Это особенность делает ее перспективной для визуализации физических законов и явлений, происходящих в реальном пространстве и времени.

AR (Augmented Reality) – это среда, в которой физические и виртуальные объекты объединяются в реальном времени. Эта технология позволяет взаимодействовать с представленными данными и воспринимать их визуально [1].

Дополненная реальность представляет собой преобразующую парадигму для образовательного и профессионального контекстов, накладывая виртуальные 3D-элементы на реальную среду, что позволяет осуществлять иммерсивную визуализацию и взаимодействие со сложными явлениями [2].

Технологии дополненной реальности широко применяются не только в бизнесе, но и в образовании и различных областях науки. В условиях стремительного роста интереса к инновационным решениям в образовании дополненная реальность занимает ведущие позиции в образовательной среде. Применение дополненной реальности в лабораторных работах по физике может сделать эксперименты более эффективными с точки зрения понимания с помощью визуализации, а также позволяет повысить их результативность, способствуя более глубокому пониманию изучаемых процессов [3].

Дополненная реальность (AR) – это область технологий, которая предполагает плавное наложение созданных компьютером виртуальных изображений на реальный мир таким образом, чтобы виртуальный контент соответствовал объектам реального мира, и с ним можно было просматривать и взаимодействовать в режиме реального времени [4].

Использование AR-средств в физике позволяет визуализировать такие трудно наблюдаемые объекты и процессы, как силовые и векторные поля, молекулярное движение, волновые процессы, электромагнитные взаимодействия, оптические явления и атомные структуры. За счёт интерактивности и пространственной наглядности AR-симуляторы способствуют формированию причинно-следственных связей между математическим описанием физических законов и их физическим смыслом, снижая когнитивную нагрузку и повышая уровень вовлеченности обучающихся.

Применяя системы визуализации в промышленной сфере, образовании и научных исследованиях, можно обеспечить наглядное представление данных, что в свою очередь повышает эффективность анализа и принятия решений. Эти системы позволяют представлять сложные рабочие процессы и данные в интуитивно понятной форме, способствуя лучшему пониманию, анализу и передаче информации [2]. В работе Liru Chen и др. исследуется потенциал систем визуализации дополненной реальности (AR), которые улучшают мультимодальное восприятие и взаимодействие для принятия сложных решений. Предлагаемая система сочетает в себе физичность и интуитивность реального мира с иммерсивными и интерактивными возможностями AR-систем. Полученные результаты призваны помочь разработчикам систем визуализации, исследователям и практикам в изучении и использовании возможностей этого комплексного подхода, что в конечном итоге приведет к более увлекательному и захватывающему взаимодействию пользователей в различных областях применения [5].

Использование дополненной реальности (AR) позволяет интегрировать цифровую информацию в наше восприятие физического мира. A.Vidak в своей работе представляет всесторонний обзор ранее опубликованной литературы по внедрению AR в образование по физике на школьном и университетском уровнях. Обзор включает анализ 96 статей из баз

данных Scopus и Eric, все из которых были опубликованы в период с 1 января 2012 года по 1 января 2023 года. Оценивается, как дополненная реальность используется для облегчения изучения физики. Были обобщены потенциальные учебные мероприятия на основе дополненной реальности по различным темам физики, а также описаны возможности и проблемы, связанные с изучением физики на основе дополненной реальности. Было показано, что технологии дополненной реальности могут облегчить изучение физики, предоставляя дополнительные визуализации, оптимизируя когнитивную нагрузку, позволяя осуществлять тактильное обучение, сокращая время выполнения задач и способствуя совместному исследованию. Потенциальные недостатки использования дополненной реальности в преподавании физики в основном связаны с недостатками программных и аппаратных технологий (например, зависание камеры, задержка визуализации) и посторонняя когнитивная нагрузка (например, уделение большего внимания второстепенным деталям, чем формированию целевых знаний) [6].

Дополненная реальность (AR) - одна из технологий, которую можно интегрировать в процесс обучения. Благодаря технологическому прогрессу, внимание к дополненной реальности будет продолжать расти, к такому выводу пришли авторы Z.Zufahmi и др. в результате своих исследований в виде систематического обзора различных исследований, посвященных использованию дополненной реальности в научном обучении и положительному влиянию ее использования, которые были опубликованы в журналах, индексируемых Scopus. В этом обзоре использованы 44 статьи, опубликованные в журналах, индексируемых Scopus, за 2019-2023 годы. Результаты этого исследования показывают рост исследований в области дополненной реальности в области научного обучения. Исследования в области дополненной реальности в области научного обучения в основном проводятся с использованием квазиэкспериментального подхода [7].

Несмотря на большое количество исследований, посвященных использованию технологий дополненной реальности в образовании, основная часть работ посвящена изучению педагогической эффективности отдельных AR-приложений или описанию отдельных практик их внедрения. Однако в научной литературе отсутствуют исследования, направленные на всестороннее сравнение AR-симуляторов в соответствии с едиными критериями оценки. AR-симуляторы, применяемые для визуализации физических законов и явлений, остаются недостаточно систематизированными, по вопросам классификации, функционального сравнения и комплексной оценки. Отсутствие системного анализа затрудняет обоснованный выбор программных решений, а также разработку методических рекомендаций по их интеграции в учебный и исследовательский процесс.

Актуальность исследования определяется противоречием между активным распространением AR-приложений для обучения физике и отсутствием единых научно обоснованных критериев их сравнительной оценки. Несмотря на наличие исследований, посвящённых отдельным AR-средствам, недостаточно раскрыты вопросы систематизации, функционального сравнения и педагогической оценки AR-симуляторов по разделам физики. Это затрудняет выбор эффективных цифровых инструментов для учебного процесса и требует проведения комплексного анализа их образовательного потенциала.

В связи с этим возникает необходимость проведения системного анализа симуляторов дополненной реальности, используемых для визуализации физических законов и явлений, с учётом их технологических характеристик, охвата разделов физики, уровня интерактивности, степени соответствия физическим моделям и педагогического потенциала. Анализ позволяет выявить как сильные стороны существующих решений, так и их ограничения, также определить направления дальнейшего развития AR-технологий используемых в обучении физике и научной визуализации.

Исследования последних лет подтверждают положительное влияние технологий дополненной реальности на мотивацию обучающихся, развитие пространственного мышления, повышение наглядности и снижение когнитивной нагрузки при изучении

естественно-научных дисциплин [1–7]. Однако большинство существующих исследований сосредоточено преимущественно на отдельных приложениях или локальных педагогических экспериментах. Вопросы комплексного сравнительного анализа AR-симуляторов, их классификации и оценки педагогической эффективности остаются недостаточно изученными.

В связи с этим возникает необходимость системного анализа существующих AR-приложений, используемых для визуализации физических законов и явлений.

Цель данного исследования — провести систематический анализ симуляторов дополненной реальности, используемых в преподавании физики, и определить их функциональные, дидактические и технологические особенности.

Для достижения желаемого результата исследования решаются следующие задачи:

- анализ существующих симуляторов дополненной реальности;
- классификация приложений по разделам физики;
- определение уровня визуализации и интерактивности;
- оценка педагогического потенциала приложений дополненной реальности;
- выявление преимуществ и ограничений использования дополненной реальности в преподавании физики.

Гипотеза исследования: использование симуляций дополненной реальности способствует более эффективному усвоению физических закономерностей за счет повышения наглядности, интерактивности и вовлеченности учащихся.

Методология исследования

В последние годы стремительно развивается разработка мобильных приложений, использующих технологию дополненной реальности, открывая путь к разнообразным методам обучения. Эти технологии позволяют передавать сложные знания в понятной и наглядной форме. Использование таких приложений помогает учащимся лучше понимать физические явления, визуализировать их и даже, в некоторых случаях, экспериментировать с ними. Дополненная реальность широко используется на разных уровнях образования и оказывает положительное влияние на процесс обучения. Использования дополненной реальности в обучении может улучшить визуализацию, мотивацию, опыт обучения и вовлеченность учащихся в изучение естественных наук [7].

Использование приложений или инструментов дополненной реальности может улучшить успеваемость и мотивацию учащихся [8-11], облегчить процесс обучения [12], повысить интерес к урокам [13], расширить участие в обучении [14], улучшить пространственные навыки, способствовать совместному обучению, сделать процесс обучения более увлекательным и снизить когнитивную нагрузку [15-16].

Исследование проводилось в период с 2025 по 2026 год, проведен анализ мобильных приложений и цифровых образовательных платформ, использующие технологии дополненной реальности в обучении физики.

Выбор методов системного, функционального и сравнительного анализа обусловлен целью исследования. Системный анализ позволил рассмотреть AR-симуляторы как комплексные образовательные системы, включающие содержательный, технологический и дидактический компоненты. Функциональный анализ использовался для выявления возможностей приложений по визуализации физических процессов, уровню интерактивности и управлению параметрами моделей. Сравнительный анализ позволил сопоставить приложения по единым критериям и определить их педагогический потенциал.

Экспертная оценка проводилась группой специалистов в области преподавания физики и цифровых образовательных технологий. В экспертной оценке приняли участие 7 экспертов. В состав экспертной группы вошли преподаватели физики, имеющие опыт использования цифровых образовательных ресурсов в учебном процессе. Критериями отбора экспертов являлись наличие высшего педагогического образования, стаж преподавания физики не менее

пяти лет, опыт применения цифровых образовательных технологий и использования интерактивных средств обучения. Каждый эксперт независимо оценивал исследуемые AR-приложения по единым критериям, что обеспечило сопоставимость полученных результатов.

Для количественной оценки использовалась пятибалльная шкала Лайкерта, широко применяемая в педагогических исследованиях для экспертной оценки качественных характеристик образовательных ресурсов. Использование данной шкалы позволило перевести качественные показатели (визуализация, интерактивность, физическая корректность моделей и педагогический потенциал) в количественную форму и обеспечить возможность сравнительного анализа исследуемых приложений.

Каждый эксперт оценивал приложение по пяти критериям: визуализация, интерактивность, физическая корректность моделей, удобство интерфейса и педагогический потенциал. Оценка проводилась по пятибалльной шкале Лайкерта, где 1 балл означал низкий уровень выраженности критерия, а 5 баллов — высокий уровень. Итоговый балл по каждому критерию рассчитывался как среднее арифметическое экспертных оценок (1)

$$\bar{x} = \frac{\sum x_i}{n} \tag{1}$$

где x_i — оценка отдельного эксперта, n — количество экспертов.

Достоверность процедуры оценки обеспечивалась использованием единых критериев для всех анализируемых приложений и многократной экспертной проверкой результатов анализа. Такой подход позволил повысить объективность сравнительной оценки AR-симуляторов и обеспечить воспроизводимость исследования.

В качестве теоретической базы использовались подходы систематических обзоров, предложенные B. Kitchenham и S. Charters.

Исследование включало три этапа:

- Планирование исследования;
- Проведение сравнительного анализа;
- Интерпретация и систематизация результатов.

На этапе отбора приложений были определены следующие критерии:

- ориентация приложения на обучение физике;
- наличие интерактивных AR-моделей;
- возможность использования в образовательном процессе;
- доступность приложения на мобильных платформах;
- распространенность в образовательной практике.

В процессе исследования было проанализировано более 20 мобильных приложений, представленных в Google Play, App Store и научных публикациях. После предварительного отбора для детального анализа были выбраны семь наиболее репрезентативных приложений:

MARVLS AR Physics 1 Mechanics;

Merge Explorer;

ARLOOPA;

ARPhysycs2;

AR.X Optics;

ObjectViewer;

Merpi AR Lab.

Технология дополненной реальности используется в разделах механики: кинематика (визуализация движения тел по криволинейным траекториям), динамика (демонстрация законов Ньютона, взаимосвязи силы, массы и ускорения), статика (моделирование простых механизмов), а также в вопросах импульса (неупругие столкновения) и механической работы (преобразование потенциальной и кинетической энергии).

Приложений, которые можно использовать в разделе механики - MARVLS AR Physics 1 Mechanics, Merge Explorer, ARLOOPA, ObjectViewer (рисунок 1) .

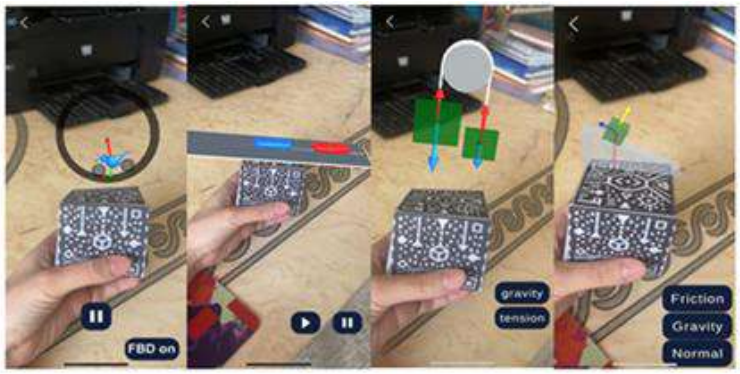


Рисунок 1. Трехмерные модели, относящиеся к главе механики в приложении MARVLS AR Physics 1 Mechanics

Раздел физики молекулярная физика и термодинамика охватывает представления о микроскопическом строении вещества, движении молекул и тепловых процессах. Поскольку явления, рассматриваемые в этом разделе физики, протекают в микромире и не видимы человеческому глазу, их невозможно напрямую наблюдать или продемонстрировать с помощью традиционных наглядных средств. Учащиеся получают теоретические знания о молекулярной структуре материи и движении молекул, но из-за отсутствия возможности увидеть эти процессы в реальности или наблюдать за ними экспериментально, они вынуждены воспринимать эти явления лишь на уровне воображения.

В такой ситуации использование технологий дополненной реальности является очень эффективным и удобным решением. С помощью дополненной реальности можно визуализировать трехмерные модели мельчайших частиц, которые невозможно увидеть или потрогать. При изучении этого раздела физики можно использовать приложения Merge Explorer и ARLOOPA (рисунок 2).

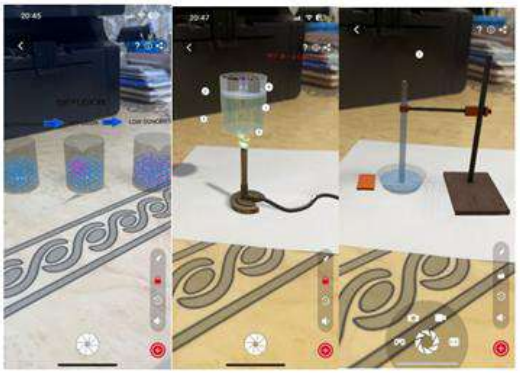


Рисунок 2. Симуляции используемые в разделах молекулярной физики и термодинамики приложения ARLOOPA

Раздел электричества и магнетизма относится к одному из важных и сложных разделов физики. В этом разделе рассматривается природа электрических явлений, возникновение электрического тока и его основные характеристики, такие как сила тока, напряжение, сопротивление, также такие важные понятия, как закон Ома. Кроме того, содержание раздела включает сложные явления, связанные с магнитным полем, силами Ампера и Лоренца, а также свойствами магнитного поля и его воздействием на другие частицы.

Приложения ARPhysycs2, Merge Explorer, ARLOOPA и Merpi AR Lab способствуют высококачественному обучению в области электричества и магнетизма (рисунок 3, 4).

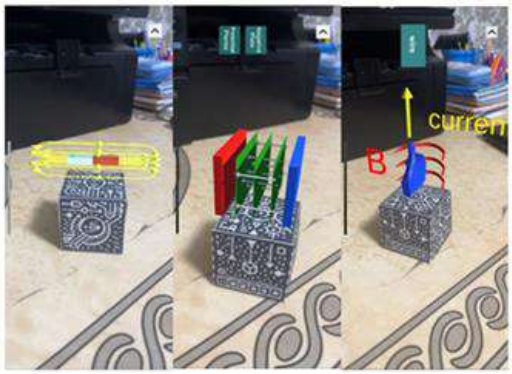


Рисунок 3. Интерфейс симуляций в приложении ARPhysycs2

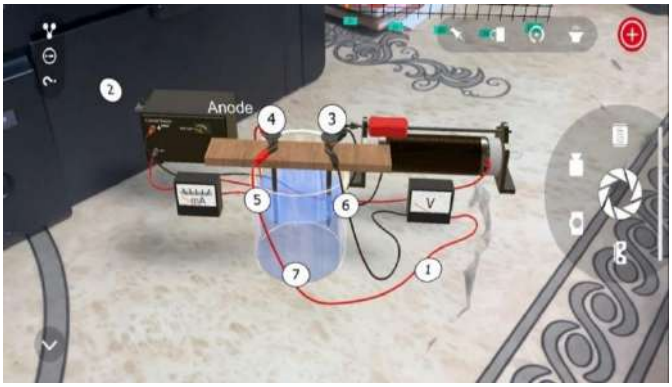


Рисунок 4. Отображение симуляции электролиза в приложении ARLOOPA

По сравнению с другими разделами физики, в оптике значительно больше приложений дополненной реальности, которые можно использовать в обучении. Такие приложения, как Merge Explorer, AR Physics 2, ARLOOPA, ObjectViewer, Assemblr EDU, Merpi AR Lab, а также специализированное приложение для оптики AR.Optics X, могут эффективно использоваться в преподавании оптики (рисунок 5).

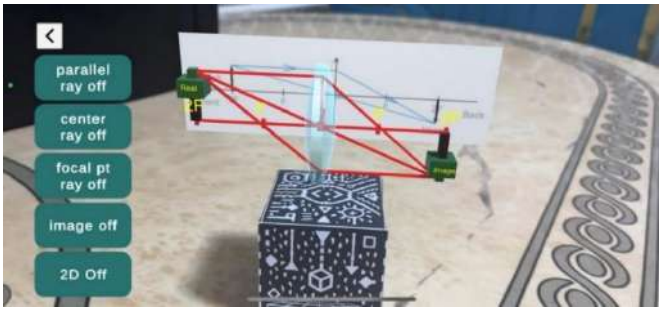


Рисунок 5. Представление симуляции «Ray Tracing» в приложении ARPhysycs2

Один из самых сложных разделов курса физики является раздел «Атом и атомное ядро». Этот раздел охватывает такие темы, как структура атома, свойства ядра, радиоактивность, ядерные реакции и их типы. Темы этого раздела наряду с теоретическим материалом требуют визуального представления. Использование визуальных и интерактивных инструментов обучения способствует повышению эффективности образовательного процесса. В этом контексте приложение ARLOOPA может стать эффективным вспомогательным средством обучения (рисунок 6).

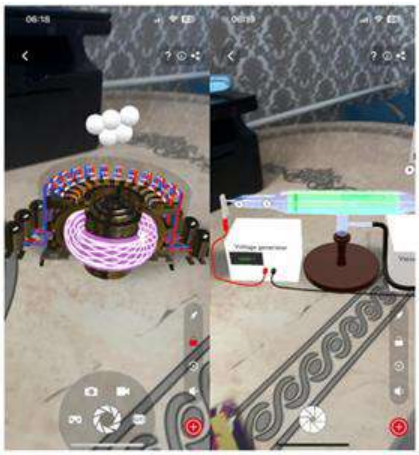


Рисунок 6. Приложение ARLOOPA

Для оценки приложений использовалась шкала Лайкерта от 1 до 5 баллов по следующим критериям:

- уровень визуализации физических процессов;
- интерактивность;
- физическая корректность моделей;
- удобство интерфейса;
- педагогический потенциал.

Провели контент-анализ приложений, который включал анализ тематического охвата разделов по физике, специфики взаимодействия пользователя с моделями, а также возможностей использования приложений в лабораторных и практических работах.

Результаты исследования

Результаты исследования показавают, что современные AR-приложения имеют большой потенциал для визуализации физических процессов и повышения эффективности обучения физике. Сравнительный анализ показал, что наивысший уровень визуализации демонстрируют приложения ARLOOPA (4,8 балла) и Merge Explorer (4,7 балла). Эти приложения дают хорошее трёхмерное представление физических процессов и охватывают широкий круг тем. Приложение AR.X Optics (4,8 балла) показал высокий уровень интерактивности. Он позволяет учащимся самостоятельно моделировать оптические эксперименты и изменять параметры взаимодействия с объектами. В разделе «Электричество и магнетизм» приложение ARPhysycs2 показало высокую физическую точность (4,7 балла), тем самым обеспечивая точное представление электрических и магнитных процессов. На основе этих параметров была проведена интегральная оценка каждого приложения (таблица 1).

Таблица 1. Оценка AR приложений

Приложение	*Раздел физики*	*Визуализация*	*Интерактивность*	*Физическая корректность*	*Педагогический потенциал*
MARVLS AR Physics	*Механика*	*4,6*	*4,2*	*4,5*	*4,4*
Merge Explorer	*Механика, Молекулярная физика, Оптика, Электричество и магнетизм*	*4,7*	*4,5*	*4,3*	*4,6*

ARLOOPA	*Все разделы*	*4,8*	*4,1*	*4,2*	*4,5*
ARPhysycs2	*Электричество и магнетизм, Оптика*	*4,5*	*4,4*	*4,7*	*4,6*
AR.X Optics	*Оптика*	*4,6*	*4,8*	*4,6*	*4,8*
ObjectViewer	*Приборы*	*4,2*	*3,2*	*4,4*	*3,9*
Merpi AR Lab	*Электричество и магнетизм, оптика*	*4,0*	*3,5*	*4,1*	*3,9*

Для повышения достоверности результатов были рассчитаны средние значения по каждому приложению и проведено ранжирование AR-симуляторов. Полученные данные показывают, что наиболее высокие интегральные показатели имеют AR.X Optics, ARPhysycs2 и Merge Explorer (таблица 2). Это связано с сочетанием высокой визуализации, интерактивности и физической корректности моделей.

Таблица 2. Ранжирование AR-симуляторов

Приложение	*Средний балл*	*Ранг*
AR.X Optics	*4,70*	*1*
Merge Explorer	*4,53*	*2*
ARPhysycs2	*4,55*	*2*
ARLOOPA	*4,40*	*4*
MARVLS AR Physics	*4,43*	*4*
ObjectViewer	*3,93*	*6*
Merpi AR Lab	*3,88*	*7*

Анализ распределения экспертных оценок показал устойчивое совпадение мнений экспертов относительно приложений AR.X Optics, ARPhysycs2 и Merge Explorer, которые получили наиболее высокие оценки по большинству критериев. Это свидетельствует о согласованности экспертных заключений и подтверждает объективность полученных результатов.

Полученные количественные оценки свидетельствуют о том, что высокий педагогический потенциал AR-приложений определяется не отдельными техническими характеристиками, а их комплексным сочетанием. Так, приложение AR.X Optics получило максимальную оценку по педагогическому потенциалу (4,8 балла) благодаря высокой интерактивности (4,8 балла). Возможность интерактивного взаимодействия позволяет обучающимся самостоятельно конструировать и изменять параметры оптических экспериментов. Это способствует формированию исследовательских навыков и развитию причинно-следственного понимания физических закономерностей.

Приложения Merge Explorer и ARPhysycs2 продемонстрировали высокие интегральные показатели вследствие оптимального сочетания визуализации физических процессов и научной корректности представляемых моделей. Напротив, более низкие показатели ObjectViewer и Merpi AR Lab связаны прежде всего с ограниченными возможностями интерактивного взаимодействия пользователя с моделями, что снижает их дидактическую эффективность.

Проведённый сравнительный анализ свидетельствует о том, что образовательная эффективность AR-приложений определяется комплексным сочетанием нескольких характеристик. Высокие итоговые оценки получили приложения, сочетающие качественную визуализацию физических процессов, интерактивное взаимодействие с виртуальными объектами и физически корректное моделирование явлений. Лидирующие позиции заняли

AR.X Optics, ARPhysycs2 и Merge Explorer, что подтверждает целесообразность их использования при изучении различных разделов школьного и вузовского курса физики.

Таким образом, результаты исследования показывают, что при выборе AR-симуляторов для обучения физике определяющее значение имеет не только качество визуализации, но и наличие механизмов активного взаимодействия обучающегося с моделируемыми объектами и процессами.

Результаты анализа показали, что AR-приложения наиболее эффективно используются при изучении следующих разделов физики: механика, оптика, электричество и магнетизм, молекулярная физика и термодинамика.

Установлено, что применение дополненной реальности способствует:

- повышению наглядности обучения;
- развитию пространственного мышления;
- развитию исследовательских навыков;
- снижению когнитивной нагрузки;
- повышению мотивации обучающихся.

Но были выявлены ограничения современных AR-систем:

- недостаточность количественного анализа;
- зависимость качества работы от технических характеристик устройств;
- недостаточная адаптация под различные уровни подготовки учащихся;
- фрагментарность отдельных проявлений.

Дискуссия

Полученные результаты подтверждают выводы современных исследований о высокой эффективности технологий дополненной реальности в естественнонаучном образовании. AR-технологии, в отличие от традиционных инструментов визуализации, позволяют не только демонстрировать физические процессы, но и интерактивное взаимодействие с моделями.

Результаты исследования показывают, что наибольший педагогический эффект достигается в тех случаях, когда у обучающегося имеется возможность изменять параметры модели и наблюдать последствия своих действий в режиме реального времени. Такой подход способствует причинно-следственному пониманию физических процессов и развитию исследовательских навыков.

Полученные данные согласуются с исследованиями A. Vidak, Z. Zufahmi, Lai J.W. и других авторов, указывающие на положительное влияние дополненной реальности на познавательную деятельность, мотивацию и пространственное восприятие учащихся.

Сравнительный анализ показал, что наиболее эффективно работают приложения с визуализацией, интерактивностью и возможностью экспериментального взаимодействия. Например, AR-приложения AR.X Optics и ARPhysycs2 предоставляют обучающимся возможность не только наблюдать за физическими процессами, но и моделировать различные сценарии путем изменения параметров системы, что приближает процесс обучения к исследовательской деятельности и помогает формировать устойчивые предметные компетенции.

Одним из важнейших преимуществ технологий дополненной реальности является возможность визуализации объектов и процессов, которые невозможно непосредственно наблюдать в реальных условиях. К таким процессам относятся распространение электромагнитных полей, движение элементарных частиц, молекулярные взаимодействия и распространение световых лучей. Пространственное представление таких явлений помогает снизить уровень абстракции учебного материала и облегчить понимание сложных физических законов.

В ходе анализа, было выявлено ряд ограничений существующих приложений дополненной реальности. Приложения в большинстве случаев ориентированы, в первую очередь, на

демонстрационный эффект, слабую интеграцию с математическим аппаратом физики. А также есть технические ограничения, связанные с производительностью мобильных устройств, стабильностью отслеживания и визуализацией качества.

Следовательно, эффективность использования дополнительной реальности зависит от методической организации образовательного процесса. Использование дополненной реальности без четко сформулированных учебных задач может привести к увеличению внешней когнитивной нагрузки, а также рассеянию внимания обучающегося на зрительные эффекты. В этой связи интеграция средств дополненной реальности должна сопровождаться разработкой конкретных методических планов, направленных на активное познание, исследовательскую деятельность и развитие аналитического мышления

Среди направлений дальнейших исследований можно выделить возможность создания AR-систем с адаптивной архитектурой, в которых элементы искусственного интеллекта могут отслеживать действия обучающегося и корректировать сценарий обучения в зависимости от его успехов и трудностей. А встроенная аналитика могла бы дать преподавателю содержательную обратную связь, но не в форме формальных оценок, а в виде данных о том, где и почему возникают пробелы в понимании.

Перспективными также являются разработка единых критериев оценки качества AR-симуляторов и разработка методических рекомендаций по их использованию в школьном и университетском образовании.

Аналитические выводы исследования:

- Наиболее высокие показатели получили AR.X Optics, Merge Explorer и ARPhysycs2;
- В разделах оптики и механики наблюдается наибольшее количество решений с дополненной реальностью;
- Большинство приложений имеют ограниченные средства количественного анализа;
- Сохраняется зависимость качества работы от технических характеристик мобильных устройств.

Педагогические выводы:

- AR способствует формированию пространственного мышления;
- Снижает когнитивную нагрузку при изучении абстрактных физических процессов;
- Повышает мотивацию и вовлеченность обучающихся;
- Наибольший эффект достигается при возможности самостоятельного управления параметрами модели.

Заключение

Системный анализ показал, что технологии дополненной реальности обладают большим потенциалом для повышения эффективности изучения физики и формирования устойчивого понимания физических закономерностей.

Исследование показало, что наиболее эффективными являются приложения дополненной реальности, которые сочетают в себе высококачественную визуализацию, высокий уровень интерактивности, физически точные модели и возможность параметрического управления. Эти приложения позволяют учащимся взаимодействовать с виртуальными объектами, моделировать физические процессы и получать обратную связь в режиме реального времени.

Результаты сравнительного анализа показали, что приложения AR.X Optics, Merge Explorer и ARPhysycs2 обладают наибольшим педагогическим потенциалом. Их использование способствует развитию пространственного мышления, повышает мотивацию учащихся, снижает когнитивную нагрузку и развивает исследовательские навыки.

Было установлено, что технология дополненной реальности более эффективна при изучении областей физики, связанных с высоким уровнем абстракции: механика, оптики, электричества и магнетизм, молекулярная физика и термодинамика. Пространственная визуализация процессов помогает преодолеть разрыв между теоретическим описанием физических явлений и их практическим восприятием.

Однако исследование выявило ряд ограничений современных решений в области дополненной реальности, таких как фрагментация образовательного контента, отсутствие интеграции с количественным анализом, зависимость качества приложения от технических характеристик устройств и отсутствие адаптации к различным уровням подготовки студентов.

Следует отметить, что настоящее исследование имеет ряд ограничений. Работа носит теоретико-аналитический характер и основана на экспертной оценке функциональных и дидактических характеристик AR-симуляторов. В рамках исследования не проводилась экспериментальная проверка эффективности использования рассматриваемых приложений с участием обучающихся. В связи с этим педагогическая апробация предложенных выводов рассматривается как перспективное направление дальнейших исследований, направленное на оценку влияния AR-симуляторов на учебные достижения, познавательную активность и мотивацию обучающихся.

Научная новизна исследования заключается в разработке и применении единой системы критериев оценки AR-симуляторов, включающей показатели визуализации, интерактивности, физической корректности моделей и педагогического потенциала. В отличие от большинства существующих исследований, ориентированных на анализ отдельных приложений, в работе выполнено комплексное сравнение AR-решений различных областях физики, основанное на едином методологическом подходе. Это позволило выявить закономерности, определяющие образовательную эффективность AR-технологий, а также определить наиболее перспективные направления их применения в физическом образовании.

Ограничения исследования заключаются в том, что анализ проводился на основе семи наиболее репрезентативных AR-приложений, доступных на мобильных платформах в период 2025–2026 гг. Исследование не включало педагогический эксперимент с участием обучающихся и основывалось на экспертной оценке функциональных характеристик приложений. В дальнейшем целесообразно дополнить результаты экспериментальной проверкой влияния выявленных характеристик на учебные достижения учащихся.

Практическая значимость исследования заключается в том, что полученные результаты могут быть использованы преподавателями физики, разработчиками цифровых образовательных ресурсов и исследователями при проектировании интерактивных учебных сред и цифровых лабораторных работ.

Таким образом, дополненная реальность может рассматриваться как эффективный инструмент современной цифровой педагогики, позволяющий глубже и более осознанно понимать физические законы и явления при условии ее методической интеграции в образовательный процесс.

Список использованных источников

[1] Lai J. W., Cheong K. H. Educational opportunities and challenges in augmented reality: Featuring implementations in physics education //Ieee Access. – 2022. – Т. 10. – С. 43143-43158 https://ieeexplore.ieee.org/abstract/document/9755165

[2] Mansour, Nasser, Aras, Ceren and other Embodied learning of science concepts through augmented reality technology. Education and Information Technologies, 2025 https://www.scopus.com/pages/publications/105002955776?origin=resultslist

[3] B. Arymbekov, K. Turekhanova, D. D. Alipbayev, Y. Tursanova DEVELOPMENT OF AUGMENTED REALITY APPLICATION FOR PHYSICS AND GEOPHYSICS LABORATORY. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 2023 https://doi.org/10.5194/isprs-archives-XLVIII-5-W2-2023-19-2023

[4] Azuma, R. T. (1997). A survey of augmented reality. Presence 6, 355–385 https://www.ronaldazuma.com/papers/ARpresence.pdf

[5] Liru Chen, Hantao Zhao, Chenhui Shi, Youbo Wu, Xuewen Yu, Wenze Ren, Ziyi Zhang, Xiaomeng Shi, Enhancing Multi-Modal Perception and Interaction: An Augmented Reality Visualization System for Complex Decision Making. Systems 2024, 12(1), 7; https://doi.org/10.3390/systems12010007

[6] A. Vidak, I. Movre Šapić, V. Mešić, V. Gomzi Augmented reality technology in teaching about physics: a systematic review of opportunities and challenges. European journal of physics, 2023 https://doi.org/10.1088/1361-6404/ad0e84

[7] Z. Zufahmi, Fatchur Rohman, Murni Sapta Sari Augmented reality in science learning: A systematic literature review. JPBI (Jurnal Pendidikan Biologi Indonesia), Vol. 11 No. 1 (2025): MARCH https://www.researchgate.net/publication/390574197_Augmented_reality_in_science_learning_A_systematic_literature_review

[8] J. Ferrer-Torregrosa, J. Torralba, M. A. Jimenez, S. García, and J. M. Barcia, "ARBOOK: Development and assessment of a tool based on augmented reality for anatomy," J. Sci. Educ. Technol., vol. 24, no. 1, pp. 119–124, 2015.https://doi.org/10.1007/s10956-014-9526-4 https://www.researchgate.net/publication/272497909_ARBOOK_Development_and_Assessment_of_a_Tool_Based_on_Augmented_Reality_for_Anatomy

[9] F. Najwa Rusli, A. N. Zulkifli, M. N. Bin Saad, and Y. Md.Yussop, "A study of students' motivation in using the Mobile Arc Welding Learning app," Int. J. Interact. Mob. Technol., vol. 13, no. 10, p. 89, 2019. https://doi.org/10.3991/ijim.v13i10.11305

[10] S. Vedadi, Z. B. Abdullah, and A. D. Cheok, "The effects of multi-sensory augmented reality on students' motivation in English language learning," in 2019 IEEE Global Engineering Education Conference (EDUCON), 2019. https://doi.org/10.1109/EDUCON.2019.8725096

[11] M. E. C. Santos et al., "Augmented reality as multimedia: the case for situated vocabulary learning," Res. pract. technol. enhanc. learn., vol.11, no.1, p.4, 2016.
https://doi.org/10.1186/s41039-016-0028-2 .

[12] K. J. Carlson and D. J. Gagnon, "Augmented reality integrated simulation education in health care," Clin. Simul. Nurs., vol. 12, no. 4, pp. 123–127, 2016 https://doi.org/10.1016/j.ecns.2015.12.005

[13] J. Zhang, Y.-T. Sung, H.-T. Hou, and K.-E. Chang, "The development and evaluation of an augmented reality-based armillary sphere for astronomical observation instruction," Com-put. Educ., vol. 73, pp. 178–188, 2014 https://doi.org/10.1016/j.compedu.2014.01.003

[14] D. M. Bressler and A. M. Bodzin, "A mixed methods assessment of students' flow experi-ences during a mobile augmented reality science game: Flow experience with mobile AR," J. Comput. Assist. Learn., vol. 29, no. 6, pp. 505–517, 2013. https://doi.org/10.1111/jcal.12008

[15] S. Küçük, S. Kapakin, and Y. Göktaş, "Learning anatomy via mobile augmented reality: Effects on achievement and cognitive load: Learning Anatomy," Anat. Sci. Educ., vol. 9, no. 5, pp. 411–421, 2016. https://doi.org/10.1002/ase.1603

[16] M. Sirakaya and D. Alsancak Sirakaya, "Trends in educational augmented reality studies: A systematic review," Malays. Online J. Educ. Technol., vol. 6, no. 2, pp. 60–74, 2018 https://doi.org/10.17220/mojet.2018.02.005

[17] Kitchenham B, Charters S (2007) Guidelines for performing systematic literature reviews in software engineering version 2.3. Engineering 45(4ve):1051.
https://doi.org/10.1145/1134285.1134500