МРНТИ 20.51.19

10.51889/2959-5894.2026.93.1.013

А.Б. Бажибаева[1*], Д.Н. Исабаева[1], С.М. Алдашев[1], С. Байпакбаева[1]

[1]Казахский национальный университет имени аль-Фараби, г. Алматы, Казахстан
[*]*e-mail:bakytbekkyzyainur@gmail.com*

КОМПЬЮТЕРНОЕ ЗРЕНИЕ ДЛЯ ВОВЛЕЧЕННОСТИ УЧАЩИХСЯ В ОНЛАЙН ОБУЧЕНИИ: ОБЗОР ЛИТЕРАТУРЫ

Аннотация

В данном обзоре систематизированы исследования 2020-2025 годов, посвящённые автоматическому обнаружению вовлечённости студентов в онлайн-обучении на основе компьютерного зрения и глубокого обучения. Вовлечённость рассматривается как многомерный конструкт, включающий эмоциональные, когнитивные и поведенческие компоненты, которые играют ключевую роль в академической успеваемости. Анализируются используемые наборы данных (публичные и специализированные), методы извлечения признаков (мимика, направление взгляда, поза головы, движения тела), архитектуры моделей (CNN, LSTM, трансформеры) и подходы к мультимодальной интеграции. Показан переход от экспериментальных решений к комплексным системам реального времени, однако отмечены сохраняющиеся проблемы обобщаемости моделей, нехватки разнообразных данных и этических рисков, связанных с конфиденциальностью и обработкой персональной информации. Делается вывод о необходимости разработки стандартизированных, этически обоснованных и адаптируемых к различным условиям методов оценки вовлечённости.

Ключевые слова: компьютерное зрение, глубокое обучение, вовлечённость студентов, онлайн-обучение, мультимодальный анализ.

А.Б. Бажибаева[1], Д.Н.Исабаева[1], С.М. Алдашев[1], С.Байпакбаева[1]
[1]Казахский национальный университет имени-аль-Фараби г. Алматы, Казахстан

КОМПЬЮТЕРЛІК КӨРУ АРҚЫЛЫ БІЛІМ АЛУШЫЛАРДЫҢ ОНЛАЙН ОҚЫТУҒА БЕЛСЕНДІ ҚАТЫСУЫН АНЫҚТАУ: ӘДЕБИЕТТЕРГЕ ШОЛУ

Аңдатпа

Бұл шолу компьютерлік көру және терең оқытуды қолдана отырып, онлайн оқуға студенттердің қатысуын автоматты түрде анықтау бойынша 2020 жылдан 2025 жылға дейінгі зерттеулерді жүйелейді. Қатысу академиялық көрсеткіштерде маңызды рөл атқаратын эмоционалды, когнитивті және мінез-құлық компоненттерін қамтитын көп өлшемді құрылым болып саналады. Шолуда пайдаланылған деректер жиынтығы (қоғамдық және мамандандырылған), ерекшеліктерді алу әдістері (бет-әлпет, көзқарас бағыты, бас позасы, дене қимылдары), модель архитектуралары (CNN, LSTM, трансформаторлар) және мультимодальды интеграцияға тәсілдер талданады. Эксперименттік шешімдерден күрделі нақты уақыт жүйелеріне көшу көрсетілгенімен, модельді жалпылаудағы тұрақты қиындықтар, әртүрлі деректердің жетіспеушілігі және құпиялылық пен жеке ақпаратты өңдеумен байланысты этикалық тәуекелдер сақталуда. Онда қатысуды бағалау үшін стандартталған, этикалық тұрғыдан негізделген және бейімделетін әдістерді әзірлеу қажеттілігі туралы қорытынды жасалады.

Түйін сөздер: компьютерлік көру, терең оқыту, студенттердің қатысуы, онлайн оқыту, мультимодальды талдау.

A.B. Bazhibayeva[1], D.N. Isabaeva[1], S.M. Aldashev[1], S.Baipakbayeva[1]
[1]Kazakh National University named after al-Farabi, Almaty, Kazakhstan

COMPUTER VISION FOR STUDENT ENGAGEMENT IN ONLINE LEARNING: LITERATURE REVIEW

Abstract

This review systematizes research on automatic detection of student engagement in online learning using computer vision and deep learning from 2020 to 2025. Engagement is considered a multidimensional construct

that includes emotional, cognitive, and behavioral components that play an important role in academic performance. The review analyzes the datasets used (public and specialized), feature extraction methods (facial expression, gaze direction, head posture, body movements), model architectures (CNN, LSTM, transformers), and approaches to multimodal integration. Although the transition from experimental solutions to complex real-time systems is demonstrated, persistent difficulties in model generalization, lack of diverse data, and ethical risks related to privacy and processing of personal information remain. It concludes that there is a need to develop standardized, ethically sound, and adaptable methods for assessing participation.

Keywords: computer vision, deep learning, student participation, online learning, multimodal analysis.

Введение

Стремительное развитие онлайн-обучения коренным образом изменило образовательный ландшафт во всем мире. Этот цифровой переход, значительно ускоренный пандемией COVID-19, утвердил онлайн-обучение не как временную альтернативу традиционному классному обучению, а как устойчивое явление [1, 5, 12]. Однако эта трансформация породила устойчивые проблемы, особенно в отношении сохранения вовлеченности студентов в виртуальных средах, где преподаватели лишены немедленной визуальной обратной связи, характерной для традиционного очного обучения [4, 12].

Вовлеченность студентов давно признана критическим фактором академического успеха, удержания знаний и завершения курсов [1, 7]. В традиционных классах учителя интуитивно оценивают вовлеченность через наблюдаемое поведение – зрительный контакт, позу, мимику и паттерны участия – и соответствующим образом корректируют стратегии обучения [4, 7].

Сама вовлеченность является многомерным конструктом, включающим поведенческий, когнитивный и эмоциональный компоненты [1, 4, 7]. Согласно последним систематическим обзорам [4], эмоциональная вовлеченность является наиболее изученным измерением (45.13%), за которым следуют многомерные подходы (38.05%), когнитивная вовлеченность (14.16%) и поведенческая вовлеченность (2.65%). Поведенческая вовлеченность относится к наблюдаемой активности и участию в задачах; когнитивная вовлеченность представляет умственные усилия и глубокую обработку информации; эмоциональная вовлеченность охватывает аффективные реакции на обучение, включая интерес, удовольствие, скуку или разочарование [4, 7]. Эти измерения динамически взаимосвязаны, однако их сложность затрудняет всестороннюю оценку с помощью традиционных методов, таких как опросники или листы наблюдения, которые являются ретроспективными, субъективными и лишены временной детализации [4, 7].

Компьютерное зрение (КЗ) стало перспективным технологическим ответом на этот оценочный разрыв. Анализируя видеоданные, снятые стандартными веб-камерами, широко используемыми в конфигурациях онлайн-обучения, системы КЗ могут неинвазивно обнаруживать мимику, направление взгляда, позу головы и движения тела, превращая эти визуальные сигналы в количественные показатели вовлеченности [4, 7, 11]. В отличие от физиологических датчиков (ЭЭГ, ЭКГ, кожно-гальваническая реакция), которые требуют специального оборудования и вызывают проблемы конфиденциальности, подходы на основе камер предлагают масштабируемость и экологическую валидность для реальных образовательных условий [4, 11, 12].

Последние пять лет (2020–2025) стали свидетелями значительных успехов в обнаружении вовлеченности на основе компьютерного зрения. Систематический обзор, проведенный Карбал и др. [4], охватил 113 исследований за период 2014-2025 гг. и показал, что этот период характеризуется развитием технологий глубокого обучения – сверточных нейронных сетей (CNN), рекуррентных архитектур (LSTM) и моделей на основе трансформеров – что значительно улучшило точность и надежность распознавания мимики, оценки направления взгляда и отслеживания позы [4, 7, 11].

На рисунке 1 представлено годовое распределение отобранных исследований, посвящённых обнаружению вовлечённости студентов с использованием методов компьютерного зрения за период 2020–2025 гг.

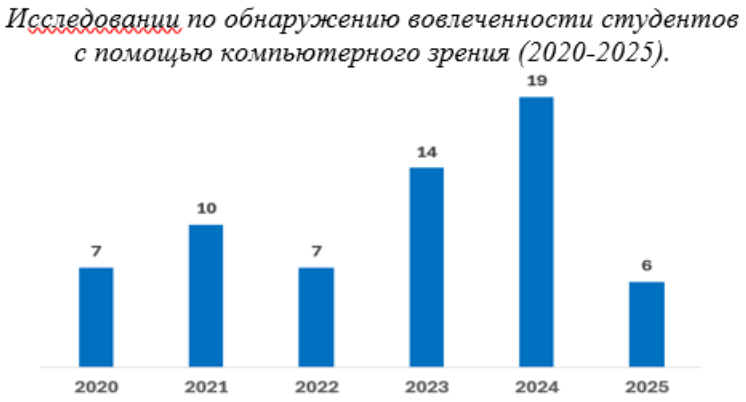

Рисунок 1. Годовое распределение отобранных исследований по обнаружению вовлеченности студентов с помощью компьютерного зрения (2020-2025).

Рисунок 1 показывает устойчивый рост интереса к данной тематике. 2023-2024 годы выделяются наибольшим количеством публикаций, что объясняется повышенным интересом к онлайн-обучению в постпандемический период [4, 12]. Доступность крупных эталонных наборов данных, таких как DAiSEE и серия челленджей EmotiW, позволила систематически сравнивать и бенчмаркить подходы [4, 13, 15]. Кроме того, глобальный переход на дистанционное обучение во время пандемии COVID-19 катализировал исследовательский интерес к инструментам, помогающим понимать и повышать вовлеченность студентов в цифровых средах [4, 5, 12]. Следовательно, литература 2020–2025 годов демонстрирует переход от исследований по доказательству концепции к сложным мультимодальным системам, которые объединяют мимические, глазные и телесные сигналы для прогнозирования вовлеченности в реальном времени [4, 7, 13].

Несмотря на эти технические достижения, остаются важные вопросы о надежности, обобщаемости и этичности развертывания таких систем в различных образовательных контекстах [4, 12]. Этот обзор литературы синтезирует последние исследования (2020–2025), посвященные анализу вовлеченности онлайн-обучающихся с помощью компьютерного зрения, рассматривая методологические подходы, технические архитектуры, наборы данных, метрики оценки и возникающие проблемы. Сосредоточившись на последних пяти годах, мы охватываем самые современные технологические разработки и гарантируем, что наши результаты отражают передовую практику. Обзор направлен на то, чтобы предоставить исследователям и педагогам всестороннее понимание текущего состояния области и определить направления будущих исследований.

Методы исследования

Этот обзор следовал структурированному подходу для выявления, отбора и анализа соответствующей литературы по обнаружению вовлеченности на основе компьютерного зрения. Стратегия поиска была направлена на рецензируемые журнальные статьи и материалы конференций, опубликованные в период с 2020 по 2025 год. Этот пятилетний период был намеренно выбран для охвата периода наиболее быстрых инноваций в применении технологий глубокого обучения и компьютерного зрения в образовании. За это время в отрасли наблюдалось появление сквозных моделей глубокого обучения, крупномасштабных наборов данных и усиление исследовательского фокуса на онлайн-обучении после пандемии COVID-19 [4, 5, 7].

Основные вопросы исследования (RQ).

1. Классификация и методы: Какие виды вовлеченности учащихся (эмоциональная, когнитивная, поведенческая) выделяются в современных исследованиях и какие методы их распознавания наиболее эффективны в различных образовательных средах?

2. Данные и алгоритмы: Какие наборы данных и архитектуры нейронных сетей (компьютерное зрение, анализ жестов и позы) являются приоритетными для точного определения состояния учащегося?

3. Оптимизация и оценка: Какие методы обработки данных и метрики эффективности позволяют достичь максимальной точности моделей при анализе вовлеченности в реальном времени?

Web of Science (WoS) и Scopus были выбраны в качестве основных баз данных из-за их широкого охвата высокорейтинговых исследований по индексам JCR и SJR [12]. В поиске использовались булевы операторы со следующими терминами: "engagement detection" И "computer vision" И "online learning"; "facial expression recognition" И "education"; "gaze estimation" И "student engagement"; и "multimodal engagement analysis". Дополнительные исследования были идентифицированы путем отслеживания цитирований и ручного поиска в соответствующих материалах конференций, включая серию челленджей EmotiW [4, 13], которые ежегодно публикуют эталонные тесты с 2018 года.

Критерии включения включали: (1) эмпирические исследования, использующие методы компьютерного зрения для обнаружения или анализа вовлеченности студентов; (2) исследования, проведенные в контексте онлайн, дистанционного или технологически опосредованного обучения; (3) статьи, сообщающие количественные показатели производительности; (4) публикации на английском языке; и (5) работы, опубликованные в период 2020–2025 гг. Критерии исключения удаляли повторяющиеся публикации, главы книг, диссертации, тезисы конференций без полного текста и исследования, сосредоточенные исключительно на традиционных классных условиях без технологического посредничества [12]. В рассмотренной литературе вовлеченность последовательно концептуализируется как многомерный конструкт, хотя операционные определения различаются между исследованиями. Опираясь на основополагающую структуру, установленную Фредрикс и др. [1], многие исследователи различают поведенческую, когнитивную и эмоциональную вовлеченность [4, 7]. На рисунке 2 представлено распределение типов вовлеченности студентов в рамках рассмотренных исследований. Наибольшую долю занимает многомерный подход (44,4%), что свидетельствует о преобладании комплексного анализа вовлеченности, учитывающего одновременно несколько ее аспектов.

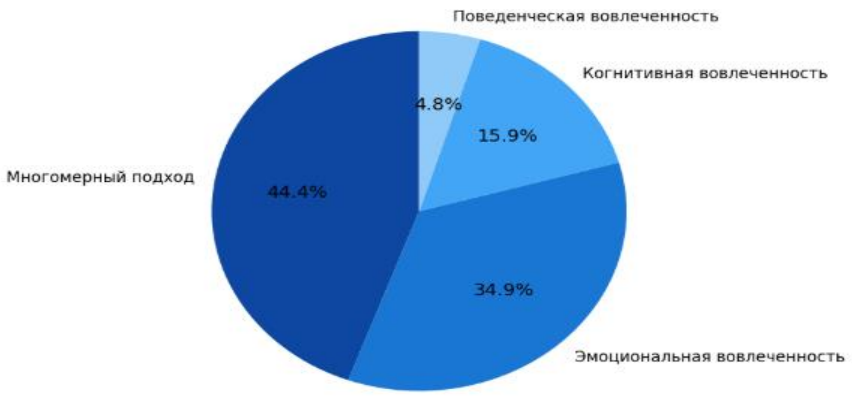

Рисунок 2. Основные типы вовлеченности

Многомерные подходы является наиболее изученным измерением (44.44%, 28 исследования), за ней следуют эмоцианальная вовлеченность (34.92%, 22 исследования), когнитивная вовлеченность (15.87%, 10 исследований) и поведенческая вовлеченность (4,76%, 3 исследования). Вовлеченность учащихся – комплексное явление, проявляющееся на нескольких уровнях. Это эмоциональный отклик от интереса до скуки, служащий индикатором восприятия контента. Это и внутренняя интеллектуальная работа, о которой судят по направлению взгляда и положению головы: сосредоточенный взгляд говорит о погружении, отведение взгляда – о потери концентрации. Внешне вовлеченность выражается в действиях (письмо, жесты), которые отслеживают современные системы анализа поз и движений. Наиболее объективную картину дают многомерные подходы, объединяющие мимику, взгляд и движения тела для целостного понимания включенности учащегося в процесс. Таблица 1 отражает распределение отобранных научных исследований по выявлению вовлеченности студентов с использованием методов компьютерного зрения в разрезе научных журналов. Представленные данные демонстрируют, в каких изданиях чаще всего публикуются работы по данной тематике, что позволяет выявить наиболее активные и авторитетные журналы в области компьютерного зрения и образовательных технологий. Анализ таблицы показывает концентрацию публикаций в ряде ведущих журналов, что свидетельствует о растущем интересе научного сообщества к вопросам автоматического определения вовлеченности студентов.

Таблица 1. Места публикации избранных исследований по выявлению вовлеченности студентов с помощью компьютерного зрения (Журнал)

Журнал	*Год*	*Ссылки*
IEEE Access	*2023-2025*	*[4], [10], [26], [27], [28], [29], [30], [31], [62]*
Multimedia Tools and Applications	*2021-2024*	*[8], [9], [14], [32], [33], [34]*
Electronics	*2020-2023*	*[3], [18], [38]*
Sustainability	*2020-2024*	*[39], [40], [61], [63]*
Computers and Electrical Engineering	*2021, 2023*	*[7], [11]*
International Journal of Cognitive Computing in Engineering	*2023, 2024*	*[36], [37]*
Applied Sciences	*2022*	*[41]*
Discover Computing (formerly Discover Systems)	*2024*	*[42]*
Discover Education	*2024*	*[43]*
Future Generation Computer Systems	*2020*	*[44]*
IEEE Transactions on Learning Technologies	*2024*	*[45]*
International Journal of Advanced Computer Science and Applications	*2023*	*[46]*
International Journal of Information and Education Technology	*2021*	*[47]*
International Journal of Web-Based Learning and Teaching Technologies	*2021*	*[48]*
Journal of Healthcare Informatics Research	*2021*	*[49]*
Optical Memory and Neural Networks	*2022*	*[50]*
SoftwareX	*2024*	*[51]*
User Modeling and User-Adapted Interaction	*2020*	*[52]*
Applied Mathematics and Nonlinear Sciences	*2024*	*[53], [54]*
International Journal of Information Technology and Computer Science	*2024*	*[55]*
Indonesian Journal of Electrical Engineering and Computer Science	*2024*	*[56]*
Journal of Robotics, Networking and Artificial Life	*2024*	*[57]*
Computer Modeling in Engineering and Sciences	*2023*	*[59]*
Springer (Journal)	*2025*	*[1]*

Таблица 2 демонстрирует распределение отобранных исследований по выявлению вовлеченности студентов с использованием методов компьютерного зрения именно в материалах научных конференций. В отличие от журнальных публикаций, представленные данные акцентируют внимание на конференциях как ключевых площадках для апробации и обсуждения новых научных результатов.

Таблица 2. Места публикации избранных исследований по выявлению вовлеченности студентов с помощью компьютерного зрения (Конференция)

Конференция	*Год*	*Ссылки*
Lecture Notes in Computer Science (LNCS)	*2020, 2024*	*[19], [58]*
IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)	*2022, 2024*	*[20], [22]*
IEEE Global Engineering Education Conference (EDUCON)	*2022, 2023*	*[23], [24]*
International Learning and Technology Conference (L&T)	*2025*	*[60]*
ACM International Conference on Multimodal Interaction (ICMI)	*2020*	*[13]*
Digital Image Computing: Techniques and Applications (DICTA)	*2020*	*[15]*
IEEE International Conference on Dependable, Autonomic and Secure Computing	*2021*	*[17]*
Springer (Engineering Applications of Neural Networks)	*2024*	*[21]*
IEEE International Conference on Big Data (BigData)	*2023*	*[2]*

Несколько общедоступных наборов данных сыграли ключевую роль в развитии исследований по обнаружению вовлеченности [4]. Набор FER-2013 содержит более 35 тысяч изображений лиц размером 48×48 пикселей, классифицированных по семи категориям эмоций, и широко используется для обучения моделей распознавания эмоций благодаря сбалансированному распределению на обучающую, валидационную и тестовую выборки [4]. Набор DAiSEE предлагает четырехуровневую аннотацию вовлеченности от полного отсутствия до высокого уровня наряду с метками эмоций, однако демонстрирует дисбаланс классов с преобладанием состояний высокой вовлеченности, что смещает модели в сторону классов большинства [4, 13]. Наборы данных челленджа EmotiW, особенно задачи прогнозирования вовлеченности 2018 и 2019 годов, стали стандартными эталонами: они включают почти 200 видео студентов, просматривающих контент MOOC в естественных условиях, с разметкой от пяти аннотаторов, достигших приемлемого уровня согласованности [4, 13, 15]. Вместе с тем ограниченная применимость общедоступных наборов данных к реальным учебным средам побудила многих исследователей разрабатывать специализированные наборы, предназначенные для захвата вовлеченности в аутентичных образовательных условиях [4].

Обнаружение вовлеченности на основе компьютерного зрения опирается на извлечение визуальных сигналов из видеопотоков, причем литература демонстрирует растущую сложность в инженерии признаков – от низкоуровневых пиксельных до высокоуровневых семантических индикаторов [4, 13, 15]. Наиболее широко используются мимические признаки: OpenFace извлекает единицы действия на основе системы кодирования лицевых движений, отслеживая до двадцати различных движений лицевых мышц с оценками интенсивности, а также предоставляет признаки позы головы и направления взгляда [4, 13, 15]. Оценка направления взгляда значительно продвинулась с помощью глубокого обучения, например, с использованием сверточной нейронной сети L2CS-Net на базе ResNet50 для прогнозирования углов поворота и наклона, где вовлеченность определяется нахождением взгляда в пределах установленных границ относительно экрана [7]. Отслеживание позы тела стало дополнительной модальностью, важной для обнаружения поведенческой вовлеченности, особенно при окклюзии лица, и OpenPose позволяет отслеживать движения головы, тела и рук из стандартного видео [13].

Ранние системы обнаружения вовлеченности использовали классические алгоритмы машинного обучения с ручными признаками, включая комбинирование мимических признаков с данными о движениях мыши, что повышало точность по сравнению с использованием только мимики [7], а также методы на основе кластеризации видеосегментов с последующим переназначением интенсивности вовлеченности [13]. Современные системы характеризуются доминированием архитектур глубокого обучения: сверточные нейронные сети применяются в различных конфигурациях – от последовательных архитектур с небольшими фильтрами и пулингом [11] до более продвинутых решений, таких как RepVGG, которая благодаря структурной репараметризации преобразует многопутевые сети во время обучения в однопутевые при выводе для повышения эффективности [7]. Рекуррентные архитектуры, особенно сети долгой краткосрочной памяти в сочетании со сверточными признаками, моделируют временные зависимости и развитие состояний вовлеченности на протяжении учебных сессий [4, 13, 15].

Для количественной оценки вовлеченности используются различные подходы: четырехуровневая дискретная шкала, различающая состояния от полного отсутствия до высокой вовлеченности [4, 7, 13, 15], и непрерывные рейтинги, обеспечивающие высокое временное разрешение через покадровые траектории [7]. Оценка систем выполняется с помощью нескольких метрик: точность классификации остается наиболее часто сообщаемой, хотя требует осторожной интерпретации из-за дисбаланса классов [1, 4, 13]; F1-мера важна для несбалансированных данных [4]; среднеквадратичная ошибка применяется для непрерывных измерений [4], а для порядковых шкал метрики корреляции, такие как корреляция Пирсона, дают лучшее представление о качестве предсказаний [13, 15].

Результаты исследования

Наборы данных FER-2013 и RAF-DB последовательно дают высокую производительность по моделям, что объясняется их структурированными аннотациями и соответствием задачам распознавания мимики, в которых CNN преуспевают. DAISEE демонстрирует более широкий диапазон производительности, отражая сложность анализа вовлеченности в видеосредах.

Важным результатом этого обзора является ограниченное внимание к безопасности и конфиденциальности в исследованиях по обнаружению вовлеченности. Унсити и др. [12] проанализировали 43 отобранные статьи и классифицировали их по уровню рассмотрения конфиденциальности:

- SP0: Безопасность или конфиденциальность вообще не упоминаются – 29 статей (67%)
- SP1: Безопасность и/или конфиденциальность упоминаются как аспекты, которые необходимо учитывать – 7 статей (16%)
- SP2: Описаны стратегии обеспечения безопасности или конфиденциальности – 3 статьи (7%)
- SP3: Безопасность и конфиденциальность всесторонне рассмотрены – 1 статья (2%)

Это пренебрежение вызывает беспокойство, поскольку системы обнаружения вовлеченности по своей природе обрабатывают личную информацию – изображения лиц, которые могут идентифицировать отдельных лиц и потенциально раскрывать конфиденциальные эмоциональные состояния [12].

В рассмотренной литературе выявлены устойчивые проблемы, ограничивающие применение систем обнаружения вовлеченности в реальном мире:

1. Изменчивость окружающей среды – условия освещения, качество камеры, позы головы и окклюзии снижают производительность моделей за пределами лабораторных условий [4, 7, 11, 12].
2. Дисбаланс классов – дисбаланс классов в наборах данных о вовлеченности смещает модели в сторону классов большинства (высокая вовлеченность), снижая чувствительность к состояниям невовлеченности, наиболее важным для образовательного вмешательства [1, 13].

3. Демографическая изменчивость – возраст, пол, культура, географическое положение влияют на паттерны мимики и поведение вовлеченности. Модели, обученные на однородных популяциях, могут не обобщаться на разнообразные группы учащихся [4, 11, 12].

4. Ограничения наборов данных – существующие общедоступные наборы данных демонстрируют дисбаланс классов, ограниченное демографическое разнообразие и узкие экологические контексты [4, 11, 13].

5. Долгосрочная динамика – большинство исследований используют поперечные дизайны или короткие сессии, давая ограниченное понимание того, как паттерны вовлеченности развиваются в течение курсов [4, 7, 11].

Этот обзор литературы синтезирует текущие исследования (2020–2025) по обнаружению вовлеченности на основе компьютерного зрения, раскрывая область на пересечении значительного технического прогресса и устойчивых концептуальных, методологических и этических проблем. Рассмотренные исследования демонстрируют, что системы КЗ могут эффективно извлекать значимые индикаторы вовлеченности из мимики, паттернов взгляда и движений тела, достигая точности, приближающейся или превосходящей человеческую производительность в контролируемых условиях [4, 7, 11, 13]. Архитектуры глубокого обучения, особенно CNN и их варианты, доказали свою способность изучать сложные представления признаков из необработанных видеоданных, а ансамблевые и мультимодальные подходы последовательно превосходят одномодальные системы [4, 13, 15].

Важным результатом этого обзора является ограниченное внимание к безопасности и конфиденциальности в исследованиях по обнаружению вовлеченности. Унсити и др. [12] проанализировали 43 отобранные статьи и классифицировали их по уровню рассмотрения конфиденциальности:

- SP0: Безопасность или конфиденциальность вообще не упоминаются – 29 статей (67%)
- SP1: Безопасность и/или конфиденциальность упоминаются как аспекты, которые необходимо учитывать – 7 статей (16%)
- SP2: Описаны стратегии безопасности или конфиденциальности – 3 статьи (7%)
- SP3: Безопасность и конфиденциальность всесторонне рассмотрены – 1 статья (2%)

Это пренебрежение вызывает беспокойство, поскольку системы обнаружения вовлеченности по своей природе обрабатывают личную информацию – изображения лиц, которые могут идентифицировать отдельных лиц и потенциально раскрывать конфиденциальные эмоциональные состояния [12].

В рассмотренной литературе выявлены устойчивые проблемы, ограничивающие применение систем обнаружения вовлеченности в реальном мире:

1. Изменчивость окружающей среды – условия освещения, качество камеры, позы головы и окклюзии снижают производительность моделей за пределами лабораторных условий [4, 7, 11, 12].

2. Дисбаланс классов – дисбаланс классов в наборах данных о вовлеченности смещает модели в сторону классов большинства (высокая вовлеченность), снижая чувствительность к состояниям невовлеченности, наиболее важным для образовательного вмешательства [1, 13].

3. Демографическая изменчивость – возраст, пол, культура, географическое положение влияют на паттерны мимики и поведение вовлеченности. Модели, обученные на однородных популяциях, могут не обобщаться на разнообразные группы учащихся [4, 11, 12].

4. Ограничения наборов данных – существующие общедоступные наборы данных демонстрируют дисбаланс классов, ограниченное демографическое разнообразие и узкие экологические контексты [4, 11, 13].

5. Долгосрочная динамика – большинство исследований используют поперечные дизайны или короткие сессии, давая ограниченное понимание того, как паттерны вовлеченности развиваются в течение курсов [4, 7, 11].

Этот обзор литературы синтезирует текущие исследования (2020-2025) по обнаружению вовлеченности на основе компьютерного зрения, раскрывая область на пересечении

значительного технического прогресса и устойчивых концептуальных, методологических и этических проблем. Рассмотренные исследования демонстрируют, что системы компьютерного зрения могут эффективно извлекать значимые индикаторы вовлеченности из мимики, паттернов взгляда и движений тела, достигая точности, приближающейся или превосходящей человеческую производительность в контролируемых условиях [4, 7, 11, 13]. Архитектуры глубокого обучения, особенно CNN и их варианты, доказали свою способность изучать сложные представления признаков из необработанных видеоданных, а ансамблевые и мультимодальные подходы последовательно превосходят одномодальные системы [4, 13, 15].

Дискуссия

Однако преобразование этих технических возможностей в практические образовательные инструменты остается неполным. Из этого синтеза вытекает несколько критических пробелов.

Во-первых, в области отсутствует стандартизированный подход к концептуализации и операционализации вовлеченности. Хотя большинство исследований ссылаются на трехкомпонентную модель (поведенческая, когнитивная, эмоциональная), конкретные индикаторы и их сопоставления значительно различаются, что затрудняет сравнение между исследованиями и мета-анализ [4, 7]. Области были бы полезны консенсусные руководящие принципы, определяющие минимальные стандарты отчетности для конструктов вовлеченности, наборов признаков и протоколов оценки.

Во-вторых, ограничения наборов данных сдерживают прогресс. Существующие общедоступные наборы данных, хотя и ценны, демонстрируют дисбаланс классов, ограниченное демографическое разнообразие и узкие экологические контексты [4, 11, 13]. Преобладание западных, образованных, молодых взрослых выборок поднимает вопросы о кросскультурной обобщаемости – паттерны мимики, нормы взгляда и поведение вовлеченности могут различаться в разных культурах способами, которые текущие наборы данных не могут уловить [4, 12]. Совместные усилия по созданию крупномасштабных, разнообразных, этично полученных наборов данных должны быть приоритетными.

В-третьих, объединение нескольких модальностей – лица, глаз, тела, голоса, физиологических сигналов – технически сложно, но концептуально необходимо [4, 7, 11, 12, 13]. Вовлеченность по своей природе мультимодальна; ограничение анализа одним каналом неизбежно упускает важные сигналы. Демонстрация Чжан и др. [7] того, что объединение экспрессии и данных мыши улучшает точность, и включение Чанг и др. [13] отслеживания тела указывают на мультимодальную фузию как на перспективное направление.

В-четвертых, необходимо больше внимания уделять временной динамике. Вовлеченность колеблется в течение секунд, минут и сессий; статические классификации теряют информацию о траекториях, паттернах и переходах [4, 7]. Посредственная производительность подходов на основе LSTM в прогнозировании вовлеченности указывает на неадекватность текущего временного моделирования [4, 13, 15].

В-пятых, разрыв в конфиденциальности и этике, задокументированный Унсити и др. [12], требует срочного исправления. Полное отсутствие рассмотрения вопросов конфиденциальности в опубликованных исследованиях неприемлемо для технологий, предназначенных для использования с уязвимыми группами населения. Будущие работы должны включать принципы конфиденциальности по дизайну: минимизацию данных, ограничение цели, прозрачность, безопасность и контроль со стороны пользователя.

В-шестых, педагогическая интеграция остается теоретически недостаточно изученной. Системы обнаружения идентифицируют состояния вовлеченности, но не предписывают соответствующие ответы. Исследования должны изучить, как преподаватели могут эффективно использовать информацию о вовлеченности [4, 12, 14].

Существует мало долгосрочных исследований, изучающих реальное применение. Большинство исследований оценивают модели в контролируемых или полуконтролируемых условиях в течение коротких периодов [4, 7, 11].

Заключение

В ходе настоящего исследования была подтверждена технологическая зрелость методов компьютерного зрения для задач обнаружения вовлеченности, а также продемонстрирован их высокий потенциал для совершенствования систем онлайн-обучения. За период 2020–2025 годов архитектуры глубокого обучения, эталонные наборы данных и подходы к мультимодальной интеграции достигли уровня развития, позволяющего говорить о готовности технологий к практическому внедрению. Разработанные системы способны надежно извлекать и интерпретировать визуальные сигналы, обеспечивая точность прогнозов, достаточную для решения прикладных образовательных задач.

Полученные результаты подтверждают, что эмоциональная вовлеченность остается доминирующим направлением анализа, при этом наблюдается устойчивый рост интереса к многомерным моделям, учитывающим сложную природу этого феномена. Анализ мимики, дополненный мультимодальными подходами, формирует основу современных решений, наиболее адаптированных к условиям цифровой образовательной среды.

Реализация всего потенциала технологий требует последовательного решения ряда концептуальных и прикладных задач, включающих обеспечение разнообразия данных, совершенствование методов временного моделирования, защиту конфиденциальности и педагогическую валидацию решений. В условиях экспансии онлайн-образования разработка точных, этичных и ориентированных на человека систем становится не просто техническим вызовом, а образовательным приоритетом. Дальнейшее развитие видится на пересечении компьютерного зрения, наук об обучении и человеко-ориентированного проектирования, что позволит создавать инструменты, реально поддерживающие учащихся и преподавателей в динамике современных образовательных процессов.

References

[1] Tingting Han, Ruqian Liu, Shuwei Dou, Wei Wang, Xiaoming Ding, Wenxia Zhang, Jihao Lang, Wenxuan Li, Jixing Han. "Balancing act: engagement detection in online learning through master-assistant models with an enhanced hierarchical attention mechanism". Published online: 30 September 2025 © The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2025

[2] D. Chiaro, D. Annuziata, S. Izzo, "Unveiling engagement in virtual classrooms: A multimodal analysis," in Proc. IEEE International Conference on Big Data (BigData), Dec. 2023, pp. 4761-4769.

[3] Y. Qi, L. Zhuang, H. Chen, X. Han, and A. Liang, "Evaluation of students' learning engagement in online classes based on multimodal vision perspective," Electronics, vol. 13, no. 1, p. 149, Dec. 2023.

[4] I. Qarbal, N. Sael, and S. Ouahabi, "Student's engagement detection based on computer vision: A systematic literature review," IEEE Access, vol. 13, pp. 140519-140543, 2025.

[5] L. Mishra, T. Gupta, and A. Shree, "Online teaching-learning in higher education during lockdown period of COVID-19 pandemic," International Journal of Educational Research Open, vol. 1, Jan. 2020, Art. no. 100012.

[6] A. Abedi and S. S. Khan, "Engagement measurement based on facial landmarks and spatial-temporal graph convolutional networks," 2024, arXiv:2403.17175.

[7] P. Bhardwaj, P. K. Gupta, H. Panwar, M. K. Siddiqui, R. Morales-Menendez, and A. Bhai, "Application of deep learning on student engagement in e-learning environments," Computers and Electrical Engineering, vol. 93, Jul. 2021, Art. no. 107277.

[8] S. Gupta, P. Kumar, and R. K. Tekchandani, "Facial emotion recognition based real-time learner engagement detection system in online learning context using deep learning models," Multimedia Tools and Applications, vol. 82, no. 8, pp. 11365-11394, Mar. 2023.

[9] S. Gupta, P. Kumar, and R. Tekchandani, "A multimodal facial cues based engagement detection system in e-learning context using deep learning approach," Multimedia Tools and Applications, vol. 82, no. 18, pp. 28589-28615, Jul. 2023.

[10] K. Watanabe, T. Sathyanarayana, A. Dengel, and S. Ishimaru, "EnGauge: Engagement gauge of meeting participants estimated by facial expression and deep neural network," IEEE Access, vol. 11, pp. 52886-52898, 2023.

[11] A. Harb, A. Gad, M. Yaghi, M. Alhalabi, H. Zia, J. Yousaf, A. Khelifi, K. Ghoudi, and M. Ghazal, "Diverse distant-students deep emotion recognition and visualization," Computers and Electrical Engineering, vol. 111, Nov. 2023, Art. no. 108963.

[12] O. Unciti, A. Martínez-Ballesté, and R. Palau, Real-time emotion recognition and its impact on educational environment, in Proc. International Conference on Computer Supported Education, 2023, pp. 1- 12.

[13] C. Chang, C. Zhang, L. Chen, and Y. Liu, "An ensemble model using face and body tracking for engagement detection," in Proc. 20th ACM International Conference on Multimodal Interaction, Oct. 2020, pp. 616-622.

[14] P. Buono, B. De Carolis, F. D'Errico, N. Macchiarulo, and G. Palestra, "Assessing student engagement from facial behavior in on-line learning," Multimedia Tools and Applications, vol. 82, no. 9, pp. 12859-12877, Apr. 2023.

[15] A. Kaur, A. Mustafa, L. Mehta, and A. Dhall, "Prediction and localization of student engagement in the wild," in Proc. Digital Image Computing: Techniques and Applications (DICTA), Dec. 2020, pp. 1-8.

[16] X. Ai, V. S. Sheng, C. Li, and Z. Cui, "Class-attention video transformer for engagement intensity prediction," 2022, arXiv:2208.07216.

[17] D. Boulanger, M. A. A. Dewan, V. S. Kumar, and F. Lin, "Lightweight and interpretable detection of affective engagement for online learners," in Proc. IEEE International Conference on Dependable, Autonomic and Secure Computing, Oct. 2021, pp. 176-184.

[18] M. U. Ucar and E. Özdemir, "Recognizing students and detecting student engagement with real-time image processing," Electronics, vol. 11, no. 9, p. 1500, May 2022.

[19] Z. A. T. Ahmed, M. E. Jadhav, A. M. Al-madani, M. Tawfik, S. N. Alsubari, and A. A. A. Shareef, "Real-time detection of student engagement: Deep learning-based system," in Proc. International Conference on Innovative Computing and Communications, Springer, 2021, pp. 313-323.

[20] A. M. Mathew, A. A. Khan, T. Khalid, and R. Souissi, "GESCAM: A dataset and method on gaze estimation for classroom attention measurement," in Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Jun. 2024, pp. 636-645.