

Қ.Т. Қожахмет¹, А.М. Өмірәлі^{1*}, Д.А. Баязитов¹

¹ НАО «Университет Нархоз», г.Алматы, Казахстан
*e-mail: aikumis.omirali@gmail.com

ИННОВАЦИОННЫЕ МЕТОДЫ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА ДЛЯ ВЫЯВЛЕНИЯ ИСТИННЫХ И ЛОЖНЫХ ВЫСКАЗЫВАНИЙ

Аннотация

В статье рассматриваются новейшие методы обнаружения ложной информации и проверки подлинности высказываний на основе искусственного интеллекта. В качестве перспективных решений по обнаружению фальсификаций обсуждается использование нейронных сетей для классификации текста, методов компьютерного зрения и обработки изображений и видео, а также алгоритмов обработки естественного языка. В статье подчеркивается точность и эффективность этих методов при выявлении истинных и ложных утверждений. В том числе, отмечаются проблемы с обеспечением точности моделей искусственного интеллекта и необходимость опережать развивающиеся технологии, используемые для создания фальсификаций. В заключении авторы пришли к выводу о том, что имеется острая необходимость в развитии методов обнаружения фальсификаций, основанных на искусственном интеллекте, что в дальнейшем даст возможность своевременного выявления дезинформации и определения методов их устранения.

Ключевые слова: искусственный интеллект, обнаружение обмана, нейронные сети, методы компьютерного зрения, методы классификации текста, алгоритмы обработки естественного языка.

Аңдатпа

Қ.Т. Қожахмет¹, А.М. Өмірәлі¹, Д.А. Баязитов¹
¹«Нархоз Университеті» КеАҚ, Алматы қ., Қазақстан

ШЫНАЙЫ ЖӘНЕ ЖАЛҒАН МӘЛІМДЕМЕЛЕРДІ АНЫҚТАУДЫҢ ИННОВАЦИЯЛЫҚ ЖАСАНДЫ ИНТЕЛЛЕКТ ӘДІСТЕРІ

Бұл мақалада жалған ақпаратты анықтаудың және жасанды интеллектке негізделген мәлімдемелердің түпнұсқалығын тексерудің соңғы әдістері қарастырылған. Жалған ақпаратты анықтаудың перспективалық шешімдері ретінде мәтінді жіктеу үшін нейрондық желілерді, компьютерлік көру әдістерін және кескіндер мен бейнелерді өңдеу әдістерін, сондай-ақ табиғи тілді өңдеу алгоритмдерін пайдалану талқыланады. Мақалада шынайы және жалған мәлімдемелерді анықтауда осы әдістердің дәлдігі мен тиімділігі көрсетілген. Оның ішінде жасанды интеллект модельдерінің дәлдігін қамтамасыз етудегі проблемалар және бұрмалануларды жасау үшін қолданылатын дамып келе жатқан технологиялардан озып кету қажеттілігі атап өтіледі. Қорытындылай келе, авторлар жасанды интеллектке негізделген жалғандықтарды анықтау әдістерін дамытудың шұғыл қажеттілігі бар деген қорытындыға келді, бұл кейіннен жалған ақпаратты уақтылы анықтауға және оларды жою әдістерін анықтауға мүмкіндік береді.

Түйін сөздер: жасанды интеллект, жалған ақпаратты анықтау, нейрондық желілер, компьютерлік көру әдістері, мәтінді жіктеу әдістері, табиғи тілді өңдеу алгоритмдері.

Abstract

INNOVATIVE ARTIFICIAL INTELLIGENCE METHODS TO IDENTIFY TRUE AND FALSE STATEMENTS

Kozhakhmet K.T.¹, Omirali A.M.¹, Bayazitov D.A.¹
¹NJSC «Narxoz University», Almaty, Kazakhstan

This article discusses the latest methods of detecting false information and verifying the authenticity of statements based on artificial intelligence. The use of neural networks for text classification, computer vision and image and video processing methods, as well as natural language processing algorithms are discussed as promising solutions for deception detection. The article emphasizes the accuracy and effectiveness of these methods in identifying true and false statements. In particular, problems with ensuring the accuracy of artificial intelligence models and the need to stay ahead of developing technologies used to create falsifications are noted. In conclusion, the authors came to the idea that there is an urgent need for the development of methods for deception detection based on artificial intelligence, which further makes it possible to detect misinformation in a timely manner and determine methods for their elimination.

Keywords: artificial intelligence, deception detection, neural networks, computer vision methods, methods for text classification, natural language processing algorithms.

Введение

В эпоху цифровых технологий информация играет важнейшую роль в принятии решений в различных областях, таких как политика, бизнес и образование. Однако с возросшей легкостью создания и распространения информации, проблема обнаружения ложной или вводящей в заблуждение информации стала серьезной проблемой. Ложная информация может принимать различные формы, включая фальсифицированные изображения, видео-манипуляции и ложные заявления.

Для борьбы с этой проблемой требуются инновационные подходы к выявлению истинных и (или) ложных утверждений. Растущее распространение информации, как истинной, так и ложной, привело к разработке инновационных подходов и баз данных для выявления подлинной или ложной информации. Использование алгоритмов искусственного интеллекта (ИИ), таких как машинное обучение и обработка естественного языка, открывает потенциал для повышения точности и эффективности проверки информации.

Один из подходов к проблеме проверки утверждений основан на использовании нейронных сетей, в частности рекуррентных нейронных сетей (Recurrent neural networks (далее – RNNs)). RNNs являются мощным инструментом для анализа последовательностей данных и позволяют выявлять ложные утверждения. Исследования показали, что RNNs эффективны в обнаружении слухов в микроблогах [1]. Этот подход позволяет автоматически анализировать текстовые данные и идентифицировать вероятность их достоверности.

Другой подход заключается в использовании методов, основанных на знаниях. Эти методы используют большие базы знаний, такие как Всемирная паутина, для поддержки процесса проверки [2]. Базы знаний содержат информацию, которая может быть использована для сопоставления и проверки фактов и утверждений.

В дополнение к алгоритмическим подходам, наличие больших и разнообразных источников данных имеет решающее значение для обучения и оценки систем обнаружения лжи. Эти источники данных могут включать новостные статьи, сообщения в социальных сетях и другие источники информации. Разработка специализированных наборов данных для обнаружения поддельных новостей [3] стала важным шагом вперед в этом направлении. Чем больше разнообразных данных доступно для обучения системы, тем точнее и эффективнее она может стать в выявлении лжи.

Однако, несмотря на потенциальные преимущества вышеуказанных подходов, существуют ограничения и предубеждения, которые необходимо устранить. Например, на классификацию достоверности автоматизированных систем проверки фактов могут влиять политические или идеологические предубеждения источников, используемых в процессе машинного обучения [4]. Это может привести к искажению результатов и неправильной классификации информации.

В целом, разработка инновационных подходов и использование ИИ в обнаружении ложной информации и проверке подлинности является важной областью исследований. Непрерывное развитие и совершенствование этих методов позволит более эффективно бороться с проблемой фальсификации и дезинформации в цифровой эпохе.

Методологические подходы

Рассматриваются следующие методологические подходы по обнаружению истинной и (или) ложной информации на основе ИИ: методы классификации текстов, методы экспертизы изображений, аудио и видео, методы обработки естественного языка, мультимодальный подход обнаружения истинных и (или) ложных высказываний.

Методы классификации текстов (Methods for Text Classification)

Методы классификации текстов являются важной частью современной обработки естественного языка и имеют огромное значение в обнаружении ложной информации.

Нейронные сети для классификации текста. В данном контексте нейронные сети представляют собой мощные алгоритмы машинного обучения, которые могут обучаться на больших объемах текстовых данных. Они способны анализировать текст и классифицировать его на категории, такие как "истинная" и "ложная" информация. Эти модели требуют больших объемов размеченных данных, где каждый текст снабжен меткой, указывающей на его правдивость.

Применение нейронных сетей. Нейронные сети могут быть обучены на разнообразных типах текстов. Например, их можно обучить на наборе данных, содержащем как правдивые, так и ложные новостные статьи. После обучения такой модели она способна делать прогнозы относительно

подлинности новой, неизвестной статьи. Это означает, что эти нейронные сети могут быть использованы для автоматического обнаружения фейковых новостей.

Сверточные нейронные сети (CNNs). CNNs - это один из видов нейронных сетей, которые успешно применяются для обработки текста. Они обычно используются для анализа коротких текстов, таких как твиты. Их особенностью является способность выявления паттернов и взаимосвязей между словами в тексте, что делает их эффективными в выявлении ложной информации в кратких текстах.

Рекуррентные нейронные сети (RNN). RNNs - это другой вид нейронных сетей, предназначенных для обработки последовательных данных, таких как текст. Они лучше справляются с более длинными текстами, такими как новостные статьи, благодаря способности "понимать" контекст и зависимости между словами и предложениями.

Обучение на размеченных данных. Для успешной классификации текстов на истинные и ложные, эти нейронные сети требуют больших объемов размеченных данных [3]. Это означает, что их необходимо обучать на больших наборах текстов, где каждый текст помечен соответствующей меткой.

Высокая точность. Одной из ключевых характеристик нейронных сетей для классификации текстов является высокая точность. Эти модели, когда правильно обучены, могут давать высокие результаты в обнаружении ложной информации.

Использование нейронных сетей, таких как CNNs и RNNs, для классификации текстов с целью обнаружения ложной информации, является мощным и эффективным методом. Эти методы могут быть применены к различным типам текстов и способны достигать высокой точности в выявлении ложной информации, что является критически важным в контексте борьбы с дезинформацией.

Методы экспертизы изображений и видео (Methods for Image and Video Forensics)

Методы экспертизы изображений и видео сегодня являются важной составной частью современной цифровой эпохи, где фальсификации и манипуляции с изображениями и видео распространяются с невиданным ранее масштабом. Искусственный интеллект предоставляет мощные инструменты для выявления таких манипуляций с высокой точностью и эффективностью.

Обнаружение глубоких подделок. Одним из выдающихся примеров в данной области является использование алгоритмов глубокого обучения, включая глубокие нейронные сети, для выявления так называемых глубоких подделок. Глубокие подделки - это видеозаписи, в которых лицо или голос человека заменены или искажены с помощью передовых методов искусственного интеллекта. Обучая модели на наборе данных, содержащем как оригинальные, так и поддельные видеоролики, можно добиться высокой точности в обнаружении таких манипуляций.

Техники обнаружения фальсификаций. Для обнаружения фальсифицированных изображений и видео применяются разнообразные техники и алгоритмы компьютерного зрения. Один из подходов основан на анализе текстуры и структуры изображений, что позволяет выявлять аномальные паттерны, свидетельствующие о фальсификации. Другие методы используют анализ стеганографических или цифровых следов, оставленных при редактировании изображений или видео.

Важным аспектом эффективности методов обнаружения фальсификаций является обучение на больших и разнообразных наборах данных. Эти наборы данных должны включать как оригинальные, так и обработанные изображения и видео, чтобы модели могли научиться распознавать характеристики фальсификации и различать их от подлинного контента.

Перспективы развития области. Область обнаружения фальсификаций в мире изображений и видео находится в постоянном состоянии эволюции и предоставляет несметные возможности для дальнейшего развития исследований и инноваций. В этом разделе мы рассмотрим более широкий спектр перспективных направлений, которые ожидают нас в ближайшие годы. Одним из наиболее интересных и активно исследуемых направлений является использование генеративных моделей. Генеративные алгоритмы, такие как генеративные состязательные сети (GANs), предоставляют новые возможности для создания и детектирования фальсификаций. С их помощью можно создавать реалистичные фейковые изображения и видео, а также разрабатывать более продвинутые методы обнаружения на основе анализа отклонений от ожидаемых генеративных моделей. В развитие области вносит вклад маркеры редактирования, которые являются своего рода цифровыми следами, оставленными при манипуляциях с изображениями и видео. Ученые активно работают над созданием инструментов и алгоритмов, которые могли бы точно выявлять такие маркеры и помогать в форензическом анализе данных, что имеет большое значение для правопорядка и судебных исследований. Один из наиболее перспективных трендов - это анализ контекста и семантики визуальных данных. Модели, способные понимать смысл и связи между объектами на изображениях

и видео, становятся более точными в выявлении фальсификаций. Это позволяет более тонко различать подлинные события и объекты от фейковых. С развитием аппаратного обеспечения и оптимизации алгоритмов, методы обнаружения фальсификаций становятся быстрее и более доступными. Это позволяет их применение в реальном времени для защиты информации и контроля за достоверностью мультимедийных данных, что актуально в сферах безопасности, медицины и журналистики. Новые исследования идут в направлении объединения различных модальностей данных, таких как изображения и аудио. Это позволяет создавать более надежные системы обнаружения, способные выявлять фальсификации, учитывая несколько аспектов информации одновременно. Важным аспектом развития области является сотрудничество между исследователями, инженерами и законодателями. Вместе с техническими аспектами также нужно обсуждать этические и правовые вопросы, связанные с использованием методов обнаружения фальсификаций.

Методы экспертизы изображений и видео на основе искусственного интеллекта представляют собой важное направление исследований и разработки в борьбе с фальсификациями и манипуляциями в мультимедийных данных. Они обеспечивают высокую точность и эффективность при выявлении разнообразных форм манипуляций, что является критически важным в современном информационном мире.

Методы обработки естественного языка (Methods for Natural Language Processing)

Алгоритмы обработки естественного языка (Natural Language Processing, NLP) являются мощным инструментом для обнаружения ложных утверждений в письменном или устном тексте. NLP-алгоритмы могут использоваться для различных целей, таких как выявление поддельных новостных статей, выявление плагиата или выявление кампаний по дезинформации на платформах социальных сетей.

Одним из примеров применения NLP для обнаружения ложных утверждений является использование алгоритмов классификации текста. Эти алгоритмы обучаются на больших наборах данных, содержащих помеченные тексты, и могут классифицировать новые тексты как истинные или ложные. Например, модель NLP может быть обучена на новостных статьях и использоваться для определения подлинности новых статей, выявляя ложные утверждения и фальсификации.

Другой подход, используемый в NLP для обнаружения ложных утверждений, - это анализ контекста и семантики текста. Алгоритмы могут исследовать связи между словами и фразами в тексте, а также использовать знания о мире, чтобы определить, является ли утверждение правдивым или ложным. Например, алгоритмы могут проверять факты и статистические данные, сопоставлять информацию с надежными источниками или определять несоответствия и противоречия в тексте.

При обучении алгоритмов NLP на данных помеченных текстовых наборов, важно иметь доступ к широкому спектру текстов, включая различные тематики и стили. Это помогает моделям обучаться на разнообразных примерах и адаптироваться к различным контекстам. Кроме того, обучение на актуальных и обновляемых данных позволяет алгоритмам быть в курсе новых форм ложной информации и дезинформации.

Визуализация результатов обработки текста также может быть полезной для анализа и интерпретации результатов. Например, можно создать облако тегов или график наиболее часто встречающихся слов или фраз в текстах, помеченных как ложные или истинные. Это позволит исследователям видеть ключевые термины, связанные с ложной информацией, и поможет лучше понять особенности ложных утверждений.

Семантический анализ и связанные события. Алгоритмы NLP могут анализировать не только отдельные утверждения, но и их связь с другими событиями. Например, они могут определять, имеются ли подтверждающие факты или события, которые противоречат данному утверждению, что помогает выявить ложные утверждения.

Анализ тональности. Анализ тональности текста является важной составной частью NLP. Этот аспект позволяет определять эмоциональную окраску текста и выявлять, был ли текст создан с целью манипуляции эмоциями аудитории.

Социальные сети и онлайн-платформы. NLP-алгоритмы могут активно применяться для обнаружения дезинформации и фейковых новостей на платформах социальных сетей. Они могут анализировать комментарии, посты и новости, выявляя нежелательное поведение, распространение ложной информации и создание ботов.

Спам и фишинг. NLP-алгоритмы также могут помочь в выявлении спама и фишинга в электронной почте и сообщениях. Они анализируют текст сообщений и идентифицируют признаки, характерные для нежелательных сообщений.

Обнаружение сентиментальных атак. Это подразумевает создание текстов, в которых сентимент или намерения автора искажены, чтобы ввести в заблуждение аудиторию. NLP-алгоритмы способны выявлять такие атаки путем анализа лингвистических и семантических аномалий.

Обработка мультимодальных данных. NLP не ограничивается только текстом. Современные методы NLP способны обрабатывать комбинации текстовых, аудио- и видеоданных, что позволяет более полно и точно выявлять манипуляции в мультимедийных контентях.

Социальная значимость. Важно также отметить социальную значимость борьбы с ложной информацией. Эффективное обнаружение ложных утверждений помогает снижать воздействие дезинформации на общество и обеспечивать информационную безопасность.

Обучение на больших данных и обновление моделей: Для обеспечения высокой точности и актуальности алгоритмов NLP, они должны быть обучены на больших, разнообразных и актуальных наборах данных. Регулярное обновление моделей позволяет адаптироваться к новым формам и методам распространения ложной информации.

Использование алгоритмов обработки естественного языка на основе искусственного интеллекта предоставляет мощный инструмент для борьбы с ложной информацией. Продолжающиеся исследования и разработки в области NLP позволяют улучшать эффективность и точность обнаружения ложных утверждений и помогают обеспечивать более достоверную информацию в различных областях.

Мультимодальный подход обнаружения истинных и (или) ложных высказываний

Мультимодальный подход объединяет несколько модальностей данных, таких как текст, изображения, аудио и видео, для обнаружения истинных и (или) ложных высказываний. Использование нескольких модальностей позволяет получить более полное представление информации и повысить точность обнаружения ложной информации.

Примером мультимодального подхода является использование комбинации текстового содержания и соответствующих изображений или видео. Например, можно анализировать текстовое содержание новостной статьи и сопоставлять его с изображениями, представленными в статье. Если содержание текста и изображения не согласуются, это может указывать на возможную ложную информацию.

Для реализации мультимодального подхода можно использовать глубокое обучение и нейронные сети. Например, можно использовать сверточные нейронные сети для обработки изображений и видео, а рекуррентные нейронные сети или алгоритмы обработки естественного языка для анализа текстового содержания. Затем результаты каждой модальности могут быть объединены и проанализированы для принятия окончательного решения о подлинности высказывания.

Применение мультимодального подхода требует доступа к разнообразным и размеченным наборам данных, содержащим информацию из различных модальностей. Эти данные могут быть получены из разных источников, таких как новостные сайты, социальные сети, видеоплатформы и другие. Обучение моделей на таких данных поможет иметь более всесторонний подход к обнаружению ложной информации.

Мультимодальный подход может быть особенно полезным в случаях, когда одна модальность данных может быть подвержена фальсификации или манипуляции, но другие модальности могут содержать дополнительные подтверждающие или противоречащие информации. В результате это позволяет создавать более надежные системы обнаружения ложной информации.

В исследовании [5] используется мультимодальный подход к обучению. В качестве набора данных используется [6]. Авторы воспользовались алгоритмом глубокой сверточной нейронной сети CNN для извлечения характеристик и для дальнейшего обучения. Визуальные характеристики извлекаются из видео с помощью 3D CNN [7]. Наилучшие результаты были получены с использованием 10-слойной архитектуры. Также, для извлечения характеристик из текста была использована модель CNN. Каждое высказывание представляется в виде вектора объединения составляющих слов. Звуковые характеристики, такие как высота тона и интенсивность голоса, извлекаются с помощью программного обеспечения OpenSmile [8]. Для мультимодального слияния было использовано 2 метода: раннее слияние и позднее слияние. При слиянии на раннем этапе характеристики извлекаются из входных данных. Затем характеристики объединяются и передаются в классификатор. Этот метод использует корреляцию между несколькими характеристиками на ранней стадии, что часто приводит к лучшему

выполнению задачи. В позднем слиянии одномодальные классификаторы используются для определения прогноза по каждому типу данных. Затем локальные прогнозы объединяются в единый вектор, который далее классифицируется для получения окончательного решения. Результаты исследования одномодального и мультимодального классификатора показаны в таблицах 1,2.

Таблица 1. Точность одномодального классификатора

Модальность	Точность (Accuracy)
Видео (B)	78.57%
Аудио (A)	87.5%
Текст (T)	83.78%

Таблица 2. Точность мультимодального классификатора

Модальность	Точность (Accuracy)	
	Позднее слияние	Раннее слияние
A + B	85%	89.1%
B + T	87%	91.8%
A + T	86%	91.9%
A + B + T	92%	96.4%

Результаты и перспективы развития

Обнаружение фальсификаций на основе искусственного интеллекта не лишено своих проблем. Одной из основных трудностей является обеспечение точности моделей искусственного интеллекта. Важно использовать большие и разнообразные наборы данных для обучения моделей, чтобы они могли точно идентифицировать фальсификации в широком диапазоне ситуаций. Кроме того, по мере развития технологий методы, используемые для фальсификации информации, также будут развиваться, и важно постоянно обновлять модели, чтобы оставаться на шаг впереди. Базы данных обнаружения обмана собираются с использованием различных модальностей, таких как, визуальные, акустические, физиологические, данные с электроэнцефалограммы (далее – ЭЭГ) и глазной взгляд.

Существующие работы по обнаружению обмана широко используют наборы данных, как показано в таблице 3.

Таблица 3. Наборы данных для обнаружения обмана

№	Набор данных	Стратегия сбора данных	Модальности	Результаты
1	База данных "Box of Lies" [9]	Реалистичный сценарий	Только видео	69 %
2	Мультимодальный набор данных для обнаружения обмана [6]	Гипотетический сценарий	Визуальные, акустические, физиологические и тепловые	-
3	"Bag of lies": мультимодальный набор данных для обнаружения обмана [10]	Реалистичный сценарий	Видео, аудио, ЭЭГ и глазной взгляд	66,17 %
4	Обнаружение обмана с высокой долей вероятности [11]	Ситуации с высокими ставками	Только видео	76,92 %
5	База данных обнаружения обмана Университета Майами [12]	Реалистичный сценарий	Только видео	52 %
6	База данных "Silesian" [13]	Гипотетический сценарий	Только видео	-
7	Реальная база данных обнаружения обмана [14]	Реалистичный сценарий	Только видео	87,7 %
8	Набор данных "UR Lying" [15]	Сценарий игры с лабораторным управлением	Только видео	75 %

В таблице 3 показаны наборы данных в виде аудио и видео материалов, собранные в разных сценариях и различных ситуациях. В целом приведены 8 наборов модальных данных, включающие аудио, видео, ЭЭГ, термо модальности. Также показаны стратегии сбора данных и результаты, которые исследователи получили с этими наборами данных. Многие из наборов данных как бы сосредоточены на одной методологии, в следствии чего уменьшается точность обучения модели, что влияет на принятие решений и ограничивает возможности искусственного интеллекта [10]. В таблице также даны результаты исследований в виде точности обучения модели, в том числе стратегии сбора данных. В частности, анализ видео показал, что записи с более высокой долей вероятности обнаружения истинности высказываний, как правило, оценивались как более честные. С другой стороны, записи с более умеренными тревожными ситуациями были оценены как менее честные [12]. Исходя из этого, можно сделать вывод, что эффективно проводить только сбор медиаданных, так как в режиме реального времени будет сложно проводить сбор данных о терме, ЭЭГ и т.д. Однако отмечается, что существует очень мало медиаданных для обучения модели ИИ.

Следует отметить в целом, что проведенные исследования представленных наборов данных рекомендуют следующее: общие характеристики видео (например, длительность) и оценки качества видео в целом (например, четкость изображения) должны быть согласованы и взаимосвязаны для правильного восприятия с целью избежания субъективных оценок. В частности, исследователями подчеркивается, что записи с более высокой вероятностью лжи (“High-Stakes Deception Detection Based on Facial Expressions”), как правило, оценивались как более правдивые, тогда как записи с низкой вероятностью лжи были оценены как менее правдивые. Исследования целевой аудитории (“База данных обнаружения обмана Университета Майами”) также подтвердили этот вывод. Более того, для повышения точности результатов необходимо включить больше характеристик в анализ аудио-видео материала [12], что является дальнейшей перспективной задачей для авторов статьи.

Заключение

Исследование и применение искусственного интеллекта в области проверки утверждений представляют собой важную и быстро развивающуюся область, которая имеет потенциал существенно повлиять на борьбу с дезинформацией и фальсификациями. Путем развития методов обнаружения фальсификаций, использующих нейронные сети, компьютерное зрение, обработку изображений и обработку естественного языка, можно достичь более эффективного выявления и предотвращения распространения ложной информации.

Однако необходимо признать, что точность и эффективность таких моделей требуют дальнейших исследований. С учетом развития технологий создания фальсифицированной информации, важно продолжать разрабатывать и совершенствовать методы обнаружения. Это включает работу по устранению ограничений и предвзятости существующих подходов, а также разработку новых алгоритмов и источников данных для определения истинных и ложных высказываний.

В целом, применение искусственного интеллекта в области проверки утверждений предоставляет возможность более эффективно бороться с дезинформацией и фальсификациями, но требует постоянного развития и усовершенствования. Только через продолжение исследований и совместных усилий мы сможем добиться более надежных и точных инструментов для выявления и борьбы с фальсификациями и поддержания информационного пространства.

Кроме того, важно обратить внимание на разработку и оценку новых алгоритмов и источников данных для определения истинных и ложных высказываний. Необходимо сосредоточиться на создании надежных и разнообразных наборов данных, которые охватывают широкий спектр тематик и типов фальсификаций. Это позволит улучшить обучение моделей и улучшить их обобщающую способность.

В этой связи необходимо продолжить работу по устранению предвзятости и ограничений, существующих в текущих моделях и алгоритмах. Находить баланс между точностью и объемом ложноположительных и ложноотрицательных срабатываний является сложной задачей, и требуется активное исследование для достижения оптимальных результатов.

Следует подчеркнуть, что совместные усилия и взаимодействие между исследовательскими группами, обществом и государственными организациями являются важными для прогресса в этой области. Постоянный обмен знаниями, распространение лучших практик и координация усилий помогут ускорить развитие методов обнаружения фальсификаций и повысить их эффективность.

Таким образом, использование искусственного интеллекта в области проверки утверждений имеет огромный потенциал для борьбы с дезинформацией и фальсификациями. Несмотря на существующие вызовы, разработка и совершенствование этих методов должны быть приоритетными задачами, чтобы обеспечить достоверность информации в нашем информационном обществе.

Исследование выполнено в рамках грантового финансирования МНУВО по проекту AP13068084 «Разработка технологии детектирования аномального (обманчивого) поведения респондента с использованием алгоритмов искусственного интеллекта (ИИ) на основе изменения характеристик голоса и речи» (конкурс молодых ученых по научным и (или) научно-техническим проектам на 2022-2024 годы).

Список использованной литературы:

- 1 Jing Ma, Wei Gao, Prasenjit Mitra, Sejeong Kwon, Bernard J. Jansen, Kam-Fai Wong, and Meeyoung Cha. 2016. Detecting rumors from microblogs with recurrent neural networks. In *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence (IJCAI'16)*. AAAI Press, 3818–3824.
- 2 Xin Luna Dong, Evgeniy Gabrilovich, Kevin Murphy, Van Dang, Wilko Horn, Camillo Lugaresi, Shaohua Sun, and Wei Zhang. 2015. Knowledge-based trust: estimating the trustworthiness of web sources. *Proc. VLDB Endow.* 8, 9 (May 2015), 938–949. <https://doi.org/10.14778/2777598.2777603>.
- 3 William Yang Wang. 2017. "Liar, Liar Pants on Fire": A New Benchmark Dataset for Fake News Detection. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 422–426, Vancouver, Canada. Association for Computational Linguistics.
- 4 Zhijiang Guo, Michael Schlichtkrull, Andreas Vlachos; A Survey on Automated Fact-Checking. *Transactions of the Association for Computational Linguistics* 2022; 10 178–206. doi: https://doi.org/10.1162/tacl_a_00454.
- 5 Mandar Gogate, Ahsan Adeel, Amir Hussain. "Deep learning driven multimodal fusion for automated deception detection". *IEEE Symposium Series on Computational Intelligence (SSCI)*. 2017.
- 6 Verónica Pérez-Rosas, Rada Mihalcea, Alexis Narvaez, and Mihai Burzo. 2014. A Multimodal Dataset for Deception Detection. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, pages 3118–3122, Reykjavik, Iceland. European Language Resources Association (ELRA).
- 7 S. Ji, W. Xu, M. Yang, and K. Yu. "3d convolutional neural networks for human action recognition". *IEEE transactions on pattern analysis and machine intelligence*. 2013. vol. 35. no. 1. pp. 221–231.
- 8 F. Eyben, F. Weninger, F. Gross, and B. Schuller. "Recent developments in opensmile, the munich open-source multimedia feature extractor". *Proceedings of the 21st ACM international conference on Multimedia*. ACM. 2013. pp. 835–838.
- 9 Felix Soldner, Verónica Pérez-Rosas, and Rada Mihalcea. 2019. Box of Lies: Multimodal Deception Detection in Dialogues. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 1768–1777, Minneapolis, Minnesota. Association for Computational Linguistics.
- 10 Viresh Gupta, Mohit Agarwal, Manik Arora, Tanmoy Chakraborty, Richa Singh and Mayank Vatsa, "Bag-of-Lies: A Multimodal Dataset for Deception Detection," 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2019, pp. 83-90, <https://doi.org/10.1109/CVPRW.2019.00016>.
- 11 Lin Su and Martin D. Levine, "High-Stakes Deception Detection Based on Facial Expressions," 2014 22nd International Conference on Pattern Recognition, 2014, pp. 2519-2524, <https://doi.org/10.1109/ICPR.2014.435>.
- 12 Lloyd E. Paige, Jason C. Deska, Kurt Hugenberg, et al. Miami University deception detection database. *Behav Res* 51, 429–439 (2019). <https://doi.org/10.3758/s13428-018-1061-4>.
- 13 Radlak, Krystian & Božek, Maciej & Smolka, Bogdan. (2015). Silesian Deception Database: Presentation and Analysis. 29-35. <https://doi.org/10.1145/2823465.2823469>.
- 14 Pérez-Rosas, Verónica & Abouelenien, Mohamed & Mihalcea, Rada & Xiao, Yao & Linton, CJ & Burzo, Mihai. (2015). Verbal and Nonverbal Clues for Real-life Deception Detection. 2336-2346. <https://doi.org/10.18653/v1/D15-1281>.
- 15 Leena Mathur and Maja J. Matarić, "Unsupervised Audio-Visual Subspace Alignment for High-Stakes Deception Detection," ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2021, pp. 2255-2259, <https://doi.org/10.1109/ICASSP39728.2021.9413550>.