

ПУБЛИЧНО-ПРАВОВЫЕ (ГОСУДАРСТВЕННО-ПРАВОВЫЕ) НАУКИ PUBLIC LAW (STATE LAW) SCIENCES

Научная статья
УДК 343.98
DOI 10.33184/pravgos-2025.3.7

Original article

ХОЛОПОВА Елена Николаевна ✉
Балтийский федеральный университет
имени И. Канта,
Калининград, Россия,
elchol@mail.ru

KHOLOPOVA Elena Nikolaevna
Immanuel Kant Baltic Federal University,
Kaliningrad, Russia

МУЛДАШЕВ Руслан Нуржанович
Балтийский федеральный университет
имени И. Канта,
Калининград, Россия,
ruslan15497@mail.ru

MULDASHEV Ruslan Nurzhanovich
Immanuel Kant Baltic Federal University,
Kaliningrad, Russia

ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ КАК СРЕДСТВО ПРОТИВОДЕЙСТВИЯ ФЕЙКАМ ПРИ ОБЕСПЕЧЕНИИ НАЦИОНАЛЬНОЙ БЕЗОПАСНОСТИ

ARTIFICIAL INTELLIGENCE AS A MEANS TO COUNTER FAKES IN NATIONAL SECURITY

Аннотация. В условиях глобальной информационной войны против России распространение фейковой информации стало одним из ключевых инструментов манипуляции общественным сознанием. Фейки угрожают национальной безопасности, подрывая доверие к государственным институтам и провоцируя социальные волнения. В связи с этим актуальным становится вопрос применения искусственного интеллекта для эффективного противодействия фейкам. Цель: обосновать значимость внедрения технологий искусственного интеллекта для борьбы с фейковой информацией, предложить алгоритм противодействия фейкам в сфере национальной безопасности России. Методы: эмпирические методы сравнения (для выявления лжи через сопоставление с фактами), описания (для классификации фейков по видам и признакам), интерпретации данных о распространении фейковой информации (с целью раскрытия их цели и воздействия на общество) использовались комплексно, что позволило не только фиксировать фейки, но и разработать меры противодействия им; теоретические методы формальной и диалектической логики использовались при анализе понятийного аппарата; частнонаучный метод юридико-догматического анализа нормативных правовых актов позволил обнаружить пробелы в регулировании, оценить эффективность законов, предложить

Abstract. In the context of the global information war against Russia, the dissemination of fake information has become one of the key tools for manipulating public consciousness. Fakes threaten national security by undermining trust in government institutions and provoking social unrest. In this regard, the issue of using artificial intelligence (AI) to effectively counter fakes becomes relevant. Purpose: to substantiate the importance of introducing AI technologies to combat fake information, as well as to propose an algorithm for countering fakes in the sphere of Russia's national security. Methods: empirical methods of comparison (in order to identify lies through comparing with facts), description (in order to classify fakes by types and characteristics), interpretation of data on the dissemination of fake information (in order to disclose their purpose and impact on society) are used comprehensively, which makes it possible not only to fix fakes, but also to develop countermeasures against them; theoretical methods of formal and dialectical logic are used in the analysis of the conceptual apparatus; the specific scientific method of legal and dogmatic analysis of regulatory legal acts allows to identify gaps in regulation, evaluate the effectiveness of laws, and propose legal solutions to the identified problems. Results: the study proves the need to

правовые решения выявленных проблем. Результаты: доказана необходимость разработки и внедрения отечественных технологий искусственного интеллекта для противодействия фейкам, направленным на подрыв безопасности государства, искажение фактов, подделку мультимедийного контента и создание недостоверных источников с целью дестабилизации общества, снижения доверия к государственным институтам и провокации различных конфликтов. Технологии искусственного интеллекта обнаруживают фейковые новости с точностью, превышающей возможности человека, и могут быстро реагировать на появление новых видов фейков. Применение искусственного интеллекта для распознавания фейков и контроля за распространением информации имеет ключевые преимущества: упреждающий подход (борьба не с последствиями, а с угрозами), адаптивность (обучение искусственного интеллекта на новых видах фейков), комплексность (сочетание технических и экспертных решений).

Ключевые слова: национальная безопасность, информационная безопасность, искусственный интеллект, фейки, фейкинг, дезинформация, машинное обучение, анализ контента, технологии, классификация фейков

Для цитирования: Холопова Е.Н. Искусственный интеллект как средство противодействия фейкам при обеспечении национальной безопасности / Е.Н. Холопова, Р.Н. Мулдашев. – DOI 10.33184/pravgos-2025.3.7 // Правовое государство: теория и практика. – 2025. – № 3. – С. 60–70.

develop and implement national AI technologies to counter fakes aimed at undermining state security, distorting facts, falsifying multimedia content and creating unreliable sources to destabilize society, reduce trust in government institutions and provoke various conflicts. AI technologies detect fake news with an accuracy exceeding human capabilities and can quickly respond to the emergence of new types of fakes. The use of AI to recognize fakes and control the dissemination of information has key advantages: a proactive approach (fighting threats rather than consequences); adaptability (training AI on new types of fakes); comprehensiveness (a combination of technical and expert solutions).

Keywords: national security, information security, artificial intelligence, fakes, faking, disinformation, machine learning, content analysis, technology, classification of fakes

For citation: Kholopova E.N., Muldashev R.N. Artificial Intelligence as a Means to Counter Fakes in National Security. *Pravovoe gosudarstvo: teoriya i praktika = The Rule-of-Law State: Theory and Practice*, 2025, no. 3, pp. 60–70. (In Russian). DOI 10.33184/pravgos-2025.3.7.

ВВЕДЕНИЕ

В последние десятилетия в отношении Российской Федерации все чаще проводятся информационно-психологические операции, информационные провокации с целью манипулирования общественным сознанием и формирования негативного отношения к управленческим решениям органов государственной власти.

Необходимость создания безопасного информационного пространства, защиты российского общества от деструктивного информационно-психологического воздействия и укрепления традиционных российских духовно-нравственных ценностей¹ подтверждается интенсивностью ведения информационной войны странами блока НАТО, попытками обвинения Вооруженных Сил РФ в действиях,

противоречащих международному гуманитарному праву (участие в специальной военной операции на Украине (далее – СВО)); разработкой и использованием систем искусственного интеллекта (далее – ИИ), создающих риск нарушения целостности информации и доступа к ней²; принятием странами блока НАТО целевых программ, направленных на нанесение ущерба России и обеспечение безопасности стран альянса³, в том числе с использованием технологий ИИ для ведения информационной войны.

² ГА ООН приняла первую резолюцию о регулировании искусственного интеллекта [Электронный ресурс]. URL: <https://tass.ru/ekonomika/20309919/amp> (дата обращения: 12.03.2025).

³ Коданева С.И. Научные и технологические тренды 2020–2040 гг.: перспективы научно-технического развития / С.И. Коданева, С.М. Пястолова [Электронный ресурс]. URL: <https://cyberleninka.ru/article/n/nauchnye-i-tehnologicheskie-trendy-2020-2040-gg-perspektivy-nauchno-tehnicheskogo-razvitiya-referat-doklada-organizatsii-nato-po> (дата обращения: 29.03.2025).

¹ Стратегия национальной безопасности Российской Федерации : Указ Президента РФ от 02.07.2021 № 400 // Собрание законодательства РФ. 2021. № 27 (ч. 2), ст. 5351.

ФЕЙК И ДЕЗИНФОРМАЦИЯ: ПОНЯТИЕ И КЛЮЧЕВЫЕ РАЗЛИЧИЯ

Одним из инструментов ведения глобальной информационной войны, развязанной в отношении Российской Федерации, являются фейки [1, с. 11]. Фейки, как цифровые инструменты медийного пространства, используются с целью обмана, введения в заблуждение или нанесения вреда отдельным категориям граждан, целым государствам.

В научной литературе фейк рассматривается как поддельная новость, которая не имеет фактической основы (представляется как факт) [2, с. 96]; ложная информация, созданная для привлечения внимания, при условии, что правда либо имеется частично, либо полностью отсутствует; сенсация и подозрительный факт, сложные для проверки и публикуемые сомнительным источником [3, с. 180]; явление медиасреды, включающее в себя поддельные текст, фото-, видео- или аудиозаписи, а также инструменты их распространения [4, с. 96]. Таким образом, общий признак фейка – намеренно созданная недостоверная информация, которая преподносится как достоверная.

Представители зарубежной науки, как правило, связывают фейк с дезинформацией [5, с. 104]. Концептуальное отличие этих понятий заключается в следующем: фейк – в основном полностью ложная информация, в отличие от дезинформации, где смешаны правда и ложь [3, с. 126]; фейк рассчитан на мгновенный эффект и имеет сенсационный и шокирующий посыл, что может быть нехарактерно для дезинформации [6, с. 117]; фейку присущ информационный «вброс» с быстрым жизненным циклом и «вирусным» распространением, а дезинформация имеет длительный эффект воздействия [7, с. 40]; фейк – целенаправленно созданный инструмент реализации политических, военных и экономических целей, а дезинформация может быть следствием непреднамеренной ошибки субъектов информационного взаимодействия [8, с. 319].

Необходимо заметить, что законодательного определения фейка нет. Вместе с тем в ст. 15.3 Федерального закона от 27 июля 2006 г. № 149-ФЗ «Об информации, информационных технологиях и о защите информации» содержится понятие «общественно значимая ложная информация» – распростра-

няемая под видом достоверных сообщений информация, создающая угрозу причинения вреда жизни и (или) здоровью граждан, имуществу, угрозу массового нарушения общественного порядка и (или) общественной безопасности⁴. Один из признаков указанного понятия такой же, как и у фейка: недостоверность информации, представляемой под видом достоверной.

Результаты исследования фейковой информации позволили определить понятие «фейк» как особый вид ложной (выдаваемой за достоверную) или искаженной информации с сенсационным содержанием, рассчитанной на быстрое распространение, с целью манипуляции общественным мнением, создания ажиотажа или дестабилизации ситуации. Фейк, как инструмент информационной войны, направлен на достижение политических, экономических или военных целей, создает угрозу безопасности личности, общества и государства.

Следует обратить внимание, что отсутствие единой классификации ложной информации объясняется различиями в подходах к проверке ее правдивости, а также разнообразием содержательных характеристик самой информации, которую необходимо проверить. В связи с этим возникают сложности при классификации фейковой информации. Так, например, отдельные исследователи выделяют три типа фейков: сфабрикованные новости (распространяемые в официальных или альтернативных СМИ), мистические новости, сатирические материалы [9, с. 429]. Существуют и иные классификации фейковой информации [4, с. 89–94].

Вместе с тем, как представляется, имеется необходимость дифференцировать фейки, созданные с целью подрыва безопасности, ориентированные на дестабилизацию и манипуляцию общественным сознанием, по функциям и видам. Такие фейки классифицируются на устрашающие (вызывающие панику и страх) – ложные сообщения о катастрофах, угрозах безопасности, эпидемиях, военных действиях, вызывающие тревогу и дезорганизующие общество); политические – направленные на дискредитацию политических оппонентов, манипуляцию общественным мнением и вмешательство во внутренние дела госу-

⁴ Собрание законодательства РФ. 2019. № 12, ст. 1221.

дарств с целью подрыва безопасности; репутационные или имиджевые – недостоверные сведения, подрывающие доверие к основным институтам государственной власти (правительству, силовым структурам, банкам), что может привести к экономическим и социальным потрясениям; «цифровые призраки» или дипфейки – сверхреалистичные подделки, созданные ИИ, которые могут использоваться для имитации заявлений или действий официальных лиц государства, такие фейки могут развлекать или угрожать обществу; «фальшресурсы» или «цифровые хамелеоны» – имитация официальных ресурсов для распространения ложной информации и провокаций; фальсификация метаданных – сообщения о событиях в неверном времени или месте с целью дезориентировать аудиторию и вызвать эмоциональную реакцию.

Создание и распространение перечисленных видов фейков наиболее опасно для национальной и информационной безопасности, так как они воздействуют на массовое сознание, подрывают доверие к государственным институтам власти и могут провоцировать конфликты и нестабильность в обществе.

ФЕЙКИ КАК УГРОЗА БЕЗОПАСНОСТИ ДЛЯ ОБЩЕСТВА И ГОСУДАРСТВА

Для обозначения различных моделей распространения фейков можно использовать следующую терминологию: контролируемое распространение фейков и неконтролируемое распространение фейков, которое различается по степени управления и последствиям. Контролируемое распространение фейков является умышленным, организованным как государственными, так и негосударственными структурами с определенной целью (например, государственная пропаганда во время политических и иных конфликтов, фейковые новости перед выборами). Неконтролируемое распространение фейков – спонтанное, неуправляемое распространение ложной информации, которое, как правило, происходит из-за психологических и социальных факторов – стресса, страха, желания поддержать социальные связи, недоверия к официальным источникам (например, слухи о катастрофах или терактах). Таким образом, контролируемое распространение фейков связано с со-

знательным действием и может иметь юридические последствия, а неконтролируемое – с массовым, стихийным распространением, вызванным психологическими и социальными причинами. И то и другое является опасным, противодействие этим видам распространения фейков имеет свою специфику.

В отечественных научных исследованиях отмечается, что неконтролируемое распространение фейковой информации создает непосредственную угрозу «ментальной безопасности» личности и всего социума в информационном пространстве [10, с. 75]. Так, О.А. Еремина пишет, что «сегодня любую катастрофу или террористический акт сопровождает "вброс" фейковых новостей, провоцирующих нагнетание страха и панику среди населения, распространение слухов и домыслов, создающих опасность политической дестабилизации общества» [11, с. 242].

Следует отметить, что увеличение интенсивности и силы информационного воздействия на население Российской Федерации является признаком информационной войны и относится к 2020–2023 гг. (устойчивая негативная тенденция). Распространение фейковой информации в Российской Федерации и анализ ее содержания свидетельствуют, что основные цели этой деятельности – подрыв доверия к власти, провоцирование общественных волнений и недовольства, введение граждан в заблуждение по поводу проводимой в стране политики.

Так, например, к фейкам относится подача и распространение недостоверной информации о пожаре в торгово-развлекательном центре «Зимняя вишня» (25 марта 2018 г.). Способ распространения негативно эмоционально окрашенной информации предполагал моментальную реакцию после ее получения и заключался в следующем: от лица представителя МЧС были направлены уведомления в морги Кемерово с указанием о подготовке мест для 300 погибших в результате пожара, что в пять раз превышало фактическое число жертв; от имени «Позднякова из Администрации Президента» в государственные структуры поступали запросы с требованием убрать цветы и игрушки, оставленные жителями города у здания торгового центра в знак скорби, с целью предотвращения формирования представления о более масштабной катастрофе.

Необходимо отметить, что к признакам, позволяющим классифицировать информацию как фейк (а не дезинформацию), относятся: подача информации лицом, имеющим большую аудиторию подписчиков (публичность); преувеличение количества погибших, имевшее мгновенный эффект воздействия на общество; содержание информации о сокрытии количества погибших органами государственной власти, носившее сенсационный характер, вызвавшее мгновенный шок; кратковременность воздействия информации, представлявшая информационный вброс, который легко мог быть опровергнут органами государственными власти, а значит, не имел возможности длительно воздействовать на общество. Распространение фейковой информации спровоцировало митинги, жители Кемерово выступали с требованиями справедливости и отставки чиновников⁵.

Проведение СВО и формирование негативного восприятия российских военнослужащих, включая дискредитацию использования Вооруженных Сил РФ в целях защиты национальных интересов, обусловили увеличение количества фейковой информации в сети Интернет. Данная информация содержала обвинения в адрес военнослужащих в совершении военных преступлений – убийство гражданского населения, применение пыток к военнопленным украинской армии и акты мародерства. В качестве примеров несанкционированных акций можно привести митинги, состоявшиеся в Москве 24 февраля 2022 г., в ходе которых выдвигались лозунги, выражающие протест против проведения СВО на территории Украины. Широкий резонанс в период СВО вызвали фейковые новости об «убийстве мирного населения» в Буче и об «атаке российской армией» Запорожской АЭС [12, с. 379].

Институтом изучения войны (СИИ) в 2023 г. был опубликован фейк о планируемой новой волне мобилизации в России (несмотря на заявления Президента РФ В.В. Путина, что данное мероприятие не планируется)⁶.

⁵ На митинге в Кемерово потребовали отставки властей региона [Электронный ресурс]. URL: <https://ria.ru/amp/20180327/1517378095.html> (дата обращения: 10.03.2025).

⁶ В Сети обсуждают доклад США о второй волне мобилизации в РФ в феврале – марте 2023 года [Электронный ресурс]. URL: <https://spbdnevnik.ru/news/2023-01-12/v-seti-obsuzhdayut-doklad-ssha-o-vtoroy-volne-mobilizatsii-v-rf-v-martefevrale-2023-goda> (дата обращения: 24.02.2025).

Практически все фейки были направлены не только на провокацию общественных беспорядков в стране, но и на обвинение органов государственной власти, на несоответствие заявленных целей СВО реально складывающейся обстановке.

Е. Мизулина⁷ отмечала: «по состоянию на 1 марта 2022 года около 1 310 000 публикаций, содержащих заведомо ложные сведения (фейки) относительно ситуации на Украине и в Донбассе, распространено в средствах массовой информации, в том числе сети Интернет»⁸.

Увеличение количества распространяемых фейков способствовало не только проведению несанкционированных мероприятий против СВО, но и внесению изменений в нормативные правовые акты (введение административной и уголовной ответственности в связи с повышенной степенью общественной опасности фейков). Федеральным законом от 1 апреля 2020 г. № 100-ФЗ «О внесении изменений в Уголовный кодекс Российской Федерации и статьи 31 и 151 Уголовно-процессуального кодекса Российской Федерации» введена уголовная ответственность за публичное распространение под видом достоверных сообщений заведомо ложной информации об обстоятельствах, представляющих угрозу жизни и безопасности граждан, и (или) о принимаемых мерах по обеспечению безопасности населения и территорий, приемах и способах защиты от указанных обстоятельств (ст. 207.1 УК РФ), за публичное распространение заведомо ложной общественно значимой информации в период пандемии (ст. 207.2 УК РФ).

Борьба с ложной информацией на законодательном уровне началась в 2019 г. после внесения поправок в Федеральный закон от 27 июля 2006 г. № 149-ФЗ «Об информации, информационных технологиях и о защите информации» и принятия Федерального закона от 18 марта 2019 г. № 27-ФЗ «О внесении изменений в Кодекс Российской Федерации об административных правонарушениях», кото-

⁷ Екатерина Мизулина – член Комиссии Общественной палаты РФ по развитию информационного сообщества, СМИ и массовых коммуникаций [Электронный ресурс] // Общественная палата РФ : сайт. URL: <https://www.oprf.ru> (дата обращения: 10.03.2025).

⁸ В ОП назвали число фейков о спецоперации на Украине [Электронный ресурс]. URL: <https://iz.ru/1298741/2022-03-01/v-op-nazvali-chislo-feikov-ospetcooperacii-na-ukraine> (дата обращения: 10.02.2025).

рыми ст. 13.15 КоАП была дополнена ч. 9–11, установившими, что распространение информации, создавшее угрозу причинения вреда жизни и здоровью граждан, имуществу, угрозу массовых беспорядков, нарушения функционирования объектов жизнеобеспечения, влечет наложение административного штрафа. Распространение фейков физическими или юридическими лицами наказывается штрафом до 1,5 млн рублей.

Необходимо подчеркнуть, что фейковая информация все чаще распространяется недружественными государствами с целью подрыва доверия к решениям органов российской власти, дестабилизации обстановки в стране, формирования деструктивных и панических настроений среди населения, целенаправленного введения в заблуждение значительной части граждан, что представляет реальную угрозу национальной безопасности Российской Федерации.

Необходимо заметить, что отсутствие доверия к основным средствам массовой информации, наличие их выбора, доступность социальных сетей в качестве источников новостей являются основными факторами, влияющими на широкое распространение фейковой информации [13, с. 241]. Интернет-платформы, включая веб-сайты, форумы, социальные сети, мессенджеры, служат информационными ресурсами, которые активно используются для стимулирования деструктивных процессов в обществе.

ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ КАК СРЕДСТВО ПРОТИВОДЕЙСТВИЯ ФЕЙКАМ

Практика борьбы с фейками в Российской Федерации основывается на самостоятельном анализе человеком публикуемых материалов на предмет наличия ложной информации. Оказать помощь в распознавании фейков может мобильное приложение «Хорошие новости», разработанное в рамках спецпроекта «Информационный щит» как онлайн-инструмент вовлечения граждан в противодействие фейкам. За две недели функционирования приложения с помощью граждан России удалось выявить более 400 ссылок на публикации с фейковой информацией⁹.

⁹ В России появился новый инструмент борьбы с фейками [Электронный ресурс]. URL: <https://ruposters.ru/news/16-05-2024/rossii-poyavilsya-novii-instrument-borbi-feikami> (дата обращения: 24.02.2025).

Вместе с тем в условиях стремительного роста объемов публикуемых фейков, скорости их распространения и сложности анализа с учетом человеческого фактора необходимы средства борьбы, которые с применением специальных лингвистических знаний могут распознать фейк как речевое действие (заведомо ложное сообщение или достоверное утверждение о фактах или событиях, оформленное с помощью языковых средств) [14, с. 287], определить достоверность сгенерированной информации, в том числе с использованием технологии ИИ самим ИИ [14, с. 219].

Как отмечает Т.А. Полякова, «неоспорима притягательность и интерес во всем мире к ИИ в свете его новых возможностей, революционных решений в самых разных областях, создания преобразующего потенциала эффективности и уникальных инноваций. Однако полагаем, что сегодня особого внимания и научного осмысления заслуживают прогнозы рисков и новых угроз для национальной безопасности, включая практически все ее виды (информационную, экономическую, транспортную, экологическую и др.)» [15, с. 86–87].

Следует заметить, что ИИ используется для создания почти 10 % всех фейков в сети Интернет (в 2023 г., по сравнению с 2022 г., количество распространяемого контента выросло в 17 раз)¹⁰. Например, компания OpenAI (одна из мировых лидеров в области разработки технологий ИИ) представила отчет о нейросетевой технологии GPT-4 и ее способности генерировать убедительную ложную информацию, запускать фишинговую атаку, скрывать признаки неправомерного поведения¹¹.

Технологии ИИ способны генерировать убедительную ложную информацию, публикации в социальных платформах и другой текстовый контент [16, с. 212], создавать дипфейки – аудиовизуальный контент, используемый для манипуляции обществом, формируемый с помощью генерации ложного звукового ряда, а также фото- и видеоизображений, представляющих «реальную»

¹⁰ Искусственно интеллектуальная собственность [Электронный ресурс]. URL: <https://www.kommersant.ru/doc/6806550> (дата обращения: 26.02.2025).

¹¹ Чат-бот GPT-4 оказался убедительным дезинформатором [Электронный ресурс]. URL: <https://news.rambler.ru/internet/50399750-chat-bot-gpt-4-okazalsya-ubeditelnym-ezinformatorem/> (дата обращения: 10.03.2025).

информацию. Настоящие факты и люди изменяются под сложные для идентификации фейки. Дипфейки практически невозможно распознать, поскольку видео отличается высокой степенью реалистичности [17, с. 67].

Таким образом, современные интернет-ресурсы способствуют широкому распространению фейков в обществе, а использование ИИ для создания ложной информации усложняет и без того непростой процесс их обнаружения, что свидетельствует о необходимости применения современных технологических решений для совершенствования механизма борьбы с распространением фейков.

Противодействие фейкам включает комплекс мер и методов, направленных на выявление, опровержение и предотвращение распространения ложной информации. Основным механизмом противодействия фейкам состоит из следующих элементов: субъекты борьбы (государственные органы, СМИ, общественные организации и граждане), методы (фактчекинг (проверка фактов), медиаграмотность (обучение критическому восприятию информации), создание антифейков (контраргументов и разоблачений, использование мемов и коротких видео для широкой аудитории, иные технологии), правовые меры (установление ответственности за дезинформацию, ограничение доступа к ресурсам, распространяющим вредоносный контент), профилактика (предупреждение информационных атак, создание и продвижение официальной информации для управления ситуацией и нейтрализации фейков).

В современных условиях актуально рассмотрение преимуществ автоматического выявления фейковой информации и эффективного внедрения в медиасферу технологий на основе ИИ. Роскомнадзор считает, что внедрение ИИ для борьбы с фейками произойдет в течение 3–5 лет¹².

Разработка и быстрое внедрение отечественных технологий ИИ, позволяющих распознавать фейки и контролировать распространение информации с учетом правовых норм, этических принципов и допустимости

ограничения свободы слова, является необходимым, так как использование зарубежных технологий ИИ, обученных на иностранных источниках, создает угрозу отнесения к фейкам достоверной информации. Так, например, китайская нейросеть Deep Seek считает фейком заявления Министерства обороны РФ об обнаружении на освобожденных территориях Украины биологических лабораторий США, разрабатывавших биологическое оружие, выделяет цель данной публикации – «оправдание СВО и провокация антиамериканских настроений», несмотря на доказательства, опубликованные начальником войск радиационной, химической и биологической защиты Вооруженных Сил РФ генерал-лейтенантом И. Кирилловым¹³. При этом нейронная сеть ссылается на американские информационные сайты Reuters и BBC News. Следует также заметить, что если ИИ обучают на данных определенных государств, то он может стать инструментом политической цензуры и предвзятости (например, в Китае ИИ блокирует любые упоминания Тибета или Тайваня как незаконные).

ИИ становится ключевым средством противодействия фейкам благодаря своим возможностям автоматизации и анализа больших объемов данных. Основные методы использования ИИ в борьбе с дезинформацией при обеспечении национальной безопасности включают: мультимодальный контент-текстовый анализ (охватывает текст, изображение, видео, аудио с использованием определенной методологии); машинное обучение (нейросети обучаются на больших массивах данных с примерами фейков, чтобы автоматически распознавать типичные паттерны дезинформации и быстро выявлять новые фейки); выявление аномалий (акцент на выявление нестандартных паттернов); распознавание ботов и автоматизированных аккаунтов (многомерный анализ поведения, контента и технических данных, сочетающий автоматизацию и экспертизу); кросс-проверка источников (системный процесс установления достоверности информации путем

12 Роскомнадзор использует искусственный интеллект в борьбе с интернет-фейками [Электронный ресурс]. URL: <https://rg.ru/2023/04/24/roskomnadzor-ispolzuet-iskusstvennyj-intellekt-v-borbe-s-internet-fejkami.html> (дата обращения: 02.03.2025).

13 Брифинг начальника войск РХБ защиты ВС РФ по результатам анализа документов, касающихся военно-биологической деятельности США на Украине [Электронный ресурс] // Министерство обороны РФ : сайт. URL: <https://structure.mil.ru/structure/forces/ground/news/more.htm?id=12526771@egNews> (дата обращения: 02.03.2025).

критического анализа содержания, проверки источников, сопоставления с подтвержденными данными); прогнозирование и оперативное реагирование (превентивно-реактивная система мониторинга анализа и нейтрализации дезинформационных угроз).

Так, например, технологии автоматического обнаружения фейков в тексте основываются на его содержании и структуре исходя из результатов семантического анализа (выявление противоречий, неестественных формулировок или стилистических особенностей) [18, с. 1569]. Необходимо отметить, что в настоящее время отечественные технологии ИИ, способные осуществлять в автоматическом режиме проверку опубликованной текстовой информации (фактчекинг), находятся на стадии разработки. Однако примеры функционирования данных технологий есть в зарубежных странах. Например, проект Factiveverse использует инструмент на основе ИИ, способный распознавать недостоверную информацию в публикуемом материале и представлять проверенные источники информации, базу данных FactiSearch для проверки фактов и контекстно-аналитический инструмент Microfacts, предназначенный для интерпретации профессиональной терминологии и жаргонизмов¹⁴.

Технологии ИИ, способные выявить погрешности в подлинном изображении и обнаружить различные типы манипуляций, определяют достоверность и целостность визуальных или аудиальных данных медиаконтента [19, с. 237]. Примеры из международной и российской практики демонстрируют накопление опыта в выявлении подобных манипуляций, при этом ключевым фактором эффективности применяемых методов является предварительное изучение признаков обработки и модификации исходных данных. Например, российская компания MTS AI разработала решение на основе ИИ для борьбы с дипфейками. Нейросеть может выявлять несоответствия, которые обычно не видны человеку – неправильные тени или неестественные движения лицевых мышц, особенности тембра голоса или шумы. Нейросеть также понимает, как один фрагмент видео- или аудио-

записи связан с другими, и учитывает интонацию голоса, паузы или длительность звуков¹⁵. Таким образом, использование имеющихся отечественных разработок для анализа публикуемых мультимедиа в основных источниках распространения дипфейков способствует снижению степени негативного влияния сгенерированного аудиовизуального контента.

Высокая, благодаря ботам, скорость распространения фейкового контента в социальных платформах [20, с. 3] свидетельствует о необходимости автоматизированного мониторинга с использованием технологий ИИ активности пользователей социальных платформ и публикуемого ими контента. Например, сотрудники Санкт-Петербургского федерального исследовательского центра РАН обучили ИИ выявлять в социальных платформах ботов, распространяющих фейки. В основе метода лежит анализ публичных данных о ботах, независимо от языка, на котором публикуются посты и комментарии¹⁶. Таким образом, мониторинг социальных платформ с использованием технологий ИИ способствует эффективному обнаружению не только фейкового контента, но и ботов, осуществляющих его массовую рассылку в отечественных социальных платформах («ВКонтакте», «Одноклассники», Rutube и др.).

Одним из элементов метода «прогнозирования и оперативного реагирования» борьбы с фейками является их опровержение. Для противодействия распространению недостоверной информации в России запустили проект «Война с фейками», цель которого – разоблачение публикуемой недостоверной новости. Так, Telegram-канал «Война с фейками» за сентябрь 2022 г. – февраль 2023 г. опубликовал опровержение 573 фейков, касающихся СВО. Однако в условиях, когда растет их количество, возникает необходимость использования технологий ИИ не только для их обнаружения, но и генерации и публикации опровержения.

15 ИИ-сервис защитит от дипфейков и запрещенного контента [Электронный ресурс]. URL: <https://mobile-review.com/all/news/ii-servis-zashhitit-ot-dipfejkov-i-zapreshhyonnogo-kontenta> (дата обращения: 07.03.2025).

16 Российские ученые разработали метод выявления ботов в соцсетях [Электронный ресурс]. URL: <https://rscf.ru/news/presidential-program/rossiyskie-uchenye-razrabotali-metod-vyyavleniya-botov-v-sotssetyakh/> (дата обращения: 05.03.2025).

14 Стартап Factiveverse разработал модель ИИ, которая способна определять фейковую информацию [Электронный ресурс]. URL: <https://truesharing.ru/news/49500/> (дата обращения: 02.03.2025).

ЗАКЛЮЧЕНИЕ

Противодействие фейкам – комплексная, многоуровневая деятельность, сочетающая технические, образовательные, правовые и информационные меры с целью защиты общества от дезинформации и манипуляций. ИИ выступает как многофункциональное средство – от выявления и анализа фейков до поддержки принятия решений и информирования пользователей, что значительно повышает эффективность борьбы с дезинформацией в современном цифровом пространстве. Алгоритм противодействия фейкам в сфере национальной безопасности состоит из следующих этапов: круглосуточный мониторинг угроз (использование технологий ИИ с аналитикой), мгновенная верификация (техническая экспертиза (метаданные, цифровые отпечатки), перекрестная проверка с закры-

тыми базами данных, консультации профильных специалистов), классификация по уровню угрозы безопасности, оперативное реагирование (автоматизированные и информационные контрмеры), координация с силовыми структурами, проверка происхождения фейка и анализ эффективности мер реагирования. Как представляется, данный алгоритм может обеспечить комплексную защиту информационного суверенитета с элементами превентивного противодействия и мгновенного реагирования на угрозы стратегического характера. Разработка отечественных технологий ИИ по борьбе с фейками – значимое направление при реализации документов стратегического планирования, таких как Стратегия национальной безопасности РФ (2021), Стратегия научно-технологического развития РФ (2024), Послание Президента РФ Федеральному Собранию (2024) и др.

СПИСОК ИСТОЧНИКОВ

1. Ахмадеев К.Н. Эффективность fake news как инструмента информационной войны в восприятии поколения Z / К.Н. Ахмадеев, М.Г. Бреслер, А.В. Манойло [Электронный ресурс] // Российский социально-гуманитарный журнал. – 2021. – № 3. – С. 8–31. – URL: <https://cyberleninka.ru/article/n/effektivnost-fake-news-kak-instrumenta-informatsionnoy-voyny-v-vospriyatii-pokoleniya-z/viewer> (дата обращения: 02.03.2025).
2. Киуру К.В. Фейковый контент в медиа как объект коммуникационных исследований / К.В. Киуру // Политическое пространство и социальное время: Правда и ложь в политике и искусстве : сборник научных трудов XXXVI Международного Харакского форума 15–17 мая 2019 г., г. Ялта / под ред. Т.А. Сеньюшкиной и Д.Танчича. – Симферополь : АРИАЛ, 2019. – С. 95–99.
3. Ильченко С.Н. Фейк-контроль или новости, которым не надо верить: как нас дурачат СМИ / С.Н. Ильченко. – Ростов-на-Дону : Феникс, 2021. – С. 188 с.
4. Суходолов А.П. Феномен «фейковых новостей» в современном медиaproстранстве / А.П. Суходолов [Электронный ресурс] // Евроазиатское сотрудничество: гуманитарные аспекты: электронный журнал. – 2017. – № 1. – С. 87–106. – URL: <https://cyberleninka.ru/article/n/fenomen-feykovyh-novostey-v-sovremennom-mediaprostranstve> (дата обращения: 10.02.2025).
5. Муратова Н. Fake news: дезинформация в медиа : учебное пособие / Н. Муратова, Н. Тошпулатова, Г. Алимова. – Ташкент, 2020. – 104 с. [Электронный ресурс]. – URL: https://www.unesco.org/sites/default/files/fake_news_rus_0.pdf (дата обращения: 20.02.2025).

REFERENCES

1. Akhmadeev K., Bresler M., Manoilo A. The Effectiveness of Fake News as an Information Warfare Tool in the Perception of Generation Z. *Russian Social and Humanitarian Journal*, 2021, no. 3, pp. 8–31. Available at: <https://cyberleninka.ru/article/n/effektivnost-fake-news-kak-instrumenta-informatsionnoy-voyny-v-vospriyatii-pokoleniya-z/viewer>.
2. Kiuru K.V. Fake Content in Media as an Object of Communication Studies. In Senyushkina T.A., D. Tanchich (eds.). *Political Space and Social Time: Truth and Lies in Politics and Art. Collection of Scientific Papers of the XXXVI International Kharak Forum, May 15–17, 2019, Yalta*. Simferopol, ARIAL Publ., 2019, pp. 95–99. (In Russian).
3. Ilchenko S.N. Fake Control or News That You Shouldn't Believe: How the Media Fools Us. *Rostov-on-Don, Phoenix Publ.*, 2021. 188 p.
4. Sukhodolov A.P. The Phenomenon of "Fake News" in the Modern Media Space. *Eurasian Cooperation: Humanitarian Aspects: Electronic Journal*, 2017, no. 1, pp. 87–106. Available at: <https://cyberleninka.ru/article/n/fenomen-feykovyh-novostey-v-sovremennom-mediaprostranstve>.
5. Muratova N. Fake News: Disinformation in the Media. Tashkent, 2020. 104 p. Available at: https://www.unesco.org/sites/default/files/fake_news_rus_0.pdf.
6. Issers O. Media Fakes: Between Truth and Mystification. *Communication Studies*, 2014, no. 2, pp. 112–123. (In Russian).

6. Иссерс О. Медиафейки: между правдой и мистификацией / О. Иссерс // Коммуникативные исследования. – 2014. – № 2. – С. 112–123.
7. Манойло А.В. «Фейковые новости» угроза национальной безопасности и инструмент информационного управления / А.В. Манойло [Электронный ресурс] // Вестник Московского университета. Серия: Политические науки. – 2019. – № 2. – С. 37–45. – URL: <https://cyberleninka.ru/article/n/fejkovyie-novosti-kak-ugroza-natsionalnoy-bezopasnosti-i-instrument-informatsionnogo-upravleniya> (дата обращения: 20.02.2025).
8. Ильченко С.Н. Как нас обманывают СМИ: манипуляция информацией / С.Н. Ильченко. – Санкт-Петербург: Питер, 2019. – 319 с.
9. Михеев Е.А. Психологические механизмы продвижения недостоверной информации в сети Интернет / Е.А. Михеев [Электронный ресурс] // Психопедагогика в правоохранительных органах. – 2021. – № 4 (87). – С. 423–434. – URL: <https://cyberleninka.ru/article/n/psihologicheskie-mehanizmy-prodvizheniya-nedostovernoy-informatsii-v-seti-internet> (дата обращения: 18.02.2025).
10. Буткевич С.А. Фейк-ньюз и медиа: проблемы дифференциации и пенализации / С.А. Буткевич // Юристы-правоведы. – 2021. – № 2. – С. 74–79.
11. Еремина О.А. Фейки как угроза информационной безопасности (на примере деятельности Вооруженных Сил России) / О.А. Еремина // Многонациональная Россия: вчера, сегодня, завтра: сборник трудов конференции. – Смоленск, 2022. – С. 242–249.
12. Степанов-Егиянц В.Г. К вопросу о криминализации распространения заведомо ложной информации об использовании Вооруженных Сил РФ / В.Г. Степанов-Егиянц, З.И. Абазехова // Евразийский юридический журнал. – 2022. – № 3. – С. 378–380.
13. Нежкина Л.Ю. Распространение фейков как способ манипулирования общественным сознанием в интернет-пространстве / Л.Ю. Нежкина, А.С. Душкин // Российский девиантологический журнал. – 2024. – № 4 (2). – С. 237–249.
14. Манойло А.В. Зарубежные научные подходы к исследованию «фейковых новостей» в мировой политике / А.В. Манойло // Россия и современный мир. – 2020. – № 2 (107). – С. 285–300.
15. Полякова Т.А. Правовые вопросы использования технологий искусственного интеллекта в информационном обществе и в государственном управлении / Т.А. Полякова, Н.А. Троян // Правовое государство: теория и практика. – 2024. – № 3. – С. 85–93.
16. Оломская Н.Н. Механизмы генерирования фейковой информации искусственным интеллектом в современном медиадискурсе / Н.Н. Оломская, А.В. Зиньковская [Электронный ресурс] // Российский социально-гуманитарный журнал. – 2024. – № 2. – С. 207–223. – URL: <https://cyberleninka.ru/article/n/mehanizmy-generirovaniya-fejkovoy-informatsii-iskusstvennym-intellektom-v-sovremennom-mediadiskurse/viewer> (дата обращения: 10.03.2025).
17. Смирнов А.А. «Глубокие фейки». Сущность и оценка потенциального влияния на национальную безопасность / А.А. Смирнов // Свободная мысль. – 2019. – № 5 (1677). – С. 63–84.
18. Хотин Д.Ю. Выявление фейковых фрагментов в текстовых новостных сообщениях с помощью машинного обучения / Д.Ю. Хотин, Г.С. Иванова // Вестник науки. – 2024. – Т. 1, № 6 (75). – С. 1568–1576.
7. Manoylo A.V. Fake News as a Threat to National Security and as a Tool of Information Management. *Moscow University Bulletin. Series: Political Science*, 2019, no. 2, pp. 37–45. Available at: <https://cyberleninka.ru/article/n/fejkovyie-novosti-kak-ugroza-natsionalnoy-bezopasnosti-i-instrument-informatsionnogo-upravleniya>.
8. Ilchenko S.N. How the Media Deceive Us: Manipulation of Information. Saint Petersburg, Piter Publ., 2019. 319 p.
9. Mikheev E.A. Psychological Mechanisms of Promotion of False Information on the Internet. *Psychopedagogy in Law Enforcement*, 2021, no. 4 (87), pp. 423–434. Available at: <https://cyberleninka.ru/article/n/psihologicheskie-mehanizmy-prodvizheniya-nedostovernoy-informatsii-v-seti-internet>.
10. Butkevich S.A. Fake News and Media: Problems of Differentiation and Penalization. *Jurist-Pravoved*, 2021, no. 2, pp. 74–79. (In Russian).
11. Eremina O.A. Fakes as a Threat to Information Security (on the Example of the Activities of the Armed Forces of Russia). *Multinational Russia: Yesterday, Today, Tomorrow. Collection of Conference Papers*. Smolensk, 2022, pp. 242–249. (In Russian).
12. Stepanov-Yegiyants V.G., Abazekhova Z.I. On the Criminalization of the Dissemination of Fake Information on the Use of the Armed Forces of the Russian Federation. *Eurasian Law Journal*, 2022, no. 3, pp. 378–380. (In Russian).
13. Nezhkina L.Y., Dushkin A. S. Dissemination of Fakes as a Way of Manipulating Public Consciousness in the Internet space. *Russian Journal of Deviant Behavior*, 2024, no. 4 (2), pp. 237–249. (In Russian).
14. Manoylo A.V., Popadiuk A.E. Foreign Scientific Approaches to the Study of “Fake News” in World Politics. *Russia and the Contemporary World*, 2020, no. 2 (107), pp. 285–300. (In Russian).
15. Polyakova T.A., Troyan N.A. Legal Issues of the Use of Artificial Intelligence Technologies in the Information Society and in Public Administration. *Pravovoe gosudarstvo: teoriya i praktika = The Rule-of-Law State: Theory and Practice*, 2024, no. 3, pp. 85–93. (In Russian).
16. Olomskaya N.N., Zinkovskaya A.V. Mechanisms of Generation of Fake Information by Artificial Intelligence in Modern Media Discourse. *Russian Social and Humanitarian Journal*, 2024, no. 2, pp. 207–223. Available at: <https://cyberleninka.ru/article/n/mehanizmy-generirovaniya-fejkovoy-informatsii-iskusstvennym-intellektom-v-sovremennom-mediadiskurse/viewer>.
17. Smirnov A. “Deepfakes”. Essence and Assessment of the Potential Impact on National Security. *Free Thought*, 2019, no. 5 (1677), pp. 63–84. (In Russian).
18. Khotin D.Yu., Ivanova G.S. Detecting Fake Fragments in Text News Messages Using Machine Learning. *Vestnik Nauki*, 2024, vol. 1, no. 6 (75), pp. 1568–1576. (In Russian).

19. Исакова А.Г. Применение искусственного интеллекта в расследовании преступлений с использованием технологии «дипфейк» / А.Г. Исакова, А.В. Осин // Вестник науки. – 2024. – Т. 3, № 1 (70). – С. 235–241.

20. Пшегорский А.С. Фейковые новости как феномен современной интернет-среды и перспективы ограничения их распространения / А.С. Пшегорский [Электронный ресурс] // Мир науки. Серия: Социология, филология, культурология. – 2023. – Т. 14, № 3. – URL: <https:sfk-mm.ru/PDF/31CSK323.pdf> (дата обращения: 02.03.2025).

19. Isakova A.G., Osin A.V. Use of Artificial Intelligence in Investigation of Crimes Using Deepfake Technology. *Vestnik Nauki*, 2024, vol. 3, no. 1 (70), pp. 235–241. (In Russian).

20. Pshegorskiy A.S. Fake News as a Phenomenon of the Modern Internet Environment and Prospects for Limiting Their Distribution. *World of Science. Series: Sociology, Philology, Cultural Studies*, 2023, vol. 14, no. 3. Available at: <https:sfk-mm.ru/PDF/31CSK323.pdf>.

ИНФОРМАЦИЯ ОБ АВТОРАХ

Холопова Елена Николаевна – доктор юридических наук, кандидат психологических наук, профессор, старший научный сотрудник ОНК «Институт управления и территориального развития»;

Мулдашев Руслан Нуржанович – Высшая школа права ОНК «Институт управления и территориального развития».

INFORMATION ABOUT THE AUTHORS

Kholopova Elena Nikolaevna – Doctor of Law, Candidate of Psychological Sciences, Professor, Senior Researcher of the Institute of Management and Territorial Development;

Muldashev Ruslan Nurzhanovich – Higher School of Law, Institute of Management and Territorial Development.

Статья поступила в редакцию 23.05.2025; одобрена после рецензирования 25.06.2025; принята к публикации 25.06.2025. The article was submitted 23.05.2025; approved after reviewing 25.06.2025; accepted for publication 25.06.2025.