Научная статья
https://doi.org/10.24412/2220-2404-2026-6-21
УДК 342.7

ПРОБЛЕМЫ ПРАВОВОГО РЕГУЛИРОВАНИЯ ИИ-УСИЛЕННОГО ФИШИНГА И СОЦИАЛЬНОЙ ИНЖЕНЕРИИ ДЛЯ ОБЕСПЕЧЕНИЯ ЦИФРОВОЙ БЕЗОПАСНОСТИ

Степаненко С.Г.
Кубанский государственный технологический университет

Аннотация. *Актуальность. Статья посвящена анализу правовых пробелов в регулировании ИИ-усиленного фишинга и социальной инженерии. Процесс глобализации и сопутствующая ему цифровизация коренным образом изменили ландшафт транснациональной преступности. Развитие генеративного искусственного интеллекта (ИИ) вывело проблему кибербезопасности на новый уровень, создав беспрецедентные угрозы. Рассматриваются риски, вызванные генеративным ИИ, включая deepfake-атаки и персонализированные scam-кампании, действующие нормативные рамки в ЕС (AI Act), США (TAKE IT DOWN Act, штатные законы) и России (экспериментальное регулирование и этический кодекс). Особое внимание уделяется распределению ответственности между провайдерами ИИ-систем, операторами (deployers) и цифровыми платформами. Выявлены ключевые проблемы: отсутствие четкой атрибуции вины при foreseeability misuse, фрагментарность норм и отставание законодательства от технологий. Предложены меры по гармонизации: обязательная маркировка синтетического контента, усиление обязанностей платформ по модерации и введение строгой ответственности за высокорисковые ИИ-инструменты.*

Ключевые слова: *искусственный интеллект, генеративный ИИ, кибербезопасность, ИИ-усиленный фишинг, социальная инженерия, deepfakes, правовое регулирование, ответственность платформ.*

Финансирование: *инициативная работа.*

Original article

PROBLEMS OF LEGAL REGULATION OF AI-ENHANCED PHISHING AND SOCIAL ENGINEERING TO ENSURE DIGITAL SECURITY

Sergey G. Stepanenko
Kuban State Technological University

Abstract. *Relevance. The article analyzes the legal gaps in the regulation of AI-enhanced phishing and social engineering. The process of globalization and the accompanying digitalization have fundamentally changed the landscape of transnational crime. The development of generative artificial intelligence (AI) has taken the issue of cybersecurity to a new level, creating unprecedented threats. The paper examines the risks caused by generative AI, including deepfake attacks and personalized scam campaigns, as well as the current regulatory framework in the EU (AI Act), the USA (TAKE IT DOWN Act, state laws), and Russia (experimental regulation and ethical code). Special attention is paid to the distribution of responsibility between AI system providers, operators (deployers), and digital platforms. The paper identifies key challenges, such as the lack of clear attribution of blame in cases of foreseeable misuse, the fragmented nature of regulations, and the lag between legislation and technology. Harmonization measures proposed: mandatory labeling of synthetic content, strengthening of platforms' responsibilities for moderation, and introduction of strict liability for high-risk AI tools.*

Keywords: *artificial intelligence, generative AI, cybersecurity, AI-enhanced phishing, social engineering, deepfakes, legal regulation, platform liability, EU AI Act, TAKE IT DOWN Act.*

Funding: *Independent work.*

Введение.

Процесс глобализации, затрагивающий все сферы жизни современного общества, в значительной мере отразился в цифровизации, уже ставшей неотъемлемой частью развития человечества. Этико-моральные, социально-экономические и иные проблемы появления и развития искусственного интеллекта (ИИ) неоднократно рассматривались [1] и, несомненно, будут рассматриваться в будущем, как одна из философских проблем современности.

Несмотря на кажущиеся очевидными плюсы данного процесса, необходимо обратить внимание на проблему неизбежных сопутствующих ему угроз, в частности, на вопрос безопасности цифровой среды.

Своеобразная ирония современных процессов в том, что в числе лидирующих направлений использования ИИ-технологий – это сфера безопасности [2, с. 7-8].

Человечеству, по сути, уже совсем скоро придётся жить в мире, где различить живых социальных существ от их цифровых двойников возможно будет лишь в том случае, когда сам ИИ будет сообщать об этом согласно заложенной в него программе [3, с. 97].

Глобализация и сопутствующая ей цифровизация не могли не повлиять на состояние и развитие транснациональной преступности: вопросы кибербез-

опасности выходят за пределы отдельных стран, перерастая в проблему мирового уровня.

Кибератака, инициированная территориально в одной стране, способна найти своих жертв по всему миру.

Преступность реально перестала иметь территориальные, этнические и другие границы. Особую актуальность проблема киберпреступности приобретает с развитием генеративного искусственного интеллекта – технологии, открывающей новые горизонты прогресса, и в то же время создавая беспрецедентные угрозы информационной безопасности.

Цель настоящей работы - выявить пробелы в ответственности платформ (социальные сети, email-сервисы, облачные провайдеры) и операторов ИИ (разработчиков и пользователей моделей), а также предложить пути совершенствования законодательства.

Обсуждение. Результаты.

Сегодня ИИ перестал быть инструментом будущего, изменяя ландшафт киберугроз в режиме реального времени. Генеративный ИИ уже сейчас активно используется злоумышленниками для реализации гиперперсонализированных фишинговых (фисшинг – кража личных данных: логинов, паролей, реквизитов банковских карт и т.п.) атак, создания голосовых клонов (*voice cloning*), генерации *deepfake*-видео (*deepfake, дипфейк* – создание реалистического изображения, видео и аудио с помощью нейросетей, фальшивого по своей сути) и автоматизации многоэтапных схем социальной инженерии.

Масштаб угрозы подтверждается статистикой: в 2025 году количество успешных фишинговых атак увеличилось на 400%, при этом 82,6% фишинговых писем содержали элементы, созданные с помощью ИИ. Прогнозируемые потери от мошенничества с использованием генеративного ИИ (*GenAI-фрод*) уже к 2027 году могут достичь 40 млрд. долларов [4].

Количество киберпреступлений с использование дипфейков возросло в России в 26 раз, только за первые пять месяцев 2026 года похищено не менее 2 млрд. рублей [5].

По данным аналитиков компании «Код Безопасности», среднее время между моментом проникновения злоумышленника в корпоративную сеть и его обнаружением обнаружения (Mean Time to Detect, MTTD) для российских компаний составляет **42 дня – вполне достаточный срок для** изучения инфраструктуры, повышения привилегий, копирования данных и установки бэкдоров (получение несанкционированного доступа к системе) [6].

Развитие генеративных моделей радикально изменило ландшафт киберугроз. Если ранее социальная инженерия опиралась на интуицию и стандартные шаблоны, то сегодня алгоритмы анализируют открытые данные для создания максимально убедительных сценариев обмана. Реальные примеры атак по всему миру:

Австралия, 2024. Злоумышленники использовали voice cloning для имитации голоса CEO крупной строительной компании. Они позвонили финансовому директору и убедили перевести 3,2 млн на «срочный контракт». Голос был настолько убедительным, что жертва не заподозрила подмены.

Индия, 2025. В ходе политической кампании распространили deepfake-видео с кандидатом, произносящим экстремистские высказывания. Ролик набрал 5 млн. просмотров за 6 часов, вызвав массовые беспорядки в нескольких штатах.

Германия, 2024. Банк потерял 1,8 млн. евро после того, как мошенники использовали deepfake-видео для подтверждения крупной транзакции. Система верификации не обнаружила подмены.

США, 2025. Компания по производству медицинского оборудования стала жертвой ИИ-фишинга: злоумышленники сгенерировали письмо от имени партнёра с просьбой обновить платёжные реквизиты. Ущерб составил 7,5 млн.

Россия, 2025. Зафиксирован рост атак с использованием голосовых клонов: в первом квартале 2025 года количество таких инцидентов выросло на 250% по сравнению с аналогичным периодом 2024 года. Особенно часто подделывали голоса руководителей банков и госорганов.

Яркий пример - инцидент с компанией Arup [7] в 2024 году, когда сотрудник перевёл 25 млн. долларов после видеозвонка с deepfake-версиями CFO и юристов. Этот случай показал: ИИ-атаки - больше не гипотетическая угроза, а реальность, с которой сталкиваются организации по всему миру.

Следует отметить некий парадокс современности: технологии, в том числе и цифровые, развиваются глобально, а правовое регулирование продолжает оставаться национально фрагментированным. В этой связи следует отметить формирование принципиальных подходов к решению данного вопроса, вот лишь некоторые из них.

Европейский Союз. ЕС принял наиболее продвинутый риск-ориентированный подход – AI Act[8] (вступает в силу поэтапно, полное применение с августа 2026):

- Deepfakes и синтетический контент отнесены к «limited-risk» системам.
- Провайдеры обязаны обеспечивать machine-readable маркировку (водяные знаки) и детектируемость (ст. 50(2)).
- Deployers (операторы) должны раскрывать факт искусственного происхождения контента пользователям (ст. 50(4)). Исключения – только для правоохранительных целей.
- Запрещены манипулятивные техники, искажающие поведение (ст. 5).

AI Act взаимодействует с DSA (Digital Services Act): платформы несут ответственность за модерацию вредоносного контента, включая ИИ-фишинг. Штрафы – до 3% общего оборота, или 15 млн евро.

Соединенные Штаты Америки. Регулирование фрагментарно (штаты + федеральный уровень). Ключевой акт, регулирующий данный вопрос – TAKE IT DOWN Act [9] (май 2025):

- криминализирует публикацию неконсенсуальных интимных deepfakes (до 2 - 3 лет лишения свободы);

- платформы обязаны удалять контент в течение 48 часов по notice-and-takedown.

Законы о deepfakes приняли 46 штатов (с 2022 года). Предложен Deepfake Liability Act (H.R.6334, 2025), ограничивающий иммунитет платформ по Section 230 за добросовестное отключение вредоносного контента [10]. FTC (Федеральная торговая комиссия, США) преследует deceptive AI-claims (обманные заявления об искусственном интеллекте) (Operation AI Comply)[11]. Однако федерального всеобъемлющего закона нет. Executive Order 2025 (**«Обеспечение подотчётности всех агентств»,** Указ Президента США Д. Трампа) лишь задаёт направление для национальной рамки регулирования проблемы.

Российская Федерация. В России регулирование ИИ носит рамочный характер – как набросок будущего здания без несущих конструкций. Среди основных документов, регулирующих данный вопрос, можно выделить:

- Указ Президента от 10 октября 2019 г. № 490[12], утвердивший Национальную стратегию развития искусственного интеллекта до 2030 года. – задают общие ориентиры, но не содержат конкретных санкций.

- Федеральный закон от 24 апреля 2020 г. № 123-ФЗ [13] – экспериментальное регулирование ИИ в Москве. Представьте: город становится полигоном для тестирования правил, которые потом могут распространиться на всю страну.

- Кодекс этики ИИ (2021) – носит рекомендательный характер [14]. Это как свод хороших манер: полезно, но не обязательно к исполнению.

Проблема в том, что отдельной ответственности за ИИ-фишинг нет. Deepfakes подпадают под:
- ст. 207.1 УК РФ (распространение заведомо ложной информации);
- ст. 152.1 ГК РФ (использование изображения гражданина без согласия).

Но эти нормы создавались до эпохи генеративного ИИ и плохо подходят для новых угроз. Например, deepfake-видео с «выступлением» чиновника может формально подпасть под ст. 207.1, но доказать умысел и связь с конкретным злоумышленником крайне сложно.

Сейчас обсуждаются законопроекты:
- об обязательной маркировке ИИ-контента (штрафы от 10 до 500 тыс. руб.);
- о запрете deepfakes для мошенничества.

Платформы несут общую ответственность в соответствии с Федеральным законом «Об информации» [15] (обязанность удалять противоправный контент), но без учёта специфики ИИ-атак.

Такая несогласованность создаёт лазейки для злоумышленников: атаки могут запускаться из юрисдикций с мягким регулированием, а экстрадиция преступников затруднена, а по сути - чаще всего и невозможна. Развитие генеративного искусственного интеллекта (ИИ) выявило ряд существенных пробелов в правовом регулировании, особенно в части ответственности операторов ИИ - разработчиков, владельцев и пользователей моделей, которые могут быть использованы для создания вредоносного контента (фишинг, deepfake, дезинформация и др.). Среди основных проблем глобальной кибербезопасности можно выделить следующие:

- *отсутствие чёткой и прямой ответственности* за foreseeability misuse (предсказуемое неправомерное использование): Операторы ИИ (разработчики моделей) могут быть привлечены к ответственности, если последствия использования их продукта были предсказуемы. Но доказать причинно-следственную связь почти невозможно. Например, компания выпустила языковую модель, которую затем использовали для генерации фишинговых писем. Утверждение разработчика: «Мы не могли предвидеть, что кто-то применит наш инструмент для мошенничества». И будет прав – ведь модель действительно создавалась для написания текстов, а не для обмана [16]. Таким образом, операторы не мотивированы внедрять дополнительные меры безопасности и контроля;

- *иммунитет платформ* и операторов за пользовательский контент*:* в ряде юрисдикций (например, США – Section 230 и аналогичные нормы) защищают платформы и операторов ИИ от ответственности за контент, созданный пользователями, даже если он вредоносен. Платформа, предоставляющая доступ к генеративной модели, не несёт ответственности за deepfake, созданный пользователем, если не доказано прямое соучастие. Пока нет обязательного proactive detection ИИ-фишинга, соцсети и email-сервисы не обязаны искать deepfakes – только удалять их по жалобе. Как следствие – отсутствие стимулов для проактивного мониторинга и удаления вредоносного ИИ-контента;

- *технические сложности:* водяные знаки можно обойти, а атрибуция атак затруднена. Например, злоумышленник арендует ИИ-сервис на даркнете (AI-as-a-service), проводит атаку и стирает следы. Найти его почти невозможно;

- *фрагментарность и отставание законодательства* от темпов технологического прогресса: законы принимаются годами, а ИИ-инструменты обновляются ежемесячно. В результате нормативная база не успевает охватить новые формы угроз. Например, в России (как и большинстве других стран) отсутствуют специальные составы преступлений за ИИ-усиленное мошенничество или создание deepfake с целью обмана. Таким образом, правоохранительные органы вынуждены применять устаревшие нормы (например, о мошенничестве или клевете), которые не учитывают специфику ИИ;

- *отсутствие международных стандартов и недостаток* в обмене данными об угрозах и координации правоприменения. Кроме того, оператор ИИ, как правило, находится в одной стране, атака соверша-

ется из другой, а жертвы – по всему миру. Экстрадиция и правоприменение осложнены различиями в законах, что создает лазейки для злоумышленников и отсутствие глобальной координации. Экстрадиция в таких случаях – редкость, а законы разных стран противоречат друг другу;

- с*ложности доказывания причинно-следственной связи:* крайне сложно доказать, что конкретный оператор ИИ несёт ответственность за ущерб, причинённый вредоносным контентом, особенно если модель была модифицирована или использована через посредников. Например**,** злоумышленник арендует ИИ-сервис на теневом рынке, проводит атаку и стирает следы. Доказать связь между оператором и преступлением практически невозможно. Отсюда – низкий уровень привлечения к ответственности и высокий уровень безнаказанности;

- *недостаточная прозрачность и отсутствие реестров:* нет обязательных реестров ИИ-моделей и требований к прозрачности их работы, поэтому невозможно отследить, кто и какие модели использует для генерации вредоносного контента, что крайне затрудняет расследование инцидентов и привлечение к ответственности;

- *иммунитет за «добросовестное» отключение контента:* даже если платформы обязаны удалять вредоносный контент по жалобам, они не несут ответственности за его появление, если действуют «добросовестно». Так, платформа удаляет deepfake после жалобы, но не обязана искать такие материалы проактивно, отсюда – замедленная реакция на угрозы и высокий риск распространения вредоносного контента до момента удаления;

- *доступность инструментов*: на даркнете deepfake-сервисы предлагаются по цене от 50 долл. за пакет из 10 видео, а фишинговые боты – от 20 долл. в месяц. И перечисленное – только знаковые проблемы из существующих.

Проведенный анализ глобальных тенденций в развитии ИИ-угроз позволяет сделать вывод о необходимости комплексного подхода в их урегулировании, который должен сочетать: 1) технические меры (маркировка ИИ-контента, AI-detection (процесс обнаружение искусственного интеллекта); 2) правовые механизмы (классификация рисков, ответственность платформ) и 3) международное сотрудничество (гармонизация норм, обмен сигнатурами угроз).

Таким образом, для решения поставленной задачи и минимизации рисков, на наш взгляд, необходимо акцентировать внимание на следующих вопросах:

1. Трансформация киберугроз под влиянием генеративного ИИ: конкретные примеры атак и их последствия.
2. Текущее состояние правового регулирования в ключевых юрисдикциях (ЕС, США, Россия) – с акцентом на пробелы и противоречия.
3. Предложения по совершенствованию законодательства на национальном и международном уровнях, включая:

- введение обязательной маркировки ИИ-контента;
- создание реестров ИИ-моделей;
- разработку механизмов ответственности платформ и операторов ИИ;
- формирование международной конвенции по борьбе с ИИ-угрозами.

Развитие генеративного ИИ знаменует собой переход к новой эре киберугроз. Раньше мошенники полагались на интуицию и общие шаблоны. Сегодня алгоритмы анализируют открытые данные из соцсетей, профессиональных сетей и публичных баз, чтобы создать убедительный сценарий атаки. То, что несколько лет назад могло произойти в фантастических шпионских фильмах, в настоящий момент уже не удивляет: звонок начальника, абсолютно узнаваемый убедительный голос с привычными интонациями с просьбой срочно перевести деньги, или сообщение от коллеги с убедительной просьбой открыть вложение – текст написан в его стиле, с привычными оборотами речи. Сегодня это реальность – и всё благодаря генеративному искусственному интеллекту [17].

Технологии, созданные для упрощения жизни, теперь помогают злоумышленникам: сущность и масштабы ИИ-усиленного фишинга и социальной инженерии поражают уже в настоящий момент.

Проблема заключается в том, что ИИ-усиленный фишинг – это не просто «более умные» письма. Это целый набор технологий:

- *генерация текста* – большие языковые модели (LLM) создают убедительные email и SMS, копируя стиль конкретного человека. Представьте: ИИ анализирует 10 писем вашего коллеги и пишет новое – так, будто его написал он сам;
- *синтез аудио и видео* – voice cloning позволяет клонировать голос за секунды, а deepfakes создают реалистичные видео;
- *автоматизированные кампании* – AI-боты рассылают тысячи персонализированных сообщений, адаптируя сценарий под реакцию жертвы;
- социальная инженерия эволюционировала в *«agentic AI»-атаки*: системы самостоятельно проводят многоэтапные схемы обмана – создают фейковые аккаунты, обходят двухфакторную аутентификацию (MFA), выстраивают доверительные отношения с жертвой.

Заключение.

Традиционные меры защиты – спам-фильтры, двухфакторная аутентификация – теряют эффективность против персонализированных атак. Представьте: вы получаете письмо от «друга» с видеосообщением. Фильтр не видит угроз – письмо не содержит вредоносных ссылок, а видео выглядит реальным. Но это deepfake, созданный за минуты.

Существующие юридические пробелы позволяют операторам ИИ избегать ответственности за создание и распространение вредоносного контента.

Для их устранения, на наш взгляд, необходимо предпринять следующие меры.

1. Введение (усиление) прямой ответственности за предсказуемое неправомерное использование: снятие иммунитета платформ за систематическое игнорирование угроз и введение ответственности платформ за непринятие мер по выявлению и удалению ИИ-фишинга и deepfake, аналогично подходу *Deepfake Liability Act* (США). В России – дополнение законодательства нормами, предусматривающими административную и гражданско-правовую ответственность за неудаление или несвоевременное удаление вредоносного ИИ-контента [18].

2. Введение субсидиарной ответственности операторов ИИ: **о**ператоры (владельцы и разработчики моделей) должны нести ответственность за отсутствие необходимых мер предосторожности (*safeguards*), если их продукт был использован для совершения преступлений. Внедрение требования о наличии *«kill switch»* (механизма экстренного отключения модели) и ведения журналов действий (*logging*).

3. Создание реестров ИИ-моделей:

во-первых, создание Государственных реестров ИИ-моделей, ведение открытых или ограниченных реестров, позволяющих отслеживать, кто и какие генеративные модели использует, с целью оперативного реагирования на инциденты;

во-вторых, создание Международной базы сигнатур ИИ-контента и формирование единой международной базы данных deepfake и других вредоносных ИИ-материалов для ускорения их обнаружения и блокировки [19].

4. Внедрение технических стандартов прозрачности и маркировки:

а) все материалы, созданные с помощью генеративного ИИ, должны содержать *машиночитаемые метки* (водяные знаки), позволяющие автоматически идентифицировать их искусственное происхождение. Исключение — только для случаев, связанных с правоохранительной деятельностью;

б) внедрение систем автоматического обнаружения ИИ-фишинга и deepfake. Платформы (социальные сети, почтовые сервисы, мессенджеры) обязаны интегрировать *AI-detection* инструменты для выявления и блокировки вредоносного контента. Внедрение механизмов *proactive detection* (проактивного обнаружения), а не только реагирования по жалобам пользователей.

5. Образовательные и научные инициативы:

а) обучение судей и следователей особенностям дел с участием ИИ, проведение специализированных курсов для представителей судебной системы и правоохранительных органов;

б) государственная поддержка разработки технологий защиты, **в**ыделение грантов и субсидий на исследования в области *deepfake detection*, атрибуции атак и создания новых методов защиты.

Однако для эффективного противодействия ИИ-угрозам требуется согласование усилий на международном уровне.

Как нам представляется, для разрешения данной проблемы, а также для формирования единой и действенной системы необходимо предпринять следующие шаги:

1.Разработать и принять международной конвенции по ИИ-безопасности и создание единого правового поля: необходимо инициировать разработку международной Конвенции (например, под эгидой ООН, G20, Совета Европы), устанавливающей минимальные стандарты ответственности для операторов и платформ, работающих с генеративным ИИ.

Нормы Конвенции должны распространяться на всех операторов и платформы, чьи услуги доступны гражданам стран-участниц, независимо от юрисдикции регистрации компании.

Данная Конвенция должна определять:

- критерии отнесения ИИ-систем к категориям риска;
- обязательные требования к прозрачности, маркировке и аудиту моделей;
- механизмы привлечения к ответственности за создание и распространение вредоносного контента [20].

2. Гармонизация национальных законодательств:

а) Страны должны привести свои законы к единому стандарту, исключив лазейки, связанные с различиями в регулировании (например, иммунитет платформ, отсутствие ответственности за *foreseeability misuse*).Внедрение единых определений *deepfake, ИИ-фишинга, генеративного контента* и связанных с ними правонарушений;

б) взаимное признание решений, создание механизмов признания судебных решений по делам, связанным с ИИ-угрозами, между странами-участницами [21].

3. Формирование глобальной системы обмена данными об угрозах:

а) создание единой базы данных вредоносных ИИ-материалов (*deepfake*, фишинговые письма, голосовые клоны), доступной для правоохранительных органов и платформ всех стран;

б) внедрение защищённых каналов для обмена данными о новых угрозах, методах атак и технических средствах их выявления.

4. Установление единых технических стандартов:

а) введение международного стандарта *машиночитаемых меток* (водяных знаков) для всего контента, созданного с помощью генеративного ИИ;

б) Обязательное внедрение систем автоматического обнаружения ИИ-фишинга и *deepfake* на всех цифровых платформах с глобальным охватом и введение единых сроков реагирования на выявленные угрозы (например, 24 часа на удаление вредоносного контента).

5. Совершенствование механизмов международного сотрудничества:

а) разработка специальных соглашений о выдаче лиц, причастных к трансграничным ИИ-преступлениям;

б) создание международных следственных групп для расследования сложных киберпреступлений с использованием ИИ.

6. Развитие образовательных и научных инициатив:

а) Организация международных программ повышения квалификации для судей, следователей, экспертов по цифровым доказательствам;

б) выделение совместных грантов на разработку технологий обнаружения вредоносного ИИ-контента и методов атрибуции атак.

Только комплексный подход, основанный на международном сотрудничестве, гармонизации законодательства и внедрении единых технических стандартов, позволит эффективно распределить ответственность между операторами и платформами, что позволит минимизировать глобальные ИИ-угрозы и сохранить доверие к цифровой среде.

Конфликт интересов

Не указан.

Рецензия

Все статьи проходят рецензирование в формате double-blind peer review (рецензенту неизвестны имя и должность автора, автору неизвестны имя и должность рецензента). Рецензия может быть предоставлена заинтересованным лицам по запросу.

Conflict of Interest

None declared.

Review

All articles are reviewed in the double-blind peer review format (the reviewer does not know the name and position of the author, the author does not know the name and position of the reviewer). The review can be provided to interested persons upon request.

Список источников:

1. Например, работы Ю. Фролова Четвертый закон робототехники // Наука и жизнь. – 2025. – № 3 . – С. 90; О.В. Паламарчук О.Т. Интеллект в помощи интеллекту // ОБЩЕСТВО: философия, история, культура: научный журнал. – Краснодар: ХОРС, 2021. – № 5. – С. 25 –30; Баррат Дж. Последнее изобретение человечества: Искусственный интеллект и конец эры Homo sapiens / пер. с англ.. – Изд. 2-е. – М.: Альпина нон-фикшн, 2019. – 396 с.; О΄Коннел М. Искусственный интеллект и будущее человечества / пер. с англ. М. Кудряшова. – М.: Эксмо, 2019. – 272 стр. и др.

2. Ларина Е.С., Овчинский В.С. Искусственный интеллект. Этика и право (Коллекция Изборского клуба). – М.: Книжный мир, 2019. – 192 с.

3. Паламарчук О.Т. Белые пятна искусственного интеллекта. Полемические очерки. – Крапснодар. Перспективы образования, 2025. 232 с.

4. Генеративный ИИ повысит риск банковских мошенничеств [Электронный ресурс]. – Режим доступа: https://big-i.ru/tekhnologii/upravlenie-innovatsiyami/prognoz-generativnyy-ii-povysit-risk-bankovskikh-dipfeykov-i-drugikh-moshennichestv-/ (дата обращения 05.06.2026).

5. Число киберпреступлений с использование дипфейков увеличилось в 26 раз [Электронный ресурс]. – Режим доступа: https://regnum.ru/news/4041741 (дата обращения 05.06.2026).

6. Как реагировать на кибератаку: пошаговый план действий при взломе [Электронный ресурс]. – Режим доступа: https://passwork.ru/blog/kak-rieaghirovat-na-kibierataku/ (дата обращения 05.06.2026).

7. В РОЦИТ рассказали о мерах защиты от злоумышленников в интернете [Электронный ресурс]. – Режим доступа: https://news.rambler.ru/tech/52864439-v-rotsit-rasskazali-o-merah-zaschity-ot-zloumyshlennikov-v-internete/ (дата обращения 05.06.2026).

8. Новая правовая архитектура регулирования ИИ в Европе и её значение для России [Электронный ресурс]. – Режим доступа: (https://alrf.ru/articles/novaya-pravovaya-arkhitektura-regulirovaniya-ii-v-evrope-i-eye-znachenie-dlya-rossii/) (дата обращения 05.06.2026).

9. Принят федеральный закон о борьбе с порнографией, направленный на месть [Электронный ресурс]. – Режим доступа: https://natlawreview.com/article/federal-take-it-down-act-targeting-revenge-porn-becomes-law (дата обращения 05.06.2026).

10. Что такое Раздел 230 Закона о порядочности в сфере коммуникаций? [Электронный ресурс]. – Режим доступа: https://legalclarity.org/what-is-section-230-of-the-communications-decency-act/ (дата обращения 05.06.2026).

11. Федеральная торговая комиссия США борется с ложными заявлениями и схемами, связанными с искусственным интеллектом, в рамках операции AI COMPLY [Электронный ресурс]. – Режим доступа: https://babl.ai/ftc-cracks-down-on-deceptive-ai-claims-and-schemes-with-operation-ai-comply/ (дата обращения 05.06.2026).

*12. Указ Президента РФ от 10 октября 2019 г. №490 «О развитии искусственного интеллекта в Российской Федерации» **[Электронный ресурс]. –** Режим доступа: https://www.consultant.ru/document/cons_doc_LAW_335184/ (дата обращения 05.06.2026).*

*13. Федеральный закон от 24 апреля 2020 г. №123-ФЗ «**О** проведении эксперимента по установлению специального регулирования в целях создания необходимых условий для разработки и внедрения технологий искусственного интеллекта в субъекте Российской Федерации - городе федерального значения Москве, об особенностях обработки персональных данных при формировании региональных составов данных и предоставления доступа к региональным составам данных и внесении изменений в статьи 6 и 10 Федерального закона «О персональных данных»» (с изменениями и дополнениями) [Электронный ресурс]. – Режим доступа: https://base.garant.ru/73945195/ (дата обращения 05.06.2026).*

14. Кодекс этики в сфере искусственного интеллекта (от 26 октября 2021 г.) Электронный ресурс]. – Режим доступа: https://base.garant.ru/406862712/ (дата обращения 05.06.2026).

15. Федеральный закон от 27 июля 2006 года №149-ФЗ «Об информации, информационных технологиях и о защите информации» (с измен. и доп.) [Электронный ресурс]. – Режим доступа: https://base.garant.ru/12148555/ (дата обращения 05.06.2026).

16. «Подлинная политика» в цифровую эпоху [Электронный ресурс]. – Режим доступа: https://www.kommersant.ru/doc/8363969 (дата обращения 05.06.2026).

*17. Обзор GenAI уязвимостей и эксплойтов за Q2 2025 [Электронный ресурс]. – Режим доступа: **https://habr.com/ru/companies/technokratos/articles/938706/** (дата обращения 05.06.2026).*

18. New Deepfake Liability Act to push against abusive AI images [Электронный ресурс]. – Режим доступа: https://www.deseret.com/politics/2025/12/03/reps-maloy-and-auchincloss-introduce-deepfake-liability-act/ (дата обращения 05.06.2026).

19. Минцифры планирует создать реестр „доверенных“ моделей ИИ [Электронный ресурс]. – Режим доступа: https://www.dp.ru/a/2026/03/26/mincifri-planiruet-sozdat (дата обращения 05.06.2026).

20. Регулирование искусственного интеллекта: почему законы не успевают за ИИ [Электронный ресурс]. – Режим доступа: https://www.anti-malware.ru/analytics/Technology_Analysis/The-world-is-rushing-to-regulate-AI (дата обращения 05.06.2026).

21. Сравнительный анализ правового регулирования искусственного интеллекта в России, Китае, США и Европейском союзе [Электронный ресурс]. – Режим доступа: https://www.techerati.com/news-hub/cross-border-genai-misuse-expected-to-cause-40-of-ai-data-breaches-by-2027/ (дата обращения 05.06.2026).

References:

1. For example, the works of Yu. Frolov, The Fourth Law of Robotics // Science and Life. – 2025. – No. 3. – P. 90; O.V. Palamarchuk, O.T. Intellect in Assistance to Intelligence // SOCIETY: Philosophy, History, Culture: Scientific Journal. – Krasnodar: KhORS, 2021. – No. 5. – P. 25–30; Barrat J. The Last Invention of Mankind: Artificial intelligence and the end of the era of Homo sapiens / translated from English.. – Ed. 2-E. – M.: Alpina non-fiction, 2019. – 396 p.; O´Connell M. Artificial Intelligence and the future of mankind / translated from English by M. Kudryashov, Moscow: Eksmo, 2019, 272 p., etc.

2. Larina E.S., Ovchinsky V.S. Artificial intelligence. Ethics and Law (Izborsk Club Collection). – Moscow: Knizhny Mir, 2019. – 192 p.

3. Palamarchuk O.T. White Spots of Artificial Intelligence. Polemical Essays. – Krasnodar. Prospects of Education, 2025. 232 p.

4. Generative AI will increase the risk of banking fraud [Electronic resource]. – Access mode: https://big-i.ru/tekhnologii/upravlenie-innovatsiyami/prognoz-generativnyy-ii-povysit-risk-bankovskikh-dipfeykov-i-drugikh-moshennichestv-/ (accessed on 05.06.2026).

5. The number of cybercrimes involving deepfakes has increased 26-fold [Electronic resource]. – Access mode: https://regnum.ru/news/4041741 (accessed on 05.06.2026).

6. How to Respond to a Cyberattack: A Step-by-Step Plan for Hackers [Electronic resource]. – Access mode: https://passwork.ru/blog/kak-rieaghirovat-na-kibierataku/ (accessed on 05.06.2026).

7. ROSTIT spoke about measures to protect against cybercriminals on the Internet [Electronic resource]. – Access mode: https://news.rambler.ru/tech/52864439-v-rotsit-rasskazali-o-merah-zaschity-ot-zloumyshlennikov-v-internete/ (accessed on 05.06.2026).

8. The new legal architecture of AI regulation in Europe and its significance for Russia [Electronic resource]. – Access mode: (https://alrf.ru/articles/novaya-pravovaya-arkhitektura-regulirovaniya-ii-v-evrope-i-eye-znachenie-dlya-rossii/ (date of access 05.06.2026).

9. A federal law on combating pornography aimed at revenge has been adopted [Electronic resource]. – Access mode: https://natlawreview.com/article/federal-take-it-down-act-targeting-revenge-porn-becomes-law (accessed on 05.06.2026).

10. What is Section 230 of the Communications Decency Act? [Electronic resource]. – Access mode: https://legalclarity.org/what-is-section-230-of-the-communications-decency-act/ (accessed on 05.06.2026).

11. The US Federal Trade Commission is fighting false claims and schemes related to artificial intelligence through Operation AI COMPLY [Electronic resource]. – Access mode: https://babl.ai/ftc-cracks-down-on-deceptive-ai-claims-and-schemes-with-operation-ai-comply/ (date of access 05.06.2026).

12. Decree of the President of the Russian Federation No. 490 dated October 10, 2019, "On the Development of Artificial Intelligence in the Russian Federation" [Electronic resource]. – Access mode: https://www.consultant.ru/document/cons_doc_LAW_335184 / (accessed 06/05/2026).

13. Federal Law No. 123-FZ of April 24, 2020 "On Conducting an Experiment to establish Special Regulation in order to create the necessary Conditions for the Development and implementation of Artificial Intelligence technologies in the Subject of the Russian Federation - the Federal city of Moscow, on the specifics of personal data processing when forming regional data structures and providing access to regional data structures and making amendments to Articles 6 and 10 of the Federal Law "On Personal Data" (with amendments and additions) [Electronic resource]. – Access mode: https://base.garant.ru/73945195/ (accessed on 05.06.2026).

14. Code of Ethics in the Field of Artificial Intelligence (dated October 26, 2021) Electronic resource]. – Access mode: https://base.garant.ru/406862712/ (accessed on June 5, 2026).

15. Federal Law No. 149-FZ dated July 27, 2006, "On Information, Information Technologies, and Information Protection" (as amended. and additional) [Electronic resource]. – Access mode: https://base.garant.ru/12148555/ (accessed on 05.06.2026).

16. "Genuine Politics" in the Digital Age [Electronic resource]. – Access mode: https://www.kommersant.ru/doc/8363969 (accessed on 05.06.2026).

17. Review of GenAI vulnerabilities and exploits for Q2 2025 [Electronic resource]. – Access mode: https://habr.com/ru/companies/technokratos/articles/938706/ (accessed on 05.06.2026).

18. New Deepfake Liability Act to push against abusive AI images [Electronic resource]. – Access mode: https://www.deseret.com/politics/2025/12/03/reps-maloy-and-auchincloss-introduce-deepfake-liability-act/ (accessed on 05.06.2026).

19. The Ministry of Digital Development plans to create a register of "trusted" AI models [Electronic resource]. – Access mode: https://www.dp.ru/a/2026/03/26/mincifri-planiruet-sozdat (accessed on 05.06.2026).

20. Regulation of Artificial Intelligence: Why Laws Are Not Keeping Up with AI [Electronic resource]. – Access mode: https://www.anti-malware.ru/analytics/Technology_Analysis/The-world-is-rushing-to-regulate-AI (accessed on 05.06.2026).

---

21. Comparative Analysis of Legal Regulation of Artificial Intelligence in Russia, China, the United States, and the European Union [Electronic resource]. – Access mode: https://www.techerati.com/news-hub/cross-border-genai-misuse-expected-to-cause-40-of-ai-data-breaches-by-2027/ (accessed on 05.06.2026).

Информация об авторах:
Степаненко Сергей Григорьевич, кандидат исторических наук, доцент кафедры социологии, ФГБОУ ВО «Кубанский государственный технологический университет», email: stepik71@mail.ru
Sergey G. Stepanenko, PhD, Associate Professor of the Department of Sociology, Kuban State Technological University.

Статья поступила в редакцию / The article was submitted 05.06.2026;
Одобрена после рецензирования / Approved after reviewing 19.06.2026;
Принята к публикации / Accepted for publication 20.06.2026.
Автором окончательный вариант рукописи одобрен.