Научная статья
УДК 343.2/7, 343.9
DOI: 10.55001/2587-9820.2023.34.41.005

УГОЛОВНО-ПРАВОВЫЕ И КРИМИНАЛИСТИЧЕСКИЕ АСПЕКТЫ ПРОТИВОДЕЙСТВИЯ РАСПРОСТРАНЕНИЮ И ИСПОЛЬЗОВАНИЮ ДИПФЕЙКОВ В РОССИЙСКОЙ ФЕДЕРАЦИИ

Николай Филиппович Бодров[1], Антонина Константиновна Лебедева[2]
[1,2]Университет имени О. Е. Кутафина (МГЮА), г. Москва, Российская Федерация
[1]bodrovnf@gmail.com.
[2]tonya109@yandex.ru.

Статья подготовлена в рамках госзадания «Российская правовая система в реалиях цифровой трансформации общества и государства: адаптация и перспективы реагирования на современные вызовы и угрозы (FSMW-2023-0006)». Регистрационный номер: 1022040700002-6-5.5.1.

Аннотация. Неконтролируемое распространение и использование технологий искусственного интеллекта, потенциальные и реальные угрозы, которые они создают, являются стимулом для научной дискуссии об отдельных запретах на некоторые технологии искусственного интеллекта. Сдержанная система контроля позволит государству урегулировать правоотношения в сфере распространения и использования генеративного контента без потери темпов технологического развития. В связи с этим в статье рассмотрены вопросы необходимости разработки механизмов правового регулирования дипфейк-технологий в уголовно-правовом и криминалистическом аспектах.

Предлагается авторское определение термина «дипфейк», обоснована необходимость его нормативного закрепления. С учетом необходимости нормативного контроля именно распространения и использования дипфейков рассмотрены критерии дифференциации дипфейков и генеративного контента.

Ключевые слова: дипфейк, генеративный контент, правовое регулирование, специальные знания, маркировка контента

Для цитирования: Бодров, Н. Ф., Лебедева, А. К. Уголовно-правовые и криминалистические аспекты противодействия распространению и использованию дипфейков в Российской Федерации // Криминалистика: вчера, сегодня, завтра : сб. науч. тр. Иркутск : Восточно-Сибирский институт МВД России. 2023. Т. 28. № 4. С. 42–55. DOI: 10.55001/2587-9820.2023.34.41.005

CRIMINAL, LEGAL AND FORENSIC ASPECTS OF COUNTERING THE SPREAD AND USE OF DEEPFAKES IN THE RUSSIAN FEDERATION

Nikolay F. Bodrov[1], Antonina K. Lebedeva[2]
[1,2]Kutafin Moscow State Law University (MSAL), Moscow, Russian Federation,
[1]bodrovnf@gmail.com.
[2]tonya109@yandex.ru

Abstract. The uncontrolled dissemination and use of artificial intelligence technologies, the potential and real threats they pose, are the impetus for a scientific discussion on individual bans on certain artificial intelligence technologies. A restrained system of control will allow the state to regulate legal relations in the sphere of distribution and use of generative content without losing the pace of technological

development. In this regard, the article considers the issues of the need to develop mechanisms of legal regulation of diphake technologies in criminal-legal and criminalistic aspects.

The author's definition of the term "dipfake" is proposed, the need for its normative consolidation is substantiated. Taking into account the need for regulatory control over the dissemination and use of dipfakes, the criteria for differentiation of dipfakes and generative content are considered.

Key words: deepfakes, generative content, legal regulation, forensic science, content labeling

For citation: Bodrov, N.F., Lebedeva, A.K. Ugolovno-pravovye i kriminalisticheskie aspekty protivodejstviya rasprostraneniyu i ispol'zovaniyu dipfejkov v Rossijskoj Federacii [Criminal, legal and forensic aspects of countering the spread and use of deepfakes in the Russian Federation]. Kriminalistika: vchera, segodnya, zavtra = Forensics: yesterday, today, tomorrow. 2023, vol. 28. no. 4, pp. 42–55 (in Russ.). DOI: 10.55001/2587-9820.2023.34.41.005

Введение

Бурное развитие технологий нейросетевого генеративного контента с 2017 года [См. например, 1; 2; 5, с. 438] и по настоящее время обусловило необходимость создания дополнительных правовых механизмов регулирования общественных отношений. Находясь на этапе цифровой трансформации общества важно сформировать механизмы адаптации правовой системы к вызовам, связанным с развитием дипфейк-технологий.

Дискуссия о необходимости правового контроля искусственного интеллекта разрастается по мере усовершенствования технологий искусственного интеллекта (далее – ИИ). Неконтролируемое распространение дипфейков представляет серьезный вызов для правовой системы любого современного государства, его информационной безопасности, в связи с чем потенциальные механизмы правового регулирования должны учитывать возможные негативные последствия распространения и использования дипфейков. Кроме того, с учетом природы дипфейков разработка правовых механизмов должна осуществляться в тесной взаимосвязи с технологическими механизмами, которые создадут дополнительные гарантии упорядочивания оборота генеративного контента.

Так, правовые механизмы должны обеспечивать защиту прав и свобод граждан, в том числе права на неприкосновенность частной жизни, защиту своей чести и доброго имени, на свободу мысли и слова; обеспечивать защищенность системы судопроизводства от фальсификации доказательств. Технологические механизмы противодействия распространению и использованию дипфейков должны, например, позволять осуществлять детекцию дипфейков без наличия у пользователя специальных знаний.

Основная часть

Разработка механизмов правового регулирования должна начинаться в первую очередь с нормативного закрепления понятия дипфейк.

Нам представляется, что ***дипфейк*** – это цифровой продукт в виде текста, графики, звука или их сочетания, сгенерированный полностью или частично при помощи нейросетевых технологий для цели введения в заблуждение или преодоления пользователем систем контроля и управления доступом.

Данное авторское определение, по нашему мнению, фиксирует как возможные виды дипфейков, так и цель их создания.

В сам момент генерации с использованием нейросетевых технологий контент не становится дипфейком, свойства дипфейка он

приобретает в момент распространения с противоправной целью.

Здесь дипфейк выступает неким аналогом «средства повышенной опасности», так как акт распространения генеративного контента, подходящего под определение дипфейка, сам по себе обладает общественной опасностью. Такая опасность заключается в том, что использование реалистичного генеративного контента без маркировки может повлечь за собой последствия в виде:

– нарушения деятельности автоматизированных систем управления и контроля различных объектов;

– нарушения работы ЭВМ и их систем;

– создания условий для несанкционированного вмешательства в информационные системы;

– тяжких и необратимых последствий, связанных с вредом физическому лицу, а также ущербом физическому, юридическому лицу или государству.

Опасность представляет не сама технология, а использование ее результатов без указания на способ получения цифрового продукта. Фактически имеет место деятельность по распространению генеративного контента, сходного с аутентичным до степени смешения за счет достаточно высокой степени развития технологий ИИ.

Критерием разграничения дипфейка и генеративного контента является то, что последний, вне зависимости от степени реалистичности генерации, не связан с сокрытием способа имитации аутентичного контента. Законный оборот генеративного контента должен сопровождаться информацией, достаточной для установления природы его происхождения.

Для упорядочивания оборота генеративного контента без ущерба информационной безопасности необходимо ограничить его использование без соответствующей маркировки. Под безопасностью информационных систем в данном аспекте исследования мы понимаем состояние информационной системы в условиях правовых запретов и ограничений, при котором обеспечивается:

– возможность субъектов правоотношений отличать генеративный контент;

– функционирование системы технологических механизмов предотвращения его противоправного использования.

В такой ситуации законный оборот фактически сводится к маркированию генеративного контента, поддержанию информационной осведомленности пользователей о таком контенте. Законодательство, таким образом, должно ограничивать оборот генеративного контента и устанавливать ответственность за преодоление механизмов информирования о генеративной природе контента.

В Российской Федерации разработаны действующие правовые механизмы маркировки различных видов контента.

Так, Федеральный закон от 29.12.2010 № 436-ФЗ "О защите детей от информации, причиняющей вред их здоровью и развитию"[1] предусматривает требования к маркировке информационной продукции, предназначенной к распространению среди детей, допуская при этом маркировку знаком информационной продукции и (или) текстовым предупреждением об ограничении распространения среди отдельных возрастных категорий.

В Постановлении Правительства Российской Федерации от 22.11.2022 № 2108 "Об утверждении Правил размещения указаний, предусмотренных частями 3 и 4 статьи 9 Федерального закона "О контроле за деятельностью лиц, находящихся под

[1] О защите детей от информации, причиняющей вред их здоровью и развитию : Федер. закон № 436-ФЗ : принят Гос. Думой 21 декабря 2010 года : одобрен Советом Федерации 24 декабря 2010 года : послед. ред. // КонсультантПлюс : сайт. URL: https://www.consultant.ru/document/cons_doc_LAW_108808/ (дата обращения: 27.10.2023).

иностранным влиянием", в том числе требований к их размещению, а также форм указаний, предусмотренных частями 3 и 4 статьи 9 Федерального закона "О контроле за деятельностью лиц, находящихся под иностранным влиянием"[2] прописаны требования к маркировке информации, распространяемой иноагентами.

Кроме того, в связи с запретом на распространение информации об общественных, религиозных и террористических организациях, которые ликвидированы или запрещены судом в соответствии с Федеральным законом от 25.07.2002 № 114-ФЗ «О противодействии экстремистской деятельности»[3], без указания о том, что организация ликвидирована или ее деятельность запрещена (ст. 4 Закона Российской Федерации от 27.12.1991 № 2124-I «О средствах массовой информации»[4]) СМИ должны маркировать любые упоминания подобных организаций.

Сам по себе результат генерации будет соответствовать требованиям закона в тех случаях, когда контент содержит информацию о факте генерации или не предназначен для распространения. Так, обязательную маркировку следует предусмотреть для генеративного контента, специально предназначенного для коммерческого использования. Например, рекламного, кинематографического и другого. Процесс некоммерческой генерации контента, таким образом, находится за пределами правовой регламентации, так как носит сугубо технический характер и без распространения или использования его результатов не несет общественной опасности.

Так, например, сгенерированный контент не отвечает критериям дипфейка, если пользователь в установленном порядке уведомляет о факте генерации (маркирует контент), а результат генерации по своему содержанию не нарушает действующее законодательство.

Контроль за выполнением требования законодателя об обязательной маркировке дипфейков может быть возложен на Федеральную службу по надзору в сфере связи, информационных технологий и массовых коммуникаций, которая уже осуществляет функции контроля и надзора, в том числе и в сфере массовой коммуникации и информационных технологий, а также за соблюдением законодательства в сфере обработки персональных данных.

Для урегулирования правоотношений, связанных с оборотом генеративного контента, требуется дать нормативные дефиниции действиям по распространению и использованию результатов генерации.

Под ***распространением*** для целей настоящей статьи мы предлагаем понимать действие, направленное на получение биометрических персональных данных или результатов их обработки, независимо от вида: изображение лица человека, полученное с помощью фотовидеоустройств; запись голоса человека, полученная с помощью звукозаписывающих устройств (Федеральный закон от 29.12.2022 г. № 572-ФЗ «Об осуществлении идентификации и (или) аутен-

[2] Об утверждении Правил размещения указаний, предусмотренных частями 3 и 4 статьи 9 Федерального закона "О контроле за деятельностью лиц, находящихся под иностранным влиянием", в том числе требований к их размещению, а также форм указаний, предусмотренных частями 3 и 4 статьи 9 Федерального закона "О контроле за деятельностью лиц, находящихся под иностранным влиянием" : Постановление Правительства Российской Федерации от 22.11.2022 № 2108 // КонсультантПлюс : сайт. URL: https://www.consultant.ru/document/cons_doc_LAW_421788/ (дата обращения: 27.10.2023).

[3] О противодействии экстремистской деятельности : Федер. закон № 114-ФЗ : принят Гос. Думой 27 июня 2002 года : одобрен Советом Федерации 10 июля 2002 года : послед. ред. // КонсультантПлюс : сайт. URL: https://www.consultant.ru/document/cons_doc_LAW_37867/ (дата обращения: 27.10.2023).

[4] О средствах массовой информации : Закон Российской Федерации от 27.12.1991 № 2124-1 : ред. от 13.06.2023 // КонсультантПлюс : сайт. URL: https://www.consultant.ru/document/cons_doc_LAW_1511/ (дата обращения: 27.10.2023).

тификации физических лиц с использованием биометрических персональных данных, о внесении изменений в отдельные законодательные акты Российской Федерации и признании утратившими силу отдельных положений законодательных актов Российской Федерации»[5]) и формы их представления (бигдата или результаты обучения нейросетей, доступные для дальнейшего использования), неопределенным кругом лиц или передачу биометрических персональных данных неопределенному кругу лиц (по аналогии с п. п. 1, 9 ст. 2 Федерального закона от 27.07.2006 № 149-ФЗ "Об информации, информационных технологиях и о защите информации"[6]), для цели введения в заблуждение или преодоления пользователем систем контроля и управления доступом.

Использование – обработка биометрических персональных данных, независимо от вида и формы их представления (бигдата или результаты обучения нейросетей, доступные для дальнейшего использования), для цели введения в заблуждение или преодоления пользователем систем контроля и управления доступом.

Правовая регламентация именно распространения и использования генеративного контента на основе биометрических персональных данных, таким образом, представляется нам эффективной с учетом природы дипфейка и невозможности установления запрета на создание генеративного контента.

Учитывая сущность дипфейка, исходя из нашего определения, нам представляется возможным включить его дефиницию в основные понятия Федерального закона от 27.07.2006 № 149-ФЗ.

Правовое регулирование распространения и использования генеративного контента уже сейчас требует формирования отдельного правового института, имеющего самостоятельный предмет, сферу распространения и методы правового регулирования. Специфика методов правового регулирования распространения и использования генеративного контента тесно связана с технологическими закономерностями развития и функционирования нейросетевых алгоритмов. Только своевременное с точки зрения темпов совершенствование технологий нейросетевой генерации развитие правового регулирования может обеспечить адекватный уровень информационной безопасности.

В качестве ключевого компонента информационной безопасности в данной сфере выступает «информационная определенность», то есть фактическая возможность выявлять дипфейки пользователями информационных ресурсов и сетей (за счет маркировки) и участниками судопроизводства (с возможностью привлечения лиц, обладающих специальными знаниями).

С этой целью необходима разработка правовых механизмов: введение норм об обязательной маркировке дипфейков и использовании водяных знаков при их создании, разработка, таким образом, общеправового запрета на распространение и использование дипфейков без соответствующей специальной маркировки.

В сфере использования технологических механизмов действенным средством могло бы стать создание

---

[5] Об осуществлении идентификации и (или) аутентификации физических лиц с использованием биометрических персональных данных, о внесении изменений в отдельные законодательные акты Российской Федерации и признании утратившими силу отдельных положений законодательных актов Российской Федерации : Федер. закон № 572-ФЗ : принят Гос. Думой 21 декабря 2022 года : одобрен Советом Федерации 23 декабря 2022 года // КонсультантПлюс : сайт. URL: https://www.consultant.ru/document/cons_doc_LAW_436110/ (дата обращения: 27.10.2023).

[6] Об информации, информационных технологиях и о защите информации : Федер. закон № 149-ФЗ : принят Гос. Думой 8 июля 2006 года : одобрен Советом Федерации 14 июля 2006 года : послед. ред // КонсультантПлюс : сайт. URL: https://www.consultant.ru/document/cons_doc_LAW_61798/ (дата обращения: 27.10.2023).

сервисов по осуществлению детекции дипфейков без наличия у пользователя специальных знаний.

Однако существующие сервисы пока не обеспечивают стабильного результата детекции. На данный момент нет данных и для прогнозирования их эффективности в будущем. Стоит отметить, что большинство из них созданы зарубежными корпорациями, алгоритмы работы которых маловероятно станут известными из-за их правовой охраны режимом коммерческой тайны.

Организации, разрабатывающие нейросети для создания генеративного контента, ведут работы по созданию систем для детекции продуктов генерации. Однако данные системы пока используют вероятностный подход к оценке результатов детекции и чаще всего эффективны в выявлении результатов генерации только одной конкретно взятой за основу нейросетью.

Информационная определенность в самом общем виде обозначает принципиальную возможность субъекта определить параметры на основе известных и доступных методов, алгоритмов.

Информационная определенность в сфере законного оборота генеративного контента, таким образом, должна быть связана с возможностью осуществления детекции дипфейков без наличия у пользователя специальных знаний.

Применительно к сфере судопроизводства информационная определенность в сфере законного оборота генеративного контента должна быть связана с возможностью осуществления детекции дипфейков с использованием специальных знаний специалиста или эксперта.

На сегодняшний день объективно имеются значительные сложности выявления и доказывания факта использования и распространения дипфейков с применением специальных знаний. Методическое обеспечение судебной экспертизы не готово к выявлению подобных технологий, хотя генеративный контент появился уже достаточно давно и сейчас получил крайне широкое распространение.

Дипфейки могут быть созданы с использованием различных по своей природе алгоритмов и технологий, что затрудняет разработку методики их выявления. В научной литературе практические вопросы использований при исследовании генеративного контента рассматриваются пока в наиболее общем виде [См., например, 9, с. 90; 10, с. 118]. Каких-либо конкретных алгоритмов действий экспертов при исследовании дипфейка ученые не предлагают, чаще всего статьи заканчиваются обзором некоторых дипфейк-программ. Более того, пока специалисты пытаются разработать методические подходы к исследованию дипфейков, созданных при помощи одних нейросетей, появляются десятки новых, которые уже учли «ошибки» предыдущих разработчиков.

Подобное отставание криминалистических походов от темпов развития технологий создания дипфейков обусловлено еще и использованием генеративных состязательных сетей (generative adversarial networks «GANS») [3, с. 53], которые работают по принципу перепроверки и коррекции результатов генерации.

Кроме того, неясным остается, как действовать гражданину для защиты своих биометрических персональных данных если, например, вследствие недостаточной методической проработанности по результатам производства судебной экспертизы эксперт придет к выводу в форме НПВ (не представляется возможным решить задачу). Таким образом, на данном этапе методического развития судебной экспертизы гражданин фактически лишается возможности доказать факт того, что нарушающий его права цифровой продукт является дипфейком.

С противоправным оборотом генеративного контента тесно связаны проблемы установления и дифференциации ответственности за правонарушения, связанные с распро-

странением и использованием дипфейков.

Некоторые специалисты, рассуждая о гражданско-правовом регулировании дипфейков, утверждают об отсутствии необходимости их регулирования, ссылаясь на существующие нормы гражданского законодательства, например на ст. 152.1 ГК РФ[7]: «изображение человека, в том числе и дипфейк, – это его персональные данные, которые законодательство тоже защищает. Гражданин имеет право потребовать удаления любой своей фотографии или видео с ним, если они были размещены в Интернете без его позволения»[8]. Позволим себе не согласиться с данной точкой зрения. Во-первых, данная статья ГК РФ не распространяется на голос лица, таким образом, гражданин лишается возможности судебной защиты своего права в случае распространения дипфейка с его голосом. На данный момент в отечественном законодательстве именно голос гражданина практически не защищен.

Во-вторых, дипфейк представляет собой не изображение гражданина, а цифровой продукт, созданный на основе его внешнего облика. Согласно вышеуказанной статье, защищается конкретное изображение, а не внешний облик лица в целом. На настоящий момент, с учетом современного уровня науки и техники, не представляется возможным доказать, какое именно изображение, видео или аудиозапись использовались для генерации цифрового продукта. А статья 152.1 ГК РФ распространяет свое действие на какое-либо конкретное изображение гражданина, которое было распространено без согласия.

Крайне непродуктивным нам представляется подход к ограничению гражданско-правовой ответственности за распространение дипфейков только рамками ст. 152 ГК РФ. Так, при исследовании продуктов речевой деятельности человека в рамках судебной лингвистической экспертизы в рамках дел о защите чести, достоинства и деловой репутации есть действующие алгоритмы производства экспертизы. Имеются методические рекомендации для экспертов для установления негативного характера распространенных о лице сведений, а также формы выражения данной информации. Применительно к дипфейкам этот алгоритм не работает. Не ясно, как будет определяться характер порочащей информации, если с помощью дипфейка создаются информативные свидетельства о событиях, которых в действительности никогда не существовало.

Первоочередной задачей гражданско-правовой регламентации является разработка норм, регулирующих механизм получения согласия на создание, использование и распространение генеративного контента на основе биометрических данных человека.

Таким образом, для урегулирования правоотношений, возникающих при распространении и использовании биометрических персональных данных (изображение лица человека, полученное с помощью фотовидеоустройств; запись голоса человека, полученная с помощью звукозаписывающих устройств) для цели правомерного использования, следует разработать правовые механизмы:

1) предложить форму согласия (по аналогии с авторским договором) на генерацию контента с изображением и/или голосом лица;

2) в качестве обязательных условий согласия необходимо предусмотреть указание целей распространения генеративного контента (про-

[7] Гражданский кодекс Российской Федерации (часть первая) : ГК : принят Гос. Думой 21 октября 1994 года : послед. ред. // КонсультантПлюс : сайт. URL: https://www.consultant.ru/document/cons_doc_LAW_5142/ (дата обращения: 30.10.2023)..

[8] Может ли Том Круз засудить свой дипфейк // Skillbox Media. URL: https://skillbox.ru/media/business/mozhet_li_nastoyashchiy_tom_kruz_zasudit_poddelnogo_kak_reguliruyutsya_dipfeyki_v_rossii_i_ssha/?ysclid=lnvnj80plr830348080 Дата публикации: 02.04.2021.

дукта на его основе), а также ограничить сферы распространения, не предусмотренные таким соглашением;

3) установить санкции за создание, распространение и использование генеративного контента без согласия лица, чьи биометрические персональные данные были использованы для генерации.

Как известно, сделка (ст. 153 ГК РФ) не может быть заключена в противоправных целях. Поэтому согласие не может предусматривать генерацию контента для изготовления дипфейков.

Таким образом, создание института согласия на создание, распространение и использование биометрических персональных данных может быть действенным механизмом гражданско-правового регулирования оборота генеративного контента и средством ограничения создания дипфейков.

Вопросы административной ответственности касаются таких аспектов, как нарушение порядка маркировки генеративного контента, аналогичной тому порядку, который предусмотрен для сфер возрастной маркировки информационной продукции, маркировки материалов экстремистских организаций, а также иноагентов.

Особый интерес представляет проблема установления уголовной ответственности за совершение преступлений с использованием дипфейков. В вопросах введения уголовной ответственности за преступления с использованием дипфейков нет единой позиции.

С одной стороны, Комиссия Правительства по законопроектной деятельности не поддержала законопроект о введении уголовной ответственности за распространение дипфейков[9].

С другой стороны, Генеральная прокуратура Российской Федерации предусмотрела учет дипфейков в качестве средств, используемых при совершении преступлений, включив в статистическую отчетность, код 049 «с использованием технологии дипфейк», в справочник № 25-ГП "Предметы, устройства и другие средства, использованные при совершении преступлений", утвержденный Приказом Генпрокуратуры России от 09.12.2022 № 746 "О государственном едином статистическом учете данных о состоянии преступности, а также о сообщениях о преступлениях, следственной работе, дознании, прокурорском надзоре"[10].

Как известно, в статьях 207, 207.1, 207.2, 207.3 УК РФ в качестве элемента состава преступления (субъективная сторона) предусмот-

[9] В правительстве не поддержали законопроект об уголовной ответственности за дипфейки // Информационное агентство ТАСС : сайт. URL: https://tass.ru/obschestvo/17922853 (дата обращения: 06.11.2023).

[10] О государственном едином статистическом учете данных о состоянии преступности, а также о сообщениях о преступлениях, следственной работе, дознании, прокурорском надзоре" (вместе с "Положением о государственном едином статистическом учёте данных о состоянии преступности, а также о сообщениях о преступлениях, следственной работе, дознании, прокурорском надзоре", "Инструкцией о порядке предоставления первичных статистических данных о состоянии преступности, о сообщениях о преступлениях, следственной работе и дознании в государственную автоматизированную систему правовой статистики", "Правилами заполнения учётных документов, используемых для предоставления первичных статистических данных о состоянии преступности, о сообщениях о преступлениях, следственной работе и дознании в государственную автоматизированную систему правовой статистики") : Приказ Генпрокуратуры России от 09.12.2022 № 746 // КонсультантПлюс : сайт. URL: https://www.consultant.ru/document/cons_doc_LAW_433986/ (дата обращения: 27.10.2023).

рена «заведомая ложность» информации.

Некоторые исследователи, формулируя определение фейка, презюмируют его заведомую ложность: «Во-первых, необходимо внести в правовую базу термин «Deepfake» или же «Заведомо ложный материал, основанный на методе синтеза с использованием искусственного интеллекта» [7, с. 90].

По нашему мнению, к дипфейкам данный подход неприменим. Для обывателя заведомая ложность материала не может быть очевидна из-за природы самого дипфейка – материала, созданного, чтобы выглядеть максимально реалистично, вводить в заблуждение.

Информация в дипфейке может быть признана заведомо ложной для пользователей только в тех случаях, когда:

– есть объективные подтверждения наличия у лица необходимой компетенции (например, монтажер на киностудии);

– лицо создало дипфейк;

– получена квалифицированная экспертная оценка дипфейка;

– была снята маркировка.

Кроме того, возможно и распространение дипфейка в случае, если гражданин не знал о его ложности, например, ознакомившись на каком-либо информационном ресурсе с дипфейком (например, с выступлением известного политического деятеля), разместил его на своей странице, имея четкое представление о реальности данного коммуникативного события.

Как и составы преступлений и иных правонарушений, относящиеся к категории вербального экстремизма (например, ст. ст. 280, 280.1, 280.3, 282 УК РФ[11], ст. ст. 20.3.1, 20.29 КоАП РФ[12] и т. д.), а также связанные с пропагандой или публичной демонстрацией (ст. 20.3 КоАП РФ), составы правонарушений, связанных с незаконным оборотом генеративного контента, по всей видимости должны быть формальными. С момента распространения и/или использования дипфейка преступление или иное правонарушение является оконченным, так как неопределенному кругу лиц (с учетом малой эффективности механизмов блокировки) предоставляется доступ к информации, способной ввести в заблуждение или способствовать в преодолении пользователем систем контроля и управления доступом.

В дискуссии об уголовной ответственности некоторые ученые говорят, что способом решения проблемы с дипфейками могла бы стать квалификация использования дипфейк-технологий как клеветы (ст. 128.1 УК РФ) при соблюдении некоторых условий: «при условии, что будет доказано, что поддельные изображения распространялись со злым умыслом. Представляется правильным квалифицировать подобные деяния по ч. 2 ст. 128.1 «Клевета» УК РФ, предусматривающей ответственность за распространение заведомо ложных сведений, порочащих честь и достоинство потерпевшего, совершенное в Интернете» [8, с. 115].

Однако позволим себе с этим не согласиться. Как мы уже указывали, говорить о заведомой ложности применительно к дипфейкам некорректно. Как мы указывали ранее при рассмотрении дел в рамках ст. 152 ГК РФ, для квалификации действий по распространению дипфейков по ст. 128.1 УК РФ невозможно применить существующие механизмы для

[11] Уголовный кодекс Российской Федерации : УК : принят Гос. Думой 24 мая 1996 года : одобрен Советом Федерации 5 июня 1996 года : послед. ред. // КонсультантПлюс : сайт. URL: https://www.consultant.ru/document/cons_doc_LAW_10699/ (дата обращения: 30.10.2023).

[12] Кодекс Российской Федерации об административных правонарушениях : КоАП : принят Гос. Думой 20 декабря 2001 года : одобрен Советом Федерации 26 декабря 2001 года : послед. ред. // КонсультантПлюс : сайт. URL: https://www.consultant.ru/document/cons_doc_LAW_34661/ (дата обращения: 30.10.2023).

обоснования порочащего характера информации.

Отдельного внимания заслуживает проблема ужесточения уголовной ответственности за фальсификацию доказательств (ст. 303 УК РФ) с использованием нейросетей. В зарубежной практике подобные случаи уже имели место. Так, например, в одном из бракоразводных процессов супруга синтезировала голос мужа, сформировав поддельную фонограмму с угрозами [4].

Угрозы, связанные с дипфейками, могут быть предотвращены некоторыми существующими уголовно-правовыми нормами. Так, например, согласно действующему законодательству, имеется ответственность за незаконные изготовление и оборот порнографических материалов или предметов (ст. 242 УК РФ), соответственно, лицо может понести ответственность по данной статье в случае создания порнодипфейка. Однако данная статья не учитывает возможный моральный ущерб в случае, если дипфейк был создан, например, с целью порномести, а также содержал внешний облик и голос реальных людей.

Должна ли быть ответственность за какое-либо преступление ужесточена, если оно было совершенно с использованием дипфейков?

Да, так как создание такого материала характеризуется:

1) Иллокутивной целью – введение в заблуждение. Реализуется посредством распространения заведомо ложной информации.

2) Особой сложностью в выявлении. Для выявления нужны программы или эксперты.

То есть большей степенью общественной опасности.

Действующее законодательство не учитывает все угрозы, связанные с дипфейками. Это приводит к тому, что многие случаи использования дипфейков не подпадают под действие закона:

– ответственность за создание и распространение дипфейков, которые могут нанести вред личности, имуществу или общественной безопасности;

– международные стандарты в области регулирования дипфейков. С помощью одной нейросети дипфейки могут создаваться по всему миру. Если в одном государстве есть обязательная, предусмотренная законом маркировка и возможность детекции, это не значит, что аналогичные меры будут применяться в другом.

В какой-то степени вопросы межгосударственного регулирования отношений, связанных с созданием, распространением и использованием генеративного контента, решаются на корпоративном уровне.

Многие крупные компании предлагают свои программные продукты для защиты от дипфейков (Adobe, Microsoft, Intel и т. д.). Однако эти инструменты, как мы отмечали ранее, пока не обеспечивают стабильного результата детекции.

На сервисе госзакупок был размещен заказ МВД России на выполнение научно-исследовательской работы «Исследование возможных способов выявления признаков внутрикадрового монтажа видеоизображений, выполненного с помощью нейронных сетей». Шифр «Зеркало (Верблюд)» (в рамках ГОЗ)[13]. Исходя из данных сайта, он был завершен в январе 2023 года и предназначен для исследования дипфейковых видео, однако никакой информации о применении данного метода исследования мы не обнаружили.

Однако подобные меры (без должного правового регулирования) работают только в мире, где развитием и разработкой нейросетей для ге-

[13] Выполнение научно-исследовательской работы «Исследование возможных способов выявления признаков внутрикадрового монтажа видеоизображений, выполненного с помощью нейронных сетей». Шифр «Зеркало (Верблюд)» // Единая информационная система в сфере закупок : офиц. сайт. URL: https://zakupki.gov.ru/epz/contract/contractCard/payment-info-and-target-of-order.html?reestrNumber=1770802535821000006#contractSubjects (дата обращения: 27.10.2023).

нативного контента занимались бы только крупные компании, которые заботятся об этике использования технологий ИИ. Как, например, у DALL-E2 имеются всевозможные фильтры, чтобы препятствовать пользователю создавать условно противоправный контент, такой как порнография или дипфейки с участием реальных людей.

На данный момент такие системы детекции дипфейков создаются крупными компаниями, которые готовы нести ответственность за сгенерированный контент, именно для детекции контента, созданного их же нейросетями. Однако существует огромное количество площадок, где выкладываются как готовые программы, так и открытый программный код, используемые для создания различных дипфейков. Создатели таких продуктов не заботятся ни об этичности созданных им технологий, ни о безопасности пользователей.

Мы уже неоднократно в своих работах указывали, что в качестве технологического механизма в противодействии дипфейкам могут стать включение в тело файлов, генерируемых нейросетями ватермарок (от англ. *Watermark* – водяные знаки) и дополнительной служебной информации в метаданных [См., например, 6, с. 12].

Ватермарки являются некими цифровыми подписями, встраиваемыми в цифровые файлы. Ватермарки позволяют идентифицировать источник происхождения цифрового продукта, предотвратить несанкционированное использование или распространение, помогают в отслеживании перемещения контента.

Сами компании, разрабатывающие нейросети для создания генеративного контента, создают системы для детекции такого контента. Однако данные системы дают ответ с определенной степенью вероятности и чаще всего работают только с конкретной нейросетью.

Ватермарки, безусловно, могут стать эффективным средством в борьбе с дипфейками, однако если правовые механизмы маркировки не будут обеспечены на законодательном уровне, то маркировать контент будут только крупные компании – разработчики продуктов на основе искусственного интеллекта, а то многообразие различных «специалистов», создающих сервисы для создания дипфейков, не будет маркировать контент и соблюдать технологические требования к программам.

Так, например, лидеры на поле генеративного контента, OpenAI и Google, заявили, что будут маркировать контент, созданный ИИ, водяными знаками, чтобы предотвратить распространение дипфейков и, как следствие, тотальную дезинформацию. Но говорить об использовании ватермарок и обеспечении на корпоративном уровне надлежащего технологического уровня его детекции можно только в случае реализации целого комплекса мер правовой регламентации.

Выводы и заключение

Комплекс таких норм в перспективе включает ряд механизмов по обеспечению обязательной маркировки генеративного контента и иных механизмов правового и технологического регулирования:

– дефиниция «дипфейк» должна быть включена в основные понятия Федерального закона от 27.07.2006 № 149-ФЗ "Об информации, информационных технологиях и о защите информации";

– нормы о маркировании дипфейков должны предусматривать применение наиболее защищенных от удаления технологических решений сферы цифровых водяных знаков;

– весьма востребованным представляется вопрос о выявлении цифровых водяных знаков без использования специальных знаний (для тех сфер, которые требуют оперативной проверки контента);

– необходима также скорейшая проработка экономико-правовых аспектов реализации технологии цифрового маркирования и распознавания цифровых водяных знаков;

– в скором будущем объективно возникнут сложности антимонопольного регулирования деятельности по генерации дипфейк-контента с учетом массового характера внедрения таких технологических решений;

– отдельного внимания заслуживает ответственность за удаление генеративного контента пользователями, получившими доступ к такому контенту.

Предлагаемые правовые и технологические механизмы регулирования общественных отношений в сфере создания, распространения и использования генеративного контента могут стать важным шагом в защите интересов государства и общества от угроз информационной безопасности, которые создают технологии ИИ. Данные механизмы, на наш взгляд, позволят найти определенный баланс в урегулировании оборота генеративного контента, защитить права и законные интересы граждан от нарушений, совершаемых с помощью дипфейков, защитить общество от дезинформации и способствовать повышению доверия граждан к информации, распространяемой в Интернете.

СПИСОК ИСТОЧНИКОВ

1. *Dekom P.* I Have Absolute Proof / P. Dekom // Ent. & Sports Law. 2017. Т. 34.

2. *Garimella K.* Image based misinformation on WhatsApp / K. Garimella, D. Eckles // Proceedings of the Thirteenth International AAAI Conference on Web and Social Media (ICWSM 2019). 2017.

3. Generative Adversarial Networks: An Overview / A. Creswell [и др.] // IEEE Signal Processing Magazine. 2018. Т. 35. Generative Adversarial Networks. № 1. С. 53–65.

4. *Ryan, P.* Deepfake Audio Evidence used in U.K. Court to Discredit Dubai Dad.The National UAE. Retrieved June 27, 2023, URL: https://www.thenationalnews.com/uae/courts/deepfake-audio-evidence-used-in-uk-court-todiscredit-dubai-dad-1.975764.

5. *Tsagourias N.* The Rule of Law in Cyberspace: A Hybrid and Networked Concept? / N. Tsagourias. 2017. The Rule of Law in Cyberspace. С. 433–451.

6. *Бодров, Н. Ф., Лебедев,а А. К.* Перспективы судебно-экспертного исследования синтезированной звучащей речи // Законы России: опыт, анализ, практика : ежемес. тематич. правовой журн. 2021. № 3. С. 9–13.

7. *Данилова, В. А., Левкин, Д. М.* Правовые аспекты регулирования "Deepfake" технологии в России // Право и государство: теория и практика : науч. журн. 2022. № 7(211). С. 88–91.

8. *Добробаба, М. Б.* Дипфейки как угроза правам человека // Lex Russica (Русский закон) : науч. юрид. рецензир. журн. 2022. Т. 75, № 11 (192). С. 112–119.

9. *Елизарова, М. Г.* Современные методы фальсификации фонограмм. Выявление признаков компьютерного синтеза голоса и речи // Теория и практика судебной экспертизы в современных условиях : мат-лы IX Междунар. науч.-практ. конф., Москва, 26–27 января 2023 года. М. : Блок-Принт, 2023. С. 88–92.

10. *Лужинская, Е. Л,. Чванкин, В. А.* Особенности исследования изображений внешнего облика человека, измененного при помощи программных средств // Вопросы криминологии, криминалистики и судебной экспертизы : науч. журн. 2022. № 2 (52). С. 116–121.

11. Толковый переводоведческий словарь / Л.Л. Нелюбин. 3-е изд., перераб. М. : Флинта: Наука, 2003.

REFERENCES

1. *Dekom P. I* Have Absolute Proof. Ent. & Sports Law. 2017, vol. 34.

2. *Garimella K.* Image based misinformation on WhatsApp. Proceedings of the Thirteenth International AAAI Conference on Web and Social Media (ICWSM 2019). 2017.

3. *Creswell, A.* Generative Adversarial Networks: An Overview. Generative Adversarial Networks. 2018, vol. 35, no. 1, pp. 53-65.

4. *Ryan, P.* Deepfake Audio Evidence used in U.K. Court to Discredit Dubai Dad.The National UAE. Retrieved June 27, 2023, https://www.thenationalnews.com/uae/courts/deepfake-audio-evidence-used-in-uk-court-todiscredit-dubai-dad-1.975764.

5. *Tsagourias, N.* The Rule of Law in Cyberspace: A Hybrid and Networked Concept? The Rule of Law in Cyberspace. 2017.

6. *Bodrov, N. F.* Perspektivy sudebno-ekspertnogo issledovaniya sintezirovannoj zvuchashchej rechi [Prospects of forensic research of synthesized sounding speech]. Zakony Rossii: opyt, analiz, praktika – Laws of Russia: experience, analysis, practice. 2021, no. 3, pp. 9-13. (in Russian).

7. *Danilova, V. A.* Pravovye aspekty regulirovaniya "Deepfake" tekhnologii v Rossii [Legal aspects of "Deepfake" technology regulation in Russia].Moscow, 2022, no. 7(211), pp. 88-91. (in Russian).

8. *Dobrobaba, M. B.* Dipfejki kak ugroza pravam cheloveka [Deepfakes as a threat to human rights]. Lex Russica (Russian law). 2022, vol. 75, no. 11(192), pp. 112-119. (in Russian).

9. *Elizarova, M.G.* Sovremennye metody fal'sifikacii fonogramm. Vyyavlenie priznakov komp'yuternogo sinteza golosa i rechi [Modern methods of falsification of phonograms. Identification of signs of computer synthesis of voice and speech]. Teorija i praktika sudebnoj jekspertizy v sovremennyh uslovijah – Theory and practice of forensic examination in modern conditions : : materialy IX Mezhdunarodnoj nauchno-prakticheskoj konferencii, Moskva, 26–27 janvarja 2023 goda. [materials of IX International Scientific and Practical Conference, Moscow, 26–27 january, 2023. Moscow, 2023, pp. 88-92. (in Russian).

10. *Luzhinskaya, E.L.* Osobennosti issledovaniya izobrazhenij vneshnego oblika cheloveka, izmenennogo pri pomoshchi programmnyh sredstv [Peculiarities of research of images of human appearance, changed by means of program means]. Voprosy kriminologii, kriminalistiki i sudebnoj jekspertizy – Issues of criminology, criminology and forensic examination. 2022, no. 2(52), pp 116-121. (in Russian).

11. *Nelyubin, L.L.* Tolkovyj perevodovedcheskij slovar' [Explanatory dictionary of translation studies]. M.: Flinta: Nauka, 2003. (in Russian).

ИНФОРМАЦИЯ ОБ АВТОРАХ

Бодров Николай Филиппович, кандидат юридических наук, доцент кафедры судебных экспертиз. Университет имени О. Е. Кутафина (МГЮА). 125993, Российская Федерация, г. Москва, ул. Садовая-Кудринская, д. 9, стр.1.

Лебедева Антонина Константиновна, кандидат юридических наук, доцент кафедры судебных экспертиз. Университет имени О. Е. Кутафина (МГЮА). 125993, Российская Федерация, г. Москва, ул. Садовая-Кудринская, д. 9, стр.1.

INFORMATION ABOUT THE AUTHORS

Nikolay F. Bodrov, Candidate Law, Lecturer of the Department of Forensic Expertise's department. Kutafin Moscow State Law University (MSAL). Sadovaya-Kudrunskaya Str., 9, Moscow, Russian Federation, 125993.

Antonina K. Lebedeva, Candidate Law, Lecturer of the Department of Forensic Expertise's department. Kutafin Moscow State Law University (MSAL). Sadovaya-Kudrunskaya Str., 9, Moscow, Russian Federation, 125993.