УДК 343.985.2

Наталья Вячеславовна ШЕПЕЛЬ,
доцент кафедры уголовного процесса
Калининградского филиала
Санкт-Петербургского университета МВД России,
кандидат юридических наук, доцент
shepelnv@mail.ru

Марина Олеговна ЯНГАЕВА,
доцент кафедры криминалистики
Барнаульского юридического института МВД России,
кандидат юридических наук, доцент
marina-ymo@mail.ru

ИСПОЛЬЗОВАНИЕ ВОЗМОЖНОСТЕЙ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА ПРИ СОЗДАНИИ ДИПФЕЙКОВ

USING THE POTENTIAL OF ARTIFICIAL INTELLIGENCE IN CREATING DEEPFAKES

Статья посвящена проблемам, связанным со стремительным развитием технологий искусственного интеллекта, в том числе при создании звуковой информации дипфейка и использовании их в преступных целях. Авторами представлен обзор алгоритмов, применяемых для создания убедительной подделки голосовых сообщений, синтезированных с помощью искусственного интеллекта. В связи с этим разработка технологий, которые могут автоматически обнаруживать и оценивать звуковую информацию является актуальной. В статье представлен обзор методов для обнаружения дипфейка, рассмотрены тенденции и направления, применяемые с указанными технологиями.

The article is devoted to the problems related to the rapid development of artificial intelligence technologies, including the creation of deepfake audio information and their use for criminal purposes. The authors present an overview of the algorithms used to create convincing forgery of voice messages synthesized using artificial intelligence. In this regard, the development of technologies that can automatically detect and evaluate audio information is relevant. The article provides an overview of methods for detecting deepfake and considers trends related to these technologies.

Ключевые слова: дипфейк, искусственный интеллект, нейросети, звуковая информация, голос, преступление.

Keywords: *deepfake, artificial intelligence, neural networks, audio information, voice, crime.*

Технологии искусственного интеллекта (далее – ИИ) с самого своего появления демонстрировали удивительные достижения в решении задач, с которыми, как традиционно считалось, способен справиться только человеческий разум. Не удивительно, что с древних времен человек старался облегчить умственный труд, создавая простейшие формы искусственного интеллекта, такие как абак, калькулятор, которые обладали способностью решения ограниченного количества задач.

В настоящее время наблюдается стремительное развитие технологий искусственного интеллекта и машинного обучения, которые все активнее проникают в нашу жизнь. По мере понимания ключевых свойств таких технологий появляются разные спектры моделей интеллектуальной деятельности мышления, в том числе применяемые в преступных целях.

Обращение к данным статистической отчетности на примере УМВД России по Калининградской области позволяют отметить актуальность преступлений, совершенных с использованием информационно-телекоммуникационных технологий. По итогам 2023 г. количество зарегистрированных в регионе преступлений в данной сфере возросло в 1,6 раза (с 2596 до 4252)[1].

Телефонные звонки по схеме «Родственник попал в беду» все еще действуют, но постепенно уходят в прошлое. С недавних пор мошенники стали чаще использовать «дипфейк» – технологию, которая позволяет подделывать изображение и голос. Подавляющее большинство (62%) россиян уверены, что умеют отличать ложь (фейк) от правды[2]. Однако в данном случае желаемое выдается за действительное.

В 2022 г. были выявлены почти 4 тысячи уникальных фейков, общее число копий фейковых сообщений составило порядка 10 миллионов. При этом специалисты прогнозируют ежегодное двукратное увеличение фейковой информации[3]. Ожидается многократный рост и «дипфейков», создаваемых с помощью технологий искусственного интеллекта[4].

Вполне очевидно, что в сложившихся условиях юридическая наука не может оставаться в стороне от осмысления технологии глубокого синтеза дипфейка. Термин «Deepfake» происходит от словосочетания «deep learning» («глубокое обучение») и слова «fake» (подделка). Дипфейк (deepfake) – методика синтеза изображения, основанная на технологии генеративно-состязательных нейросетей (GAN) [6, с. 93].

С использованием технологий искусственного интеллекта можно изготовить убедительные подделки изображения и голоса человека, которые в силу развития алгоритмов глубинного обучения становятся все более реалистическими [7, с. 144]. Одним из наиболее актуальных способов совершения преступлений является использование ИИ для подделки голоса и дальнейшей рассылки таких сообщений с просьбами о переводе денег. Так, злоумышленники подделали голос и видео жительницы г. Барнаула В. и смогли совершить кражу денег ее знакомых. Лже-сообщения получили десятки ее контактов из Телеграм-аккаунта, который был взломан. Пять человек откликнулись на просьбу, и на неизвестную карту отправили суммарно свыше 100 тысяч рублей[5].

Развитие и внедрение в повседневную жизнь технологий искусственного интеллекта, в свою очередь, создает условия для расширения поля деятельности преступников. Актуальность таких технологий и проблемные аспекты их использования требуют совершенствования знаний правоохранителей в сфере цифровых технологий, в том числе при создании звуковой информации. В настоящее время в Государственную думу внесен законопроект, устанавливающий право на защиту голоса гражданина как объекта его личных неимущественных прав. Инициатива направлена на предотвращение недобросо-

1 Комплексный анализ состояния оперативной обстановки и результатов оперативно-служебной деятельности УМВД России по Калининградской области за 2023 год. URL: https://39.мвд.рф/activities/reports/main_reports/2023-год/информационно-аналитическая-записка (дата обращения: 03.09.2024).

2 Исследование: более 60% россиян уверены, что умеют отличать фейки от правдивых новостей // ТАСС. URL: https://tass.ru/obschestvo/14005711.

3 В России к 2024 году число фейков может вырасти более чем в два раза. URL: https://www.ap22.ru/paper/V-Rossii-k-2024-godu-chislo-feykov-mozhet-vyrasti-bolee-chem-v-dva-raza.html.

4 Топ 5 фейков 2023 года: как не дать себя обмануть. URL: https://www.gosnews.ru/news/obshchestvo/top_5_feykov_2023_goda_kak_ne_dat_sebya_obmanut/ (дата обращения: 08.09.2024).

5 На Алтае зарегистрирован первый случай мошенничества с помощью искусственного интеллекта. URL: https://vk.com/wall-76916936_88591 (дата обращения: 26.09.2024).

вестного использования голоса. В документе подчеркивается: использование и распространение голоса гражданина, включая записи с голосом или воссозданные с помощью технологий голосовые имитации, возможны только с его согласия. После смерти это право переходит к его детям или супругу, а при их отсутствии – к родителям.

Вместе с тем в законопроекте указано, что если голос человека был использован без согласия и распространен в Интернете, гражданин может требовать удаления записи, запрета на ее дальнейшее использование и распространение. Пояснительная записка к законопроекту предлагает дополнить Гражданский кодекс РФ новой статьей 152.3, которая установит охрану голоса как объекта личных неимущественных прав гражданина по аналогии с его изображением. Кроме того, планируется ввести уголовное наказание за дипфейки, что нашло отражение в законопроекте N 506240-8 «О внесении изменений в Уголовный кодекс Российской Федерации». Предложенной законодательной инициативой планируется установить санкции за использование технологии подмены лица и голоса. Подмена голоса станет отягощающим обстоятельством.

Тот факт, что технология подделки голоса все чаще используется обычными людьми, а не специалистами, открывает путь для ее неправомерного использования и заставляет нас рассматривать дипфейк через призму уголовного закона. В уголовно-правовом контексте можно уверенно говорить о том, что дипфейк может выступать средством совершения любых преступлений, связанных с распространением недостоверной либо иной вредоносной информации. Рассмотрим некоторые из них.

1. Финансовые преступления. Существует несколько методов, которые используются преступниками для этой цели:

– фальсификация фотографий или видео в компрометирующей манере, когда преступники вымогают деньги у жертвы. Если деньги не будут уплачены, материалы будут отправлены всем их близким. Такая тактика все чаще используется в форме атаки программ-вымогателей Deepfake;

– обход аутентификации по Face ID, в частности, на сайтах онлайн-знакомств. Преступник получает доступ к конфиденциальной информации и реквизитам кредитной карты, что позволяет ему совершать платежи удаленно;

– использование программного обеспечения deepfake audio для имитации голоса высокопоставленного человека, например руководителя организации. Голосовой звонок или телефонный звонок с конкретными инструкциями по переводу денег на банковский счет преступника – основная форма этого преступления;

– использование инсайдерской информации от высокопоставленных должностных лиц организации или политических деятелей с целью манипулирования рынком. Выдавая себя за должностное лицо, преступник получает доступ к определенной информации, которая может дать ему преимущество в сфере торговли.

2. Вымогательство. Использование так называемых «deepnudes» фейковых интимных фотографий жертвы на основе ее реальных фото из социальных сетей, с целью шантажа и дальнейшего требования выкупа.

3. Кража личных данных. Использование изображения и голоса другого лица. Этот метод часто сочетается с мошенничеством, когда преступник выдает себя за другое лицо. Например, в марте 2019 г. с помощью дипфейковых видео управляющего директора британской энергетической компании было похищено около 240 миллионов долларов[1]. В 2021 г. в Китае разоблачена группа мошенников, которые два года обманывали госсистему распознавания лиц с помощью технологии дипфейк, создающей реалистичные замены лиц на видео или заставляющей

1 Мошенник подделал голос CEO и украл $243 тыс. при помощи технологии deepfake // Inc. Russia. URL: https://incrussia.ru/news/deepfake-moshennik-ukral-243-tys/.

фотографии двигаться[1]. Так, злоумышленники покупали фотографии реальных людей в высоком качестве в даркнете, затем «оживляли» их с помощью технологии дипфейк. Через специально перепрошитые смартфоны, у которых система распознавания лиц работает некорректно и принимает дипфейк за реальное лицо, подделывали налоговые накладные.

4. Политические манипуляции репутацией и мнением. Поддельный контент может быстро распространяться с помощью ботов и ферм троллей[2] для манипуляции будущим целой страны.

5. Не исключена возможность использования дипфейка при совершении должностных преступлений. Здесь наиболее наглядным примером может выступать использование технологии при служебном подлоге (например, для фальсификации приложений к официальному документу). Технология глубокого синтеза[3] может быть также использована при совершении такого преступления, как фальсификация доказательств и результатов оперативно-разыскной деятельности. В таких случаях современная редакция ст. 303 УК РФ позволяет надлежащим образом квалифицировать содеянное.

Для полного понимания технологии изготовления голосового дипфейка сотруднику правоохранительных органов необходимо обладать некоторыми техническими познаниями. Глубокое обучение успешно применяется для решения различных сложных задач, начиная от анализа больших данных и заканчивая компьютерным зрением [4, с. 83]. Рассмотрим некоторые аспекты копирования голоса на примере доступного цифрового хранилища. Связывание речи и текста – одна из функций искусственного интеллекта, где в качестве прочтения текста на основе наборных данных содержится пара «текст-звуковая файловая запись». Проблемным вопросом в связывании текста и речи является сохранение характерных особенностей голоса. Когда человек общается по телефону, мозг, получая звуковой сигнал, оценивает тембр голоса, манеру разговора и интонацию, возникает уверенность в узнаваемости голоса. Отдельное внимание таким особенностям голоса уделяется при создании цифровой копии голоса, «переносе голоса» или «клонировании голоса». В числе основных форм переноса голоса используются автокодировщики, это тип нейронных сетей, сжимающие входные данные (часть Encoder) до компактного внутреннего представления, а затем посредством обучения учатся разжимать их из этого представления обратно (часть Dekoder), чтобы восстановить исходные данные. На вход модели заносятся две аудиозаписи, причем звук со второй записи наносится на первую. Из первого звука с помощью Content Encoder выделяется, что было произнесено, из второго с помощью Speaker Encoder выделяются характеристики получаемого голоса, т.е. речь определенного объекта. Декодером генерируется, что говорит объект и как он это говорит. В итоге получается, что имеющееся в первой записи произносится голосом объекта из второй записи. Существуют и другие способы, например, основанные на генеративно-состязательных сетях (GAN) или диффузионных моделях[4].

В сети Интернет имеется достаточное количество бесплатных и открытых инструментов, но качественный звуковой дипфеик создать при помощи доступного инструмен-

1 Мошенники в Китае с помощью дипфейков обманули госсистему распознавания лиц на $76,2 млн // Право на vc.ru/ URL: https://vc.ru/legal/228953-moshenniki-v-kitae-s-pomoshyu-dipfeikov-obmanuli-gossistemu-raspoznavaniya-lic-na-762-mln.

2 «Ферма троллей» или «фабрика троллей» – это организованная группа интернет-троллей, которая стремится повлиять на политические взгляды и принятие решений. Интернет-тролль – это человек, который намеренно публикует оскорбительные или провокационные сообщения в интернет (в социальных сетях, новостной группе, форуме, чате, онлайн-игре).

3 Технология глубокого синтеза (Deep Synthesis) – это технология, позволяющая использовать алгоритмы глубокого обучения и иные алгоритмы генерации и синтеза для создания текста, изображений, аудио, видео, виртуальных сцен и других информационных объектов.

4 МВД занялось системой распознавания фейковых видео. URL: https://www.tadviser.ru/index.php/Проект:-Зеркало_(Верблюд)_система_МВД_для_распознавания_фейковых_видео (дата обращения: 23.09.2024).

та совсем не просто. Человеку необходимо владеть навыками программирования на «Python» и иметь опыт работы в программах по обработке звука, но и в этом случае результат будет не сильно убедительным. Для создания высококачественного звукового дипфейка необходимо обратиться к закрытым и платным сервисам. Например, Microsoft в 2023 г. представил алгоритм, способный по звуку продолжительностью всего три секунды воспроизвести голос человека, причем на разных языках. Этот сервис пока еще не доступен пользователям, так как Microsoft продолжает исследования по совершенствованию сервиса. В рамках реализации подобных программ действует платформа ElevenLabs, которая на своем сайте дает возможность изготовления голосовых дипфейков. Для этого лицу при подготовке к преступлению достаточно загрузить аудиозапись голоса потенциальной жертвы продолжительностью не менее пяти минут и необходимый для озвучивания текст. Коротких, отрывочных фраз для этой платформы будет недостаточно, поскольку ElevenLabs приняли меры для ограничения доступа к своему сервису и изменили правила его пользования, блокируя доступ анонимным пользователям, которые создают дипфейки на основе самостоятельно загруженных голосов [5, с. 58].

Особое место занимает способ генерации звука с помощью искусственного интеллекта старых голосовых сообщений владельцев аккаунтов сетей Telegram, WhatsApp. Этот способ позволяет убедительно просить в долг или на благотворительность деньги у списка контактов. Эта схема считается новой для нашей страны. Сначала злоумышленники получают доступ к аккаунту, затем присылают потенциальным потерпевшим из круга контактов его владельца просьбу о переводе определенной суммы денег с генерируемым голосом, отправляют банковскую карту с нужными реквизитами. Аудиосообщение дублируется в личную переписку и во все чаты, где состоит владелец акаунта. Схема опасна тем, что мошенники используют несколько факторов идентификации жертвы – аккаунт, голос и банковскую карту, соответственно, для приготовления к преступлению злоумышленнику необходимо иметь готовый текст и образец голоса потерпевшего, образцы фотографий и банковских карт, что в дальнейшем может стать доказательством приготовления к преступлению [2, с. 50].

В обязательном порядке при совершении преступления лицу необходимо донести звуковую информацию до потерпевшего, для этого используется телефонная связь, посредством IP (Internet Protokol) телефонии, основанная на принципах передачи голосовой информации в сетях с пакетной коммутацией. Устройствам в сети присваивается IP-адрес, уникальный код, по которому разные устройства опознают друг друга, передовая голос. Компания провайдер владеет и подключает пользователей к IP-телефонии, которая соединяет абоненты с IP-адресом с помощью специального оборудования. Суть заключается в том, что абонент номер один, набирая абонента номер два, направляет запрос провайдеру со своим IP-адресом. Далее провайдер находит его в сети и устанавливает интернет-связь между ними. Голос абонента номер один переводится в цифровой аудиофайл. Провайдер отправляет аудиофайл на компьютер или телефон абонента. Для того чтобы воспользоваться IP-телефонией, нужно подобрать поставщика услуги, заключить с ним договор, зарегистрировать личный кабинет, выбрать тарифы и опции, приобрести IP-телефон, который преобразует аналоговый сигнал в цифровой и отправляет его через Интернет. В заключение установить софтфон, специальное программное обеспечение, с помощью которого можно звонить с компьютера или смартфона. Установить автоматическую телефонную станцию с помощью, которой можно объединить в работу сразу нескольких сотрудников или подразделений. Это даст возможность принимать большое количество звонков одновременно при возможности сохранения записи разговора. В настоящее время IP-телефония активно используется злоумышленниками с функцией подмены номера телефона, когда абонент, получающий вызов, видит известный ему номер [3, с. 204].

Масштабный механизм по вовлечению граждан в преступную деятельность с использованием ИИ для подделки голоса и рассылки голосовых сообщений с просьбами о переводе денег предопределил необходимость разработки мер борьбы с ними. Так, особую актуальность на данный момент приобретает разработанная система «Антифрод», которая используется для блокировки звонков и SMS с подменных номеров. В 2023 г., по данным Роскомнадзора, система предотвратила более 756 миллионов вызовов телефонных мошенников, самый большой объем которых производился через операторов связи, имеющих небольшое количество абонентов, в основном направляющих эти вызовы транзитом абонентам крупных операторов[1].

В феврале 2024 г. крупные мобильные операторы стали пропускать мошеннические звонки чаще, в отличие от небольших операторов. Как пояснил мобильный оператор «Мегафон», мошенники продолжают звонить, как с подменой номера на нумерацию малых операторов, не подключенных к «Антифроду», так и без подмены с SIM-карт крупных операторов. В «ВымпелКоме» отметили, что мошенники научились получать доступ к личному кабинету абонента, создавать и использовать несколько виртуальных номеров. В начале февраля 2024 г. к «Антифроду» были подключены более 700 телекоммуникационных операторов, это чуть выше половины всех имеющихся компаний в России [1, с. 138].

На фоне сложившейся ситуации для решения проблемы распознавания дипфейков МВД России начало использовать разработку АО «Научно-промышленной компании «Высокие технологии и стратегические системы» «Зеркало» («Верблюд»). Эта разработка предназначена для выявления поддельных голосовых сообщений и видеороликов в рамках проведения видеотехнической экспертизы. Кроме того, задачами технологии будут являться:

– проведение теоретических и экспериментальных исследований по изучению вопроса выявления признаков внутрикадрового монтажа видеоизображений, выполненного с помощью нейронных сетей;

– проведение анализа информации о технологиях применения искусственного интеллекта при выполнении монтажа видеоизображений;

– изучение и систематизирование данных о зарубежном опыте исследований в области выявления признаков монтажа видеозаписей, в том числе выполненного с помощью нейронных сетей;

– проведение анализа имеющихся на российском рынке разработок в области выявления признаков внутрикадрового монтажа видеоизображений, в том числе выполненного с помощью нейронных сетей;

– определение наличия возможности проведения экспертного исследования цифровых видеозаписей, созданных с помощью нейронных сетей, содержащихся в видеофайлах распространенных форматов и представленных при отсутствии информации об обстоятельствах их получения, в том числе из интернет-ресурсов;

– определение комплекса диагностических признаков внутрикадрового монтажа видеоизображений, выполненного с помощью нейронных сетей, в том числе визуального синтеза человеческого образа;

– определение способа выявления признаков внутрикадрового монтажа видеоизображений, выполненного с помощью нейронных сетей, учитывая дальнейшее развитие и совершенствование искусственного интеллекта;

– определение критериев пригодности видеоизображений, ограничивающие использование установленных способов анализа. Одновременно АНО «Диолог Регионы» разработал систему «Зефир», которая распознает дипфейк и реальные видео- или аудиозаписи[2].

1 Система Роскомнадзора «Антифрод» заблокировала свыше 260 млн звонков мошенников. URL: https://tass.ru/obschestvo/17953897 (дата обращения: 02.09.2024).

2 Не верь ушам своим: голосовые дипфейки. URL: https://www.kaspersky.ru/blog/audio-deepfake-technology/35694 (дата обращения: 20.08.2024).

Таким образом, противодействие преступной деятельности с использованием возможностей искусственного интеллекта, в том числе с созданием звуковой информации, требует системного подхода и изучения различных аспектов данного явления.

Анализ способов и механизмов формирования дипфейков позволит выявить уязвимые места в его технологии, способствовать более эффективному выявлению и расследованию преступлений. Также необходимо отметить проблемные аспекты перспективы развития дипфейков, они таковы, что в скором времени появится генерация голоса в реальном времени, в результате дистанционное мошенничество и другие виды преступлений, в совершении которых используется клонированный голос, возрастет в разы. Модели для создания дипфейков постоянно совершенствуются, а качество неуклонно растет. Востребованность в выявлении таких технологий искусственного интеллекта у правоохранительных органов достаточно велика, необходимо создавать методики для предотвращения противоправного использования таких технологий.

Библиографический список

1. Абдразяпов, Р.Р. Привлечение к ответственности операторов связи за подмену абонентского номера как одно из средств профилактики телефонных мошенничеств / Р.Р. Абдразяпов // Аграрное и земельное право. – 2023. – N 6 (222). – С. 138-139.

2. Буграева, А.Р. Особенности уголовно-правовой квалификации преступлений, совершенных с использованием DEEPFAKE технологий / А.Р. Буграева // Уголовное законодательство: вчера, сегодня, завтра : материалы ежегодной всероссийской научно-практической конференции. Санкт-Петербург, 18-19 мая 2021 г. / под ред. Т.А. Огарь, Д.М. Кокина. – СПб.: Санкт-Петербургский университет МВД России. 2021. – Ч. 2. – С. 48-51.

3. Гайдин, А.И. Механизм хищений денежных средств, совершаемых с использованием технологий IP–телефонии и программ подмены номеров / А.И. Гайдин, И.С. Звягин, И.С. Садырин // Вестник Воронежского института МВД России. – 2022. – N 3. – С. 202-206.

4. Довгаль, В.А. Применение глубокого обучения для создания и обнаружения поддельных изображений, синтезированных с помощью искусственного интеллекта /В.А. Довгаль // Вестник АГУ. – 2021. – N 4 (291). – С. 82-94.

5. Киселев, А.С. О необходимости правового регулирования в сфере искусственного интеллекта: дипфейк как угроза национальной безопасности / А.С. Киселев // Вестник Московского государственного областного университета. – 2021. – N 3. – С. 54-64.

6. Красовская, Н.Р. Технологии манипуляции сознанием при использовании дипфейков как инструмента информационной войны в политической сфере / Н.Р. Красовская, А.А. Гуляев // Власть. – 2020. – N 4. – С. 93-98.

7. Лемайкина, С.В. Проблемы противодействия использования дипфейков в преступных целях / С.В. Лемайкина // Юристъ-Правоведъ. – 2023. – N 2 (105). – С.143-148.