УДК 343.985.2

ИСПОЛЬЗОВАНИЕ ДИПФЕЙКОВ КАК СОВРЕМЕННЫЙ СПОСОБ СОВЕРШЕНИЯ ПРЕСТУПЛЕНИЙ

Шепель Наталья Вячеславовна

Санкт-Петербургский университет МВД России, г. Калининград
e-mail: shepelnv@mail.ru

USING DEEPFAKES AS A MODERN WAY TO COMMIT CRIMES

Shepel Natalia V.

St. Petersburg University of the Ministry of Internal Affairs of Russia, Kaliningrad

Аннотация: Статья посвящена проблемам, связанным со стремительным развитием технологий искусственного интеллекта, в частности использованием дипфейков в преступных целях. Автором представлен некоторый перечень преступлений, где дипфейки выступают ка средство их совершения, поскольку основной проблемой в настоящее время является не осведомленность граждан о наличии такого способа совершения преступления, что порождает веру в содержание полученной информации и новые случаи мошенничества. Представлен обзор технологии для обнаружения дипфейка, рассмотрены тенденции и направления ее применения.

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

Abstract: The article is devoted to the problems associated with the rapid development of artificial intelligence technologies, in particular the use of deepfakes for criminal purposes. The author presents a certain list of crimes where deepfakes act as a means of committing them, since the main problem currently is the lack of awareness of citizens about the existence of such a method of committing a crime, which gives rise to faith in the content of the information received and new cases of fraud. An overview of the technology for deepfake detection is presented, trends and directions of its application are considered.

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

Для цитирования: Шепель Н.В. Использование дипфейков как современный способ совершения преступлений // Проблемы правовой и технической защиты информации. 2024. №12. С.103-106.

For citation: Shepel N.V. Using deepfakes as a modern way to commit crimes // Legal and Technical Problems of Information Security. 2024. No. 12. P.103-106.

С развитием тенденций искусственного интеллекта (далее – ИИ) возникают новые технологии, основанные на применении генеративно-состязательных нейросетей (GAN), так называемые «дипфейки». 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) [2, с.93]. На сегодняшний день существует несколько разновидностей дипфейков: звуковые, фото- и видео дипфейки (сочетающие комбинацию динамического изображения и голоса человека). Вполне очевидно, что в сложившихся условиях юридическая наука

не может оставаться в стороне от осмысления технологии глубокого синтеза дипфейка.

Так, согласно статистике 65% граждан нашей страны уверены, что умеют отличать фейк от правды. Однако, проведенным опросом установлено обратное – 52% тестируемых не смогли отличить ложь от правды. По прогнозам специалистов в области информационных технологий, негативная статистика будет только расти, так, в 2022 году выявлено 10 миллионов фейков, в 2023 – 12,5 млн., к 2025 году прогнозируется рост до 15 миллионов. По прогнозам ожидается многократный рост в сети фейков, в том числе и «дипфейков», создаваемых с помощью алгоритмов глубокого обучения [4].

Для дипфейков с изображением лица человека генератор должен изучить исходные медиафайлы человека, чтобы узнать геометрию его лица, текстуру кожи, особенности мимики и стиль речи. Далее дискриминатор выявляет неестественные факторы в поддельном файле, который ему предоставляет генератор. Этот состязательный процесс двух сетей приводит к быстрому улучшению выходных мощностей генератора. При наличии достаточного количества данных, предоставленных генеративно-состязательной сетью, гиперреалистично переставляет лицо человека в кадр видео или изображения. Клонирование голоса происходит по аналогичному принципу с использованием программ синтеза речи.

В этой связи выявить и раскрыть преступления с использованием дипфейков, крайне сложно, преступники имеют хорошее техническое оснащение, что обуславливает сложность в организации расследования. Однако, несмотря на то что технология дипфейков достигла того уровня, когда практически невозможно отличить подлинный файл от поддельного, нынешние алгоритмы по-прежнему оставляют после себя тонкие подсказки, обнаруживаемые человеческим глазом. Эти видимые недостатки действуют как сигналы, которые помогают разобраться в подлинности предоставляемого материала.

В уголовно-правовом контексте можно уверенно говорить о том, что дипфейк может выступать средством совершения любых преступлений, связанных с распространением недостоверной 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 в виде криптовалюты у своей 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 – самая 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.

– используют ChatGPT и другие чат-боты с ИИ, чтобы сгенерировать грамотный «человеческий» текст для фишингового письма или рассылки.

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 и выдача его за себя. Этот метод часто сочетается с мошенничеством, когда преступник выдает себя за другое лицо. Например, в марте 2019 года с помощью дипфейковых видео управляющего директора британской энергетической компании было похищено около 240 миллионов долларов. В 2021 году в Китае разоблачена группа мошенников, которые два года обманывали госсистему распознавания лиц с помощью технологии дипфейк, создающей реалистичные замены лиц на видео или заставляющей фотографии двигаться. 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 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 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.

Следует помнить, что алгоритмы, наборы данных и модели глубоких нейронных сетей продолжают совершенствоваться и даже специалисты могут испытывать трудности с тем, чтобы отличить подлинные носители информации от сфабрикованных. Для того, чтобы пресекать преступления в данной сфере граждане должны обладать здоровым скептицизмом, обладать критическим мышлением, быть осмотрительным и осторожным при оценке любой поступающей информации.

Необходимо помнить, что мошенники активно используют фактор внезапности, погружая жертву в эмоциональное напряжение, вызывая яркую реакцию, не предоставляя время на обдумывание происходящего, для того чтобы легче получить желаемое [1 с.72]. Но основная проблема состоит в том, что граждане массово не осведомлены о присутствии такого способа совершения преступления как использование дипфейков в мошеннических схемах, что порождает безусловную веру в содержание полученной информации и новые случаи мошенничества.

Понимание методов, лежащих в основе дипфейков, имеет значение для выявления их уязвимостей и разработки мер защиты от их вреда. 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 от реального видео или аудио записи [3].

Таким образом, модели для создания дипфейков постоянно совершенствуются, а качество неуклонно растет. Востребованность в выявлении таких технологий искусственного интеллекта у правоохранительных органов достаточно велика, со способами совершения таких преступлений необходимо создавать методики для предотвращения противоправного использования таких технологий.

Библиографический список

1. Доув М. Психология мошенничества. Методы убеждения и мошенничества. Издательство: Гуманитарный центр, 2022. 224 с.

2. Красовская Н.Р., Гуляев А.А. Технологии манипуляции сознанием при использовании дипфейков как инструмента информационной войны в политической сфере // Власть. 2020. № 4. С. 93-98.

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

4. Топ 5 фейков 2023 года: как не дать себя обмануть: сайт. URL: https://www.gosnews.ru/news/obshchestvo/top_5_feykov_2023_goda_kak_ne_dat_sebya_obmanut/ (дата обращения: 08.08.2024).