Секция «Финансовый государственный контроль»

Развитие дипфейков как угроза современному обществу

Научный руководитель – Фонотов Андрей Георгиевич

Бондарев Александр Владимирович
Студент (магистр)
Национальный исследовательский университет «Высшая школа экономики», Факультет социальных наук, Москва, Россия
E-mail: alexbond2001@yandex.ru

В современности, с развитием технологий, цифровизацией,
и появлением новых методов оптимизации трудоемких процессов посредством применения искусственного интеллекта (далее – ИИ), нейросетей, перед которыми ставятся задачи облегчить монотонную работу, снизить долю погрешности в виде ошибок пользователей, человеческого фактора, появляются и новые угрозы, требующие скорейшего реагирования и регулирования.

Одной из современных угроз, с которой столкнулось общество, явлется угроза со стороны дипфейков, а точнее злоумышленников-пользователей данной технологией с целью причинения вреда посредством фишинговых атак, дезинформации и т.д. Термин дипфейк происходит от английских понятий «глубинное обучение» (англ. deep learning) и «подделка» (англ. fake). Таким образом, дипфейк – это синтез аудио и визуальной информации с последующим наложением на исходное фото или видеоизображение.

На сегодняшний день существует три вида дипфейков:

- подмена звукового сопровождения;
- подмена фотоизображения;
- подмена видеоизображения.

При этом все виды дипфейков могут сочетаться и создавать единый подменный образ. Примером бесплатного российского программного обеспечения (далее – ПО) отечественного производства является DeepFaceLab, причем компания предоставляет разработку бесплатно. ПО способно как создавать дипфейки, так и распознавать их [2]. Данный факт подтверждает доступность данной технологии для населения и, в том числе, для злоумышленников.

В процессе создания дипфейков используется технология ИИ
и нейросетей, это позволяет воссоздавать мимические движения лица, аудиоданные и, с развитием технологии, отличить дипфейк от реального изображения становится все сложнее, швы и нечеткости изображения все более детально прорабатываются и совершенствуются.

Как пример, компания Vera voice уже подменивает аудио так, что распознать подмену без помощи сторонних программ становится невозможно. По отдельным фрагментам воссоздаются голоса известных певцов и поэтов своего времени [6]. Нейросеть даже может генерировать новые композиции лучше человека, изучая музыку классических или современных композиторов. Например, в будущем злоумышленник сможет продать «затерянную» симфонию известного композитора.

В данной ситуации верно высказывание М. А. Желудкова: «Подобные технологии в условиях удалённого доступа могут быть использованы для оформления подложных товарно-денежных операций, изменения доказательств по реальным уголовным делам. Если сегодня такие программы пока ещё недостаточно совершенны, то пройдет небольшой промежуток времени, и технология дипфейков с открытым кодом создаст серьёзные трудности в идентификации аудио- и видеоинформации в интернет-пространстве. В этом случае

увеличится количество мошеннических действий, где от имени руководства или собственников предприятий будут поступать указания на перевод денежных средств или продажу активов, проведение по телефону банковских операций и др.» [3, с. 66-67].

В сентябре 2021 г. злоумышленники использовали дипфейк с лицом основателя ПАО «Тинькофф банк», предлагая перейти по ссылке с регистрацией, обещая клиентам 50% выгоду к вложенной сумме денежных средств. Однако целью злоумышленника было хищение персональных данных и денежных средств. Благодаря несовершенству технологии на данном этапе, дипфейк можно было распознать.

Другой пример: использование поддельного лица и голоса исполнительного директора компании Dbrain. Злоумышленник создал рекламное видео от лица Д. Мацкевича и призывал, вкладывая средства, получить высокий доход от инвестиций.

В предыдущих случаях факты недостоверности были обнаружены еще в процессе просмотра, однако, в феврале 2021 г. в сервисе TikTok был использован дипфейк Тома Круза, видео с которым за месяц собрало 11 млн. просмотров. На инцидент отреагировали даже некоторые зарубежные издания, отметив свои опасения насчет развития технологий. Данный пример не был отрицательного характера, однако подобное использование дипфейков знаменитостей может иметь злой умысел и оказывать влияние на общество, и даже быть оружием в руках враждующих государств. При этом угроза возникает как для лиц, на которых воздействует дипфейк, так и для лиц, чьи данные использовались для создания данного образа.

14 декабря 2023 г. был использован даже дипфейк Президента Российской Федерации. Во время прямой линии В.В. Путин смог поговорить со своим «двойником», представившимся студентом Санкт-Петербургского государственного университета.

Во II квартале 2021 г. было украдено 3 млрд. руб. посредством 236,9 тыс. денежных переводов без согласия клиентов. Чаще всего для своих целей преступники используют поддельные интернет-ресурсы или телефонные звонки. Но с появлением возможности создавать дипфейки стали возрастать факты совершения вымогательства денежных средств через шантаж лиц, подверженных манипуляции, используя откровенные изображения и даже видео, созданные нейросетью.

Как отмечает компания Positive technologies, в 2023 дипфейки все чаще используются как оружие пропаганды на фоне геополитической обстановки [1]. Данная проблема требует ужесточения правового регулирования и контрольных (надзорных) мероприятий. В настоящий момент в нормативном правовом регулировании отсутствует определение «дипфейк» или «нейростеть», однако они активно используются.

На текущий момент существует единственный нормативный правовой акт, регулирующий данную сферу, Федеральный закон от 24.04.2020 № 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 и внесении изменений в статьи 6 и 10 Федерального закона «О персональных данных» (дата обращения: 11.02.2024), однако в нем есть информация только об ИИ.

1. Необходимо нормативно закрепить термин «дипфейк».

2. Ввести отдельную юридическую ответственность за использование дипфейков для распространения заведомо ложных сведений, несущих вред обществу либо отдельным лицам, и (или) сведений порочащих честь и достоинство лица, чьи данные были использованы в процессе создания образа.

3. Ввести запрет на использование материалов умерших лиц для создания дипфейков без разрешения родственников. Следует нормативно закрепить право наследования цифрового образа члена семьи в рамках главы 63 Гражданского кодекса Российской Фе-

дерации (часть третья) от 26.11.2001 № 146-ФЗ (ред. от 24.07.2023) (с изм. и доп., вступ. в силу с 04.08.2023) и усложнить процедуру использования дипфейков, созданных на основе данных материалов.

Бесспорно, развитие нормативной правовой базы для регулирования дипфейков является сложным процессом. Однако использование и распространение дипфейков может нанести как угрозу Национальной безопасности и финансовой системе, так и оказывать негативное влияние на лиц, чьи данные были использованы в процессе создания дипфейка [4].

Источники и литература

1) Аналитическая статья компании Positive Technologies // Специалисты Positive Technologies выделили угрозы 2023 года: кибершпионаж, двойное вымогательство и двукратный рост атак на телеком [Электронный ресурс]. - Режим доступа: https://goo.su/LQp7nC (дата обращения: 11.02.2024)
2) Гарифуллин И. М. Использование нейросетей для выявления мошеннеческих транзакций // Инновационная наука. 2021. № 3. С. 30–32.
3) Желудков М. А. Обоснование необходимости адаптации деятельности правоохранительных органов к условиям цифровой трансформации преступной среды // Lex russica. 2021. № 4 (173). С. 63–70.
4) Иванов В. Г., Игнатовский Я. Р. Deepfakes: перспективы применения в политике и угрозы для личности и национальной безопасности // Вестник Российского университета дружбы народов. Серия: Государственное и муниципальное управление. 2020. № 4. С. 379–386
5) Капитонова Е. А. Шантаж «нюдсами» и смежные деяния: проблемы уголовно-правовой квалификации // Уголовное право. 2021. № 6. С. 19–27.
6) DNS Клуб // Нейросеть Vera Voice точно имитирует голоса людей [Электронный ресурс]. - Режим доступа: https://club.dns-shop.ru/digest/22282-neiroset-vera-voice-tochnoimitiruetgolosa-ludei (дата обращения: 11.02.2024)