DOI: 10.17803/1729-5920.2024.216.11.021-031

Яо Ли
Восточно-китайский университет политических наук и права
г. Шанхай, Китайская Народная Республика

Использование технологии «дипфейк» в Китае: проблемы правового регулирования и пути их решения

Резюме. Ядром технологии «дипфейк» является генеративно-состязательная сеть, построенная на комбинации двух нейронных сетей: генеративная сеть (сеть G) создает образцы, дискриминативная сеть (сеть D) старается отличить правильные («подлинные») образцы от неправильных. Сети G и D конкурируют друг с другом тысячи или даже миллионы раз, пока сеть G не улучшит свою производительность, так что сеть D уже не сможет отличить настоящие данные от поддельных. С развитием больших данных и технологий машинного обучения сценарий применения технологии «дипфейк» постепенно изменился: от создания звуковых моделей и имитации текста до глубокой подделки видео. Долгое время изображения, измененные с помощью традиционных Photoshop-технологий и других, легко распознавались, технология «дипфейк» изменила эту ситуацию, усложнив выявление подделок. Будучи важной технологической инновацией в области искусственного интеллекта, технология «дипфейк» широко используется в различных сферах жизни общества, создавая огромную прикладную ценность. Однако любая технология — это обоюдоострый меч. Применение технологии «дипфейк» несет большую угрозу личной конфиденциальности, безопасности имущества и даже национальной безопасности. Для того чтобы найти баланс между технологическими инновациями и предотвращением и контролем рисков, страны по всему миру активно изучают различные пути управления. В статье описаны основные риски, которые несет в себе современная технология «дипфейк», приведен обзор правового регулирования в данной области в Китае и предложен эффективный путь решения проблем.

Ключевые слова: дипфейк; генеративная состязательная сеть; личная жизнь; мошенничество; национальная безопасность; правовое регулирование; законодательство Китая

Для цитирования: Яо Ли. Использование технологии «дипфейк» в Китае: проблемы правового регулирования и пути их решения. *Lex russica.* 2024. Т. 77. № 11. С. 21–31. DOI: 10.17803/1729-5920.2024.216.11.021-031

The Use of Deepfake Technology in China: Problems of Legal Regulation and Ways to Solve Them

Yao Li
East China University of Political Science and Law
Shanghai, People's Republic of China

Abstract. The core of the deepfake technology is based on a generative-adversarial network built on a combination of two neural networks: a generative network (network G) creates samples, a discriminative network (network D) tries to distinguish correct («genuine») samples from incorrect ones. Networks G and D compete with each other thousands or even millions of times until network G improves its performance. Thus, the network D will no longer be able to distinguish real data from fake data. With the development of big data and machine learning

technologies, the scenario for using deepfake technology has gradually changed from creating sound models and imitating text to deep video forgery. For a long time, images modified using traditional Photoshop and other technologies were easily recognized. Deepfake technology changed this situation, making it more difficult to identify fakes. As an important technological innovation in the field of artificial intelligence, deepfake technology is widely used in various areas of society, creating enormous applied value. However, any technology is a double-edged sword. The use of deepfake technology poses a great threat to personal privacy, property security and even national security. In order to find a balance between technological innovation and risk prevention and control, countries around the world are actively exploring various ways to manage. The paper describes the main risks posed by modern deepfake technology, provides an overview of legal regulation in this area in China and offers an effective way to solve problems.

Keywords: deepfake; generative adversarial network; personal life; fraud; national security; legal regulation; Chinese law

Cite as: Yao Li. The Use of Deepfake Technology in China: Problems of Legal Regulation and Ways to Solve Them. *Lex russica.* 2024;77(11):21-31. (In Russ.). DOI: 10.17803/1729-5920.2024.216.11.021-031

Введение

Термин «дипфейк» (deepfake) происходит из области искусственного интеллекта, использующего алгоритмы ИИ deep learning («глубинное обучение») и fake («подделка»). Это основанная на ИИ технология синтеза изображений, аудио и видео, а также генерации текста. С помощью технологии «дипфейк» можно заменить, синтезировать или наложить целевое изображение и аудио/видео, текст и т.д. на другое изображение и аудио/видео, текст и т.д., чтобы создать более реалистичные поддельные изображения, текст и аудио/видео, чем традиционная технология синтеза.

Реализация этой технологии в основном опирается на новейшие алгоритмы искусственного интеллекта, включая сеть кодер-декодер (Encoder-Decoder), конволюционную нейронную сеть (Convolutional neural network), генеративную состязательную сеть (Generative adversarial network), архитектуру Pix2Pix, циклическую генеративную состязательную сеть (CycleGAN), рекуррентную нейронную сеть (Recurrent neural network). Среди них генеративная состязательная сеть — одна из наиболее широко используемых генеративных моделей, она включает генеративную сеть и дискриминативную сеть, эти две нейросетевые модели противостоят друг другу: генеративная сеть отвечает за генерацию синтетических данных с высокой степенью сходства, а дискриминативная сеть отвечает за сравнение бывших синтетических данных и реальных данных, чтобы определить подлинность[1].

Генеративная сеть, в частности, генерирует данные для создания поддельных звуков, изображений и т.д.; дискриминативная сеть берет на себя задачу идентификации поддельных данных, сравнения оригинальных звуков и изображений с поддельными звуками и изображениями и определения, какие из них настоящие. Генеративная сеть постоянно корректирует свой метод генерации в соответствии с обратной связью от дискриминативной сети; цикл повторяется, обе сети учатся друг у друга и противостоят друг другу, пока генеративная сеть не создаст звуки и изображения, которые практически невозможно определить как настоящие или поддельные. Китайские ученые считают, что «всё более совершенный алгоритм GANS может полностью поддерживать автоматический синтез и подделку образцов изображений, аудио и видео, а индивидуальный голос, микровыражения, движения тела и другие биологические маркеры и поведенческие особенности могут быть пересажены для достижения эффекта подделки настоящего»[2].

Коды и алгоритмы технологии «дипфейк» размещены в открытом доступе на технических сайтах, ими могут пользоваться и обмениваться технические специалисты и компьютерные энтузиасты, а алгоритмические команды постоянно обновляются в процессе реального использования, чтобы сделать эффект машинного обучения более качественным. Пользователи,

[1] *Цао Сюлянь.* Исследование состояния развития технологии обнаружения дипфейков // Технология и применение сетевой безопасности. 2022. № 5. С. 49–51.

[2] *Мао Нин, Ян Хуэй.* Нормативная дилемма технологии «дипфейк» и ее правовой ответ // Чанбай. 2021. № 5. С. 94–101.

обладающие профессиональными знаниями, могут создавать поддельные аудио- и видеоматериалы, изучая открытый код веб-сайтов, что создает серьезную угрозу сетевой безопасности.

В то же время с непрерывным развитием больших данных и технологий машинного обучения сценарий применения технологии «дипфейк» постепенно перешел от создания одной звуковой модели и имитации текста к глубокой подделке видео, ускоряя подрывные разработки интеллектуальных технологий. «Дипфейк» является продуктом развития технологий искусственного интеллекта, а киберпреступления, основанные на технологии «дипфейк», представляют собой «технологический риск». Хотя сама по себе технология нейтральна, ее применение может привести к совершенно разным последствиям из-за различий в ценностях и целях пользователей[3]. Лица с разными ценностями используют одну и ту же технику, но могут совершать совершенно разные действия. Поэтому мы должны своевременно обращать внимание на угрозы и риски, возникающие при использовании технологии «дипфейк», и находить правовые решения.

Риски и угрозы, связанные с технологией «дипфейк»

С развитием больших данных и машинного обучения дипфейк-контент становится всё ближе к реальному, и отличить подлинник от фальшивки трудно не только человеку, но и искусственному интеллекту[4]. Стремительное развитие технологии «дипфейк» открывает множество возможностей для социального развития, но широкомасштабное злоупотребление ею также порождает угрозы в области личной жизни, сетевой безопасности, национальной безопасности и в других аспектах.

1. Угроза личной жизни. Информация о лице, задействованная в технологии подмены лиц ИИ, относится к персональной биометрической информации, и неправильное использование информации о лице несет в себе риск нарушения безопасности личной информации граждан. Кроме того, без согласия лиц использование их изображения для AI face-swapping и публичного распространения и даже скандальное, клеветническое поведение подозреваются в нарушении права на изображение, права на частную жизнь. Если в процессе присутствуют такие факторы, как «злонамеренная клевета», «злонамеренное редактирование», «использование оскорбительных слов», которые делают социальную оценку репутации, заслуг и т.д. других людей явно негативной, это может представлять собой нарушение права людей на репутацию.

Возьмем в качестве примера программное обеспечение Deepnude, которое обучает алгоритмы искусственного интеллекта на основе большой базы данных реальных фотографий обнаженной натуры и генерирует фотографии обнаженных жертв на основе глубокого обучения и технологии GAN. Пользователям достаточно загрузить любую фотографию женщины, и программа автоматически «снимает с нее одежду» в течение 30 секунд[5]. В июне 2024 г. Бюро сетевой безопасности Министерства общественной безопасности КНР раскрыло ряд типичных случаев использования технологии «дипфейк», подчеркнув серьезную социальную опасность, вызванную злоупотреблением технологиями ИИ со стороны недобросовестных элементов. В одном из случаев подозреваемый использовал программное обеспечение искусственного интеллекта для «одноключевой обработки фотографий жертвы», создав около 7 000 непристойных фотографий и продавая их через Сеть[6]. Глубоко подделанные порнографические видеоролики подвергали жертв «виртуальной угрозе сексуального насилия», что оказывало глубокое психологическое воздействие на человека.

2. Риск нарушения имущественных прав. С одной стороны, такой вид риска сосредоточен на взломе различных типов систем распозна-

---

[3] *Лю Цзе.* Разведывательный риск технологии «дипфейк» и ее контрмеры // Вестник Полицейского колледжа Цзянси. 2022. № 1. С. 32–38.

[4] *Цай Шилинь.* Техническая логика и правовые изменения дипфейков // Политика и право. 2020. № 3. С. 131–140.

[5] *Сюй Яньпин.* Этические размышления о технологии «дипфейк» — на примере программы для стриптиза в один клик Deepnude // Компьютерная эпоха. 2021. № 11. С. 118–121.

[6] URL: https://baijiahao.baidu.com/s?id=1803729989604660419&wfr=spider&for=pc (дата обращения: 02.07.2024).

вания лиц. С развитием технологий 3D-печати и 3D-масок постоянно появляются новости об успешном взломе платежных систем с распознаванием лиц[7]. Лицо гражданина является ключом к базе данных в платежных приложениях, правонарушители подделывают реалистичные биометрические данные лица в реальном времени. Атаки на платформы онлайн-кредитования, супермаркеты и разблокировка мобильных телефонов, которые имеют низкий коэффициент информационной безопасности, наносят прямой материальный ущерб.

С другой стороны, преступники используют лазейки для захвата камеры мобильного телефона, активацию фотокамеры, манипуляцию выражением лица и другие технологии «дипфейк» для синтеза видео, клонируют голос жертвы и осуществляют мошенничество в отношении родственников или руководителей жертвы через телефонные звонки, онлайн-видеочаты и т.д.

В 2021 г. Бюро общественной безопасности провинции Шаньдун объявило о нескольких случаях мошенничества, совершенных с использованием дипфейков. Например, бухгалтеру компании поступил телефонный звонок от руководителя с просьбой немедленно перевести 20 000 юаней поставщику и одновременно пришло электронное письмо с инофрмацией о переводе. Благодаря очень реалистичному акценту руководителя в телефонном разговоре бухгалтер беспрекословно выполнил перевод[8].

По сообщениям гонконгских СМИ, полиция Гонконга (Китай) обнаружила, что мошенники использовали технологию «дипфейк» с помощью искусственного интеллекта, чтобы успешно имитировать изображения и голоса руководителей британской компании, выдавать себя за нескольких человек на онлайн-встречах и обмануть финансовых сотрудников на 200 млн гонконгских долларов, используя видеоролики компании на YouTube и материалы СМИ, полученные из других общедоступных источников[9].

Отчет о безопасности искусственного интеллекта за 2024 г., опубликованный компанией Qianxin, показывает, что в 2023 г. число случаев мошенничества с использованием технологии «дипфейк» вырастет на 3 000 %[10].

3. Национальная безопасность. Технологии искусственного интеллекта могут создавать голосовые модели на основе публичных интервью или иных записей голоса жертвы, генерируя аудио, чрезвычайно похожее на настоящий голос персонажа, а содержание поддельного аудио может быть разработано в соответствии с потребностями пользователя[11]. «Дипфейк» как перспективная сетевая технология обладает такими характеристиками, как мощное распространение, высокая реалистичность и простота эксплуатации, и легко используется преступниками для подрыва международных отношений, подстрекательства к терроризму и обострения национальных и социальных конфликтов.

В частности, исходя из политических, религиозных и экономических требований, преступники выбирают в качестве мишени важных политических деятелей или даже национальных лидеров, синтезируют их аватары и голоса в политические видеоролики враждебных сепаратистов, фабрикуют, редактируют и пересылают ложные сообщения и новости, созданные программным обеспечением, в попытке создать политические слухи, подстрекать к подрыву государственной власти, опорочить имидж национальных лидеров. В результате этого возникает риск нарушения социальной безопасности и финансового порядка, страдает имидж страны в международных отношениях, провоцируются международные геополитические конфликты.

Сетевая информация, основанная на технологии «дипфейк», может представлять угрозу политической безопасности стран и даже изменить политический процесс некоторых государств с точки зрения национального управле-

---

7 URL: https://www.xianjichina.com/special/detail_435812.html (дата обращения: 03.07.2024).

8 URL: https://baijiahao.baidu.com/s?id=1738851566361078810&wfr=spider&for=pc (дата обращения: 02.07.2024).

9 URL: https://baijiahao.baidu.com/s?id=1790106358293240290&wfr=spider&for=pc (дата обращения: 03.07.2024).

10 URL: https://www.qianxin.com/threat/reportdetail?report_id=311 (дата обращения: 01.07.2024).

11 *Чжан Юаньтин.* Правовое регулирование злоупотребления дипфейками в эпоху искусственного интеллекта // Теория : ежемесячник. 2022. № 9. С. 118–130.

12 *Лю Гочжу.* «Дипфейк» и национальная безопасность: перспектива, основанная на общей концепции национальной безопасности // Дайджест социальных наук. 2022. № 6. С. 65–67.

ния[12]. Виды преступлений, о которых идет речь, в основном включают подстрекательство к расколу страны (п. 2 ст. 103 УК КНР), пропаганду терроризма, экстремизма и подстрекательство к осуществлению террористической деятельности (ст. 120.3 УК КНР), а также подстрекательство к подрыву государственной власти (ст. 105 УК КНР), предоставлениюе заведомо ложной информации о противнике в военное время (ст. 377 УК КНР), создание слухов и нарушение морального духа армии в военное время (ст. 378 УК КНР), создание слухов и запутывание общественности в военное время (ст. 433 УК КНР)[13].

Правовое регулирование технологии «дипфейк» в Китае и его недостаточность

Дипфейк в Китае может привести к нарушению права физического лица на изображение, на защиту личной информации, ущербу репутации, имущественным и другим правам, основанным на нормах Гражданского кодекса КНР[14] и Закона о защите персональных данных[15], а преступления, которые могут представлять собой диффамацию, мошенничество, кражу, распространение непристойных материалов и умышленное распространение ложной информации, также могут быть охвачены системой уголовного права. В отношении интернет-платформ, распространяющих дипфейк-контент, провайдеров услуг глубокого синтеза и производителей глубоко подделанного контента, помимо Закона о кибербезопасности[16], действует Закон о безопасности данных[17], существуют также Положения об экологичном управлении сетевым информационным контентом[18], Положения об управлении глубоким синтезом информационных услуг Интернета[19], Положения об управлении сетевыми аудио- и видеоинформационными услугами[20] и другие требования по управлению платформами.

Положения об экологичном управлении сетевым информационным контентом, опубликованные 1 марта 2020 г., устанавливают требования к производителям контента: в ст. 6 сказано, что «не должен производиться контент, оскорбляющий или клеветнический по отношению к другим или нарушающий честь, частную жизнь и другие законные права и интересы других», а в ст. 7 указано, что «не должен производиться контент, содержащий сексуальные намеки, сексуальные авансы и т.д.». Положения могут использоваться для регулирования дипфейков или фальшивого контента, нарушающего права и интересы других, а также распространения порнографии.

В Положении об управлении сетевыми аудио- и видеоинформационными услугами, вступившем в силу в январе 2020 г., содержатся подробные правила для поставщиков сетевых аудио- и видеоинформационных услуг, например: «Использование новых приложений, основанных на новых технологиях, таких как глубокое обучение и виртуальная реальность, для производства, публикации и распространения неаутентичной аудио- и видеоинформации, должно быть обозначено на видном месте», использование технологий глубокого обучения и виртуальной реальности запрещено в области СМИ, что отражает уважение к правдивости новостей и предотвращает недоверие общественности к новостным СМИ. В статье 12 говорится о внедрении технологии идентификации и своевременном прекращении распространения нелегального контента, что свидетельствует об усилении идентификации и аудите контента в отношении платформ распространения информации. В настоящее время технология обнаружения дипфейков всё еще разрабатывается и совершенствуется, и в ближайшей перспективе

---

13 URL: https://flk.npc.gov.cn/detail2. html?ZmY4MDgxODE3OTZhNjM2YTAxNzk4MjJhMTk2NDBjOTI (дата обращения: 01.07.2024).

14 URL: https://www.gov.cn/xinwen/2020-06/01/content_5516649.htm (дата обращения: 01.07.2024).

15 URL: http://www.npc.gov.cn/npc/c2/c30834/202108/t20210820_313088.html (дата обращения: 01.07.2024).

16 URL: https://www.cac.gov.cn/2016-11/07/c_1119867116.htm (дата обращения: 01.07.2024).

17 URL: http://www.npc.gov.cn/npc/c2/c30834/202106/t20210610_311888.html (дата обращения: 01.07.2024).

18 URL: https://www.cac.gov.cn/2019-12/20/c_1578375159509309.htm (дата обращения: 01.07.2024).

19 URL: https://www.cac.gov.cn/2022-12/11/c_1672221949354811.htm (дата обращения: 01.07.2024).

20 URL: https://www.cac.gov.cn/2019-11/29/c_1576561820967678.htm (дата обращения: 01.07.2024).

не все онлайн-платформы смогут ее применять, поэтому данное положение имеет перспективный характер[21].

Положения об управлении глубоким синтезом информационных услуг Интернета, которые вступили в силу 10 января 2023 г., охватывают возможные негативные последствия технологии «дипфейк». В документе сформулирован ряд правил, которым должны следовать поставщики услуг глубокого синтеза, в частности: внедрение основной ответственности за информационную безопасность; разработка и раскрытие правил управления и конвенций платформы; проверка подлинности информации о реальной личности; усиление управления контентом глубокого синтеза; создание и совершенствование механизмов разглашения слухов.

Введение данного Положения знаменует собой тот факт, что глубокий синтез стал первым типом алгоритмических услуг, который был специально законодательно закреплен в области алгоритмического управления в Китае в связи с его важностью и высоким риском[22]. Более всеобъемлющее положение для поставщиков и пользователей услуг глубокого синтеза является одним из наиболее важных способов регулирования услуг глубокого синтеза и технологии «дипфейк»[23]. В целом Китай установил более полное регулирование контента, включающего «дипфейк», и быстро отреагировал на влияние технологии искусственного интеллекта с позитивным отношением, но всё еще есть некоторые детали, которые необходимо улучшить.

Во-первых, в Положении об управлении сетевыми аудио- и видеоинформационными услугами содержится требование к провайдерам, то есть организациям или частным лицам, предоставляющим населению сетевые аудио- и видеоинформационные услуги, и пользователям о том, что «использование новых приложений, основанных на новых технологиях, таких как глубокое обучение и виртуальная реальность, для производства, публикации и распространения неаутентичной аудио- и видеоинформации должно быть обозначено на видном месте», и такие фразы содержатся также в Положении об управлении глубоким синтезом информационных услуг Интернета, хотя в последнем содержатся более подробные положения о сценариях, в которых должна применяться маркировка. Но ни в одном из этих двух Положений не упоминается, что такое «заметная маркировка» либо «обозначение на видном месте», поэтому возможны различные трактовки.

Во-вторых, в законодательстве недостаточно четко определено нарушение, связанное с технологией «дипфейк». Например, статья 1024 ГК КНР и статья 6 Положения об экологичном управлении сетевым информационным контентом устанавливают, что ни одна организация или частное лицо не должны нарушать право на репутацию других лиц путем оскорбления или клеветы, однако на практике из-за анонимности пользователей, создающих дипфейк-контент, существует сложность в определении личности нарушителя, а также трудно определить каналы распространения и сферу распространения.

Статья 1027 ГК КНР устанавливает, что «если литературное или художественное произведение, опубликованное виновным лицом, содержит оскорбительное или клеветническое содержание, нарушающее право другого лица на репутацию путем описания реального или конкретного лица, потерпевший вправе требовать от виновного привлечения его к гражданско-правовой ответственности в соответствии с законом», но не предусматривает «использование средств информационных технологий для подделки содержания», что затрудняет подачу иска потерпевшим.

Пути предотвращения рисков и решения проблем

В Соединенных Штатах было принято законодательство, регулирующее технологии «дипфейк»: Конгресс принял Deepfakes Report Act of

[21] *Чу Ибинь.* Углубленный синтез подделок «нереального аудио и видео» и предложения по улучшению — с точки зрения Положений об управлении сетевыми аудио- и видеоинформационными услугами // Тяньшуйский институт государственного управления. 2021. № 6. С. 120–124.

[22] *Чжан Линьхань.* Обновление логики и итерация системы управления глубокого синтеза — китайский путь генеративного управления искусственного интеллекта, такого как ChatGPT // Юридическая наука (журнал Северо-Западного университета политики и права). 2023. № 3. С. 38–51.

[23] *Ма Лися, Гао Геге, Вэй Юньцзе, Хэ Жан.* Новый метод оценки социальных рисков на основе технологии «дипфейк» // Обзор менеджмента. 2022. № 10. С. 14–26.

2019[24], и 20 сентября 2023 г. Палата представителей передала в Подкомитет по управлению чрезвычайными ситуациями и технологиям H.R.5586 — Deepfakes Accountability Act[25], в котором предлагается усилить защиту от технологических атак и говорится о необходимости разработки технологий обнаружения и идентификации, а также контрмер. Компания Facebook (Meta[26]) потратила огромные средства на проведение конкурса Deepfake Detection Challenge совместно с Массачусетским технологическим институтом, Оксфордским университетом и другими колледжами и университетами, который получил широкую поддержку со стороны академического сообщества.

В 2018 г. Европейская комиссия созвала представителей крупнейших технологических компаний и индустрии онлайн-рекламы для разработки добровольных рамок отраслевого саморегулирования и опубликовала Свод практических правил по дезинформации (Code of Practice on Disinformation), согласно которому интернет-компании должны укреплять самоцензуру и бороться с ложным контентом в Сети у источника. Усиленный Свод практических правил по дезинформации был подписан и представлен 16 июня 2022 г. 34 подписантами, которые присоединились к процессу пересмотра Свода от 2018 г.[27]

Взятые вместе меры по управлению технологией «дипфейк», принятые Соединенными Штатами и Европейским Союзом, включают: а) регулирование посредством законодательства, разъяснение прав, обязанностей и юридической ответственности субъектов; б) усиление саморегулирования отрасли и требование к платформам выполнять свои обязательства по регулированию; в) усиление технологических исследований и разработок и управление технологией «дипфейк»[28].

На основе объединения опыта Соединенных Штатов и Европейского Союза Китай может принять следующие меры для дальнейшего противодействия рискам, связанным с технологией «дипфейк».

1. Поставщики услуг технологии «дипфейк» должны активно сотрудничать с государственными ведомствами, чтобы изучить законные сценарии применения технологии «дипфейк». В настоящее время основные интернет-платформы и ключевые инфраструктуры в странах мира монополизированы несколькими технологическими гигантами, а большой объем профессиональных и технических знаний и ресурсов, необходимых для управления технологией «дипфейк», не принадлежит государству, и все сложные задачи по управлению технологией «дипфейк» не могут быть полностью возложены на правительство, поэтому необходимо предоставить поставщикам услуг технологии «дипфейк» (далее — поставщики услуг) ограниченные полномочия по саморегулированию, чтобы они могли осуществлять саморегулирование под надзором и руководством государственных органов. В частности, поставщики услуг должны прежде всего провести самооценку, чтобы доказать, что используемые ими технические решения соответствуют законам, правилам и стандартам, и представить отчет о самооценке в государственный орган, а после получения сертификата органа они могут быть освобождены от ответственности.

Это может стимулировать поставщиков услуг вносить положительный вклад в соблюдение правовых норм. С другой стороны, поскольку объем цензурных обязанностей поставщиков услуг всё еще неясен, государственные органы должны как можно скорее определить сценарии неправильного использования технологии «дипфейк» и сформировать список, ограничивающий ее применение[29]. Поставщики услуг должны регулярно предоставлять органам обратную связь о результатах проверки дипфейк-контента, чтобы

---

[24] URL: https://www.congress.gov/bill/116thcongress/house-bill/3600/ (дата обращения: 29.06.2024).

[25] URL: https://www.congress.gov/bill/118th-congress/house-bill/5586/text (дата обращения: 01.07.2024).

[26] Решением Тверского районного суда города Москвы от 21.03.2022 компания «Мета Плэтформс, Инк.» (Meta Platforms, Inc.) признана экстремистской организацией и ее деятельность запрещена на территории Российской Федерации.

[27] URL: https://digital-strategy.ec.europa.eu/en/policies/code-practice-disinformation (дата обращения: 02.07.2024).

[28] *Чжан Тао.* Правовые риски «дипфейк» и ее регулирование в эпоху постправды // Электронное правительство. 2020. № 4. С. 91–101.

[29] *Лонь Кун, Ма Е, Чжу Цичао.* Вызовы дипфейков для национальной безопасности и ответные меры // Информационная безопасность и хранение тайны связи. 2019. № 10. С. 21–34.

динамично корректировать список ограничений и формировать механизм совместного управления. Правительство также должно поощрять платформы и организации к разработке отраслевых руководств и стандартов в рамках, разрешенных законами и правилами.

Согласно Закону КНР о стандартизации, отраслевой стандарт — это стандарт, совместно разработанный общественными организациями, созданными в соответствии с законом для удовлетворения потребностей рынка и инноваций и координации участников рынка, который принимается членами группы или добровольно принимается обществом в соответствии с положениями группы, а участниками разработки стандарта являются в основном ведущие организации, имеющие влияние в отраслевых областях[30].

2. Уточнение принципа информированного согласия. В настоящее время статья 1035 ГК КНР, статьи 41–42 Закона о кибербезопасности и статьи 13–15 Закона о защите персональных данных предусматривают правило информированного согласия, но в реальности оно является простой формальностью. Это связано с тем, что, с одной стороны, если пользователь не согласен с условиями предоставления персональной информации, он не сможет воспользоваться услугой, а поставщик услуг связывает «согласие» с «использованием», так что поставщик услуг может получить персональные данные пользователя только на основании его одноразового согласия.

Для этого необходимо создать механизм динамического информированного согласия, когда согласие пользователя требуется на всех этапах оказания услуги. Кроме того, когда пользователь просматривает дипфейк-контент или ставит лайк, комментирует или совершает другие действия с таким контентом, должно автоматически всплывать окно предупреждения, чтобы проинформировать пользователя о том, что это дипфейк-контент. Пользователи могут отозвать свое согласие при необходимости, и поставщики услуг должны предоставить возможность сделать это на видном месте, а также не должны скрывать кнопку отзы́ва.

3. Совершенствование методов выявления дипфейков или фальшивого контента. Подтверждение подлинности информации приобретает исключительную важность в наши дни, когда технологии «дипфейк» становятся всё более изощренными[31]. Следует поощрять научно-технические предприятия к развитию технологии обнаружения дипфейков, а также к исследованию и разработке технологии различения реального и поддельного контента.

Законодательство Китая может перенять опыт американского Deepfakes Accountability Act и усовершенствовать концепцию маркировки «заметным образом», включая использование голосовых или письменных заявлений с подробным описанием степени подделки в дипфейк-контенте, на которую можно указать в течение всего процесса или в начале и конце аудио- и видеопроизведений, а также использование технологии, которую нельзя изменить или заменить. Например, блокчейн для постоянной записи метаданных. Любой дипфейк-контент может быть записан и отслежен в режиме реального времени с помощью технологии блокчейн, что обеспечивает бóльшую прозрачность и возможность отслеживания такого контента. В частности, блокчейн может использоваться для добавления временны́х меток к оригинальным видео, аудио, фотографиям и другим материалам, создавая неизменяемые записи метаданных в распределенном реестре, так что весь процесс производства и распространения контента может быть отслежен в режиме реального времени. Кроме того, могут быть разработаны алгоритмические модели для идентификации поддельных лиц и т.д. путем распознавания того, согласовано ли поддельное лицо с остальными кадрами изображения, или обнаружения отсутствия физиологических сигналов в сгенерированном контенте.

4. Раскрытие фальшивых данных с помощью реальных. С точки зрения расширения возможностей контроля человека над своей информацией важно иметь надежные доказательства подлинности или подделки контента. Технология «дипфейк» часто пытается исказить то, куда человек перемещается, как он себя ведет и что говорит. Надежность доказательств обеспечивают реальные траектории движения человека. В данном случае речь идет о

[30] URL: http://www.npc.gov.cn/zgrdw/npc/xinwen/2017-11/04/content_2031446.htm (дата обращения: 02.07.2024).

[31] *Цао Цзяньфэн, Фан Линьман.* Исследование рисков и контрмер в отношении технологии «дипфейк» // Информационная безопасность и хранение тайны связи. 2019. № 2. С. 89–97.

технологии, которая отслеживает и записывает жизнь человека на основе его местоположения, коммуникаций и действий, помогая выявлять и разоблачать дипфейки, повышая контроль над цифровым контентом, связанным с ним.

Носимые умные устройства и умные IoT-устройства могут собирать эти данные для подтверждения подлинности информации с помощью мониторинга жизни и здоровья, чтобы помочь доказать целостность данных человека и в конечном счете — подлинность его частного контента. Такие данные могут быть зашифрованы, сохранены и удалены владельцем для предотвращения нарушения конфиденциальности.

Заключение

Технология «дипфейк» нейтральна по отношению к ценностям, но поведение человека в отношении технологии не является нейтральным, и эффекты добра и зла не симметричны. Технологический прогресс должен не только использовать интеллект машин, но и демонстрировать мудрость человека в освоении технологий. В последние годы технология «дипфейк» в кино и телевидении, СМИ, образовании и других областях демонстрирует большое количество положительных применений, наряду с появлением спроса на использование всех видов коммерческих продуктов, количество контента с использованием технологии «дипфейк» значительно увеличилось. Однако аудио- и видеоматериалы, созданные в результате злонамеренного использования этой технологии, демонстрируют большую разрушительную силу, нанося ущерб репутации и имуществу частных лиц и предприятий и даже представляя угрозу для общества и национальной безопасности.

Несмотря на то что онлайн-платформы уже принимают меры по проверке информации, из-за недостатков правового регулирования платформы часто допускают создание и распространение дипфейков, также недостатки конструкции системы платформы могут привести к неспособности фильтровать или отслеживать глубокие фейки. Комплексное управление технологией «дипфейк» должно включать в себя сотрудничество между государством и предприятиями, подробные законодательные положения и меры по стимулированию разработки предприятиями технологий для выявления дипфейков.

СПИСОК ЛИТЕРАТУРЫ

Лонь Кун, Ма Е, Чжу Цичао. Вызовы дипфейков для национальной безопасности и ответные меры // Информационная безопасность и хранение тайны связи. 2019. № 10. С. 21–34. [龙坤，马钺，朱启超：《深度伪造对国家安全的挑战及应对》，载《信息安全与通信保密》2019年第10期，第21-34页。].

Лю Гочжу. «Дипфейк» и национальная безопасность: перспектива, основанная на общей концепции национальной безопасности // Дайджест социальных наук. 2022. № 6. С. 65–67. [刘国柱：《深度伪造与国家安全：基于总体国家安全观的视角》，载《社会科学文摘》2022年第6期，第65-67页。].

Лю Цзе. Разведывательный риск технологии «дипфейк» и ее контрмеры // Вестник Полицейского колледжа Цзянси. 2022. № 1. С. 32–38. [刘杰：《"深度伪造"技术的情报风险及其应对》，载《江西警察学院学报》2022年第1期，第32-38页。].

Ма Лися, Гао Гэге, Вэй Юньцзе, Хэ Жан. Новый метод оценки социальных рисков на основе технологии «дипфейк» // Обзор менеджмента. 2022. № 10. С. 14–26. [马丽霞，高格格，魏云捷，赫然：一个新的关于深度伪造技术的社会风险评估方法，载《管理评论》2022年第10期，第14-26页。].

Мао Нин, Ян Хуэй. Нормативная дилемма технологии «дипфейк» и ее правовой ответ // Чанбай. 2021. № 5. С. 94–101. [毛宁，杨会：《深度伪造技术的监管困境及其法律应对》，载《长白学刊》2021年第5期，第94-101页。].

Сюй Яньпин. Этические размышления о технологии «дипфейк» — на примере программы для стриптиза в один клик Deepnude // Компьютерная эпоха. 2021. № 11. С. 118–121. [徐燕萍：《"深度伪造"技术的伦理反思———以一键脱衣软件 "Deepnude" 为例》，载《计算机时代》2021年第11期，第118-121页。].

Цай Шилинь. Техническая логика и правовые изменения дипфейков // Политика и право. 2020. № 3. С. 131–140. [蔡士林：《"深度伪造"的技术逻辑与法律变革》，载《政法论丛》2020年第3期，第131-140页。].

Цао Сюлянь. Исследование состояния развития технологии обнаружения дипфейков // Технология и применение сетевой безопасности. 2022. № 5. С. 49–51. [曹秀莲：《深度伪造检测技术发展现状研究》，载《网络安全技术与应用》2022年第5期，第49-51页。].

Цао Цзяньфэн, Фан Линьман. Исследование рисков и контрмер в отношении технологии «дипфейк» // Информационная безопасность и хранение тайны связи. 2019. № 2. С. 89–97. [曹建峰, 方玲曼：《"深度伪造"的风险及对策研究》, 载《信息安全与通信保密》2020年第2期, 第89-97页。].

Чжан Линьхань. Обновление логики и итерация системы управления глубокого синтеза — китайский путь генеративного управления искусственного интеллекта, такого как ChatGPT // Юридическая наука (Журнал Северо-Западного университета политики и права). 2023. № 3. С. 38–51. [张凌寒：《深度合成治理的逻辑更新与体系迭代——ChatGPT 等生成型人工智能治理的中国路径》, 载《 法律科学（西北政法大学学报）》2023年第3期, 第38-51页。].

Чжан Тао. Правовые риски технологии «дипфейк» и ее регулирование в эпоху постправды // Электронное правительство. 2020. № 4. С. 91–101. [张涛：《后真相时代深度伪造的法律风险及其规制》, 载《电子政务》2020年第4期, 第91-101页。].

Чжан Юаньтин. Правовое регулирование злоупотребления дипфейками в эпоху искусственного интеллекта // Теория : ежемесячник. 2022. № 9. С. 118–130. [张远婷：《人工智能时代"深度伪造"滥用行为的法律规制》, 载《理论月刊》2022年第9期, 第118-130页。].

Чу Ибинь. Углубленный синтез подделок «нереального аудио и видео» и предложения по улучшению — с точки зрения Положений об управлении сетевыми аудио- и видеоинформационными услугами // Тяньшуйский институт государственного управления. 2021. № 6. С. 120–124. [储宜彬：《深度伪造合成"非真实音视频"的规制现状与完善建议——以〈网络音视频信息服务管理规定〉为切入点》, 载《天水行政学院学报》2021年第6期, 第120-124页。].

REFERENCES

Cai Shilin. Technical logic and legal changes «deepfake». *Politics and Law*. 2020;3:131-140. [蔡士林：《"深度伪造"的技术逻辑与法律变革》, 载《政法论丛》2020年第3期, 第131-140页。]. (In Chinese).

Cao Jianfeng, Fang Linman. Deepfake Risk and Countermeasures Study. *Information Security and Secrecy.* 2019;2:89-97. [曹建峰, 方玲曼：《"深度伪造"的风险及对策研究》, 载《信息安全与通信保密》2020年第2期, 第89-97页。]. (In Chinese).

Cao Xulian. Research on the state of development of deepfake detection technology. *Network Security Technology & Application*. 2022;5:49-51. [曹秀莲：《深度伪造检测技术发展现状研究》, 载《网络安全技术与应用》2022年第5期, 第49-51页。]. (In Chinese).

Chu Yibin. Advanced synthesis of fake «unrealistic audio and video» and suggestions for improvement — from the point of view of «Rules for managing network audio and video information services». *Tianshui Institute of Public Administration*. 2021;6:120-124. [储宜彬：《深度伪造合成"非真实音视频"的规制现状与完善建议——以〈网络音视频信息服务管理规定〉为切入点》, 载《天水行政学院学报》2021年第6期, 第120-124页。]. (In Chinese).

Long Kun, Ma Ye, Zhu Qichao. National Security Deepfake Challenges and Responses. *Information Security and Secracy*. 2019;10:21-34. [龙坤, 马钺, 朱启超：《深度伪造对国家安全的挑战及应对》, 载《信息安全与通信保密》2019年第10期, 第21-34页。]. (In Chinese).

Liu Jie. Intelligence risk of deepfake technology and its countermeasures. *Journal of Jiangxi Police College*. 2022;1:32-38. [刘杰：《"深度伪造"技术的情报风险及其应对》, 载《江西警察学院学报》2022年第1期, 第32-38页。]. (In Chinese).

Liu Guozhu. *«Deepfake»* and national security: a perspective based on a common concept of national security. *Social Science Digest*. 2022;6:65-67. [刘国柱：《深度伪造与国家安全：基于总体国家安全观的视角》, 载《社会科学文摘》2022年第6期, 第65-67页。]. (In Chinese).

Ma Lixia, Gao Gege, Wei Yunjie, He Jean. New method for assessing social risks based on deepfake technology. *Management Review.* 2022;10:14-26. [马丽霞, 高格格, 魏云捷, 赫然：一个新的关于深度伪造技术的社会风险评估方法, 载《管理评论》2022年第10期, 第14-26页。]. (In Chinese).

Mao Ning, Yang Hui. The regulatory dilemma of deepfake technology and its legal response. *Changbai*. 2021;5:94-101. [毛宁, 杨会：《深度伪造技术的监管困境及其法律应对》, 载《长白学刊》2021年第5期, 第94-101页。]. (In Chinese).

Xu Yanping. Ethical reflections on deepfake technology — using Deepnude as an example of a one-click striptease program. *The Computer Age*. 2021;11:118-121. [徐燕萍：《"深度伪造"技术的伦理反思———以一键脱衣软件 "Deepnude» 为例》, 载《计算机时代》2021年第11期, 第118-121页。]. (In Chinese).

Zhang Linhan. Logic Update and Iteration of Deep Synthesis Control System — China's Way of Control over Generative AI like ChatGPT. *Journal of Northwest University of Political Science and Law.* 2023;3:38-51. [张凌寒：

《深度合成治理的逻辑更新与体系迭代——ChatGPT 等生成型人工智能治理的中国路径》，载《 法律科学（西北政法大学学报）》2023年第3期，第38-51页。]. (In Chinese).

Zhang Tao. Legal risks of deepfake technology and its regulation in the post-truth era. *E-government*. 2020;4:91-101. [张涛：《后真相时代深度伪造的法律风险及其规制》，载《电子政务》2020年第4期，第91-101页。]. (In Chinese).

Zhang Yuanting. Legal regulation of deepfake abuse in the era of artificial intelligence. *Theory: Monthly Journal*. 2022;9:118-130. [张远婷：《人工智能时代"深度伪造"滥用行为的法律规制》，载《理论月刊》2022年第9期，第118-130页。]. (In Chinese).

ИНФОРМАЦИЯ ОБ АВТОРЕ

Яо Ли, доктор юридических наук, Институт докторантуры международного права при Восточно-китайском университете политических наук и права
д. 555, Лунюйань ул., г. Шанхай 201620, Китайская Народная Республика
yaozaihenmang@mail.ru

INFORMATION ABOUT THE AUTHOR

Yao Li, Dr. Sci. (Law), Institute of Doctoral Studies in International Law, East China University of Political Science and Law, Shanghai, People's Republic of China
yaozaihenmang@mail.ru

Материал поступил в редакцию 16 мая 2024 г.
Статья получена после рецензирования 26 августа 2024 г.
Принята к печати 15 октября 2024 г.

Received 16.05.2024.
Revised 26.08.2024.
Accepted 15.10.2024.