Deepfake: новая реальность

Бычинская Анна Александровна, Северо-Западный институт управления Российской академии народного хозяйства и государственной службы при Президенте Российской Федерации; факультет экономики и финансов (Санкт-Петербург, Российская Федерация)
студентка 4-го курса бакалавриата;
e-mail: abychinskaya-20@edu.ranepa.ru
Иванова Софья Игоревна, Северо-Западный институт управления Российской академии народного хозяйства и государственной службы при Президенте Российской Федерации; факультет экономики и финансов (Санкт-Петербург, Российская Федерация)
студентка 4-го курса бакалавриата;
e-mail: sivanova-20-01@edu.ranepa.ru

Научный руководитель:
Гетман Анастасия Геннадьевна, Северо-Западный институт управления Российской академии народного хозяйства и государственной службы при Президенте Российской Федерации; кафедра таможенного администрирования (Санкт-Петербург, Российская Федерация)
заведующая кафедрой, кандидат экономических наук, доцент;
e-mail: getman-ag@ranepa.ru

Аннотация
Целью научной работы является исследование влияния технологии deepfake на современное общество. Для достижения цели были поставлены следующие задачи: сформулировать определение deepfake и представить классификацию; сформулировать предложения для снижения отрицательных влияний deepfake на современное общество. Новизна исследования заключается в выявлении проблемы влияния deepfake на современное общество и выработке подходов к ее решению. Методология работы включает в себя систематизацию теоретических данных, классификацию, анализ совокупности факторов и прогнозирование. В результате авторами было дано определение технологии deepfake, выделены положительные и отрицательные аспекты технологии, а также обозначены дальнейшие направления технологических исследований и инноваций в данной области.
Ключевые слова: технология deepfake, права интеллектуальной собственности, нарушение приватности, цифровые данные, искусственный интеллект, информационная безопасность

Deepfake: New Reality

Anna A. Bychinskaya, North-Western Institute of Management, Russian Academy of National Economy and Public Administration under the President of the Russian Federation; Faculty of Economics and Finance (Saint Petersburg, Russian Federation)
BA student;
e-mail: abychinskaya-20@edu.ranepa.ru.
Sofya I. Ivanova, North-Western Institute of Management, Russian Academy of National Economy and Public Administration under the President of the Russian Federation; Faculty of Economics and Finance (Saint Petersburg, Russian Federation)
BA student;
e-mail: sivanova-20-01@edu.ranepa.ru

Academic Supervisor:
Anastasia G. Getman, North-Western Institute of Management, Russian Academy of National Economy and Public Administration under the President of the Russian Federation; Department of Customs Administration (Saint Petersburg, Russian Federation)
Head of the Department, PhD of Economic Sciences, Associate Professor;
e-mail: getman-ag@ranepa.ru

Abstract
The objective of the academic paper is to investigate the impact of deepfake technology on modern society. To achieve the goal the following aims were set: to formulate the definition of deepfake and present the classification; to state proposals for reducing the negative influence of deepfake on modern society. The novelty of the study consists in identifying the issue of deepfake influence on modern society and developing approaches to its solution. The methodology of the work involves systematizing theoretical data, classifying, analyzing a complex of factors and forecasting. As a result, the authors provide a definition of deepfake technology, highlight its positive and negative aspects, as well as identify further trends of technological research and innovation in this area.

Keywords: deepfake technology, intellectual property rights, privacy violation, digital data, artificial intelligence, information security

Современное общество характеризуется стремительным развитием цифровых технологий не только в сфере глобальных экономических процессов, но и в сфере моделей социальных и общественных взаимодействий на различных уровнях. Совершенствованию deepfake-технологий способствует процесс быстрой модернизации программного обеспечения и объединения профессионального сообщества программистов.

Новые технологические возможности, связанные с deepfake, несут серьезные риски. Список угроз начинается с отдельного человека или группы пользователей и заканчивается государственными интересами.

Технология deepfake развивается быстрее, чем законодательная база. Этим объясняются существующие правовые проблемы, нарушение авторских и интеллектуальных прав собственности. Именно эти факты и обуславливают актуальность проблемы использования deepfake.

Объект исследования — deepfake-технологии.

Предметом исследования является deepfake как специфическая форма развития современного общества.

Цель исследования — изучение влияния технологии deepfake на современное общество.

Задачи исследования:

1) сформулировать определение deepfake и провести классификацию;

2) смоделировать предложения для снижения отрицательных влияний deepfake на современное общество.

Новизна исследования заключается в формулировании определения поставленной проблемы и классификации ее видов, а также в выработке подходов к ее решению.

Методология работы включает в себя систематизацию теоретических данных, классификацию, анализ совокупности факторов и прогнозирование.

Методологическую базу исследования составляют публикации российских и зарубежных специалистов в области интеллектуальной собственности (А. Г. Гетман, В. Г. Иванов) и нейронных сетей (Кэти Уорр).

1. ТЕХНОЛОГИЯ DEEPFAKE

Deepfake — синтетический контент, в котором человек в оригинальном фото, аудио или видео заменяется другим при помощи искусственного интеллекта. Термин заимствован из английского языка и объединяет два понятия: «deep learning — глубокое обучение» и «fake — подделка». Для создания deepfake используются нейронные сети глубокого обучения, такие как глубокие сверточные нейронные сети (CNN) и генеративно-состязательные сети (GAN), чтобы анализировать и синтезировать изображения и видео [1, с. 31]. Для создания аудио используются технологии, клонирующие голос человека и генерирующие фразы, которые он на самом деле не произносил. Это может включать в себя обучение моделей голосового синтеза на основе образцов голоса целевой личности [7, с. 89–91].

Точную дату создания технологии deepfake установить невозможно. Популярность она приобрела в 2017 г., когда в Интернете появился обработанный порнографический фильм, в котором якобы главную роль сыграла голливудская актриса Галь Гадот, наиболее известная по роли Чудо-женщины.

В последние несколько лет технология deepfake получила широкое распространение среди пользователей. Большинство пользователей воспринимают deepfake как новую форму онлайн-юмора и развлечений, а не как способ обмануть или шантажировать людей. Технология deepfake с каждым годом становится совершенней, и всё труднее отличить настоящие видеоролики от поддельных. Страдают не только знаменитости и публичные деятели, но и обычные пользователи. Карьера и жизнь человека могут быть скомпрометированы и полностью разрушены, а поддельные видеоролики с мировыми лидерами могут привести к дипломатическим конфликтам и даже военным столкновениям. На законодательном уровне проблематично учесть все тонкости развития и реализации современных технологий, поэтому социальные сети пытаются самостоятельно внедрить технологии, которые будут ограничивать использование технологии deepfake.

2. КЛАССИФИКАЦИЯ ТЕХНОЛОГИИ DEEPFAKE

Технологию deepfake можно классифицировать по методам создания, автору, степени полезности, наносимому ущербу.

Классификация deepfake по методам создания:

1. Изменение лиц (face swap) — замена лица одного человека лицом другого человека, применение фильтров, изменяющих внешность. Это отличительное свойство социальных сетей и приложений для видеоконференций.

2. Кукловодство (puppeteering) — рендеринг характерных движений человека при помощи создания 3D-модели лица и тела.

3. Синхронизация губ (lip-syncing) — технология создания определенных мимических движений рта под аудиозапись. Благодаря этому происходит моделирование речи человека, даже при условии, что он никогда этого не говорил.

4. Клонирование голоса (voice cloning) — технология создания синтетического голоса при помощи искусственного интеллекта.

5. Синтез изображений — технология создания визуальных компонентов при помощи метода компьютерного зрения, глубокого обучения и GAN (генеративно-состязательных сетей). В результате создается компьютерное изображение, которое никогда не существовало в реальной жизни.

6. Генерация текста — технология создания текста при помощи рекуррентных нейронных сетей или GAN.

7. Подрисовывание — технология создания дополнительных элементов на исходном изображении [5, с. 16–17].

Классификацию deepfake по авторам можно разделить на пять основных групп:

- объединение любителей;
- политические игроки;
- злоумышленники;
- коммерческие организации;
- средства массовой информации.

Большое количество deepfake обусловлено наличием базового программного обеспечения для простого процесса их создания и распространения [5, с. 17].

Классификация deepfake по степени полезности:

- Полезные и безвредные deepfake создаются с развлекательной целью и распространены в сфере коммерции, здравоохранения, образования, коммуникации между человеком и машиной, видеоконференций, творческого выражения пользователей.
- Вредоносные deepfake создаются с целью мошенничества, распространения ложной информации или нанесения вреда на индивидуальном и коллективном уровнях, на уровне организаций и в масштабе страны.

Классификация deepfake по наносимому ущербу:

- психологический;
- финансовый;
- социальный.

Таблица

Классификация deepfake по типу наносимого ущерба[1]

Психологический ущерб	Финансовый ущерб	Социальный ущерб
Клевета	Вымогательство	Ущерб экономической стабильности
Шантаж	Мошенничество	Ущерб системе правосудия
Издевательство	Манипулирование финансовым рынком	Ущерб научной системе
Дискредитация	Кража личных данных	Ущерб политическому строю
	Репутационный ущерб	Ущерб международным отношениям
		Ущерб национальной безопасности

3. ПОЛОЖИТЕЛЬНЫЕ АСПЕКТЫ ИСПОЛЬЗОВАНИЯ ТЕХНОЛОГИИ DEEPFAKE

При правильном использовании технология deepfake может быть полезна в различных сферах жизни. Благодаря такому контенту можно расширить границы реальности и визуализировать любую идею.

1. Помощь людям с ограниченными возможностями

При помощи deepfake можно создавать специальные инструменты, которые дополняют основные способности человека — слух и зрение. Например, компания Microsoft создала приложение SeeingAI, которое помогает слабовидящим или незрячим людям превращать визуальный мир в звуки и тем самым взаимодействовать с окружающим миром.

2. Искусство и творчество

Благодаря deepfake люди могут взаимодействовать с популярными личностями, которых давно нет в живых. Так, в 2019 г. Музей Сальвадора Дали во Флориде открыл выставку Dalí Lives и при помощи искусственного интеллекта оживил художника. Посетители увидели большое количество роликов, где Дали рассказывает о своей жизни и о погоде, а также с ним можно было сделать селфи.

[1] Источник: составлено (разработано) авторами.

3. Прогнозирование и визуализация

При помощи deepfake люди могут увидеть происходящее как миллионы лет назад, так и то, что происходит в соседних странах. Например, в 2017 г. в MIT был создан алгоритм, который переносил военные разрушения на фотографии мирных городов. Алгоритм обучался на реальных фотографиях сирийских городов до и после войны. Проект был создан для того, чтобы пользователи, кого не коснулась война, увидели реальный масштаб разрушений инфраструктуры городов.

4. ОТРИЦАТЕЛЬНЫЕ АСПЕКТЫ ИСПОЛЬЗОВАНИЯ ТЕХНОЛОГИИ DEEPFAKE

Технология deepfake имеет много положительных моментов. Однако нужно принимать во внимание тот факт, что существуют отрицательные аспекты этой технологии, которые побуждают разрабатывать законодательную базу для предотвращения и борьбы с возникновением аспектов.

1. Политические манипуляции и обман

Технология deepfake получила широкое распространение в ситуациях отсутствия нормативного регулирования. Так она подрывает существующие основы демократии, дестабилизирует доверие граждан к органам государственной власти и правовым институтам, вводит в заблуждение избирателей, а также ведет к угрозе общественной безопасности и обострению социальных разногласий [2, с. 96].

В 2019 г. в Интернете опубликовали deepfake-видео со спикером Палаты представителей Конгресса США Нэнси Пелоси. Автор ролика с помощью технологий искусственного интеллекта изменил речь Пелоси так, что она плохо выговаривала слова. Пользователи, посмотревшие видео, посчитали, что политик находится в состоянии алкогольного опьянения. Всё это переросло в большой скандал, и лишь спустя некоторое время было доказано, что речь Пелоси была сгенерирована компьютером. Несмотря на это, Нэнси Пелоси была скомпрометирована, и было утрачено доверие к ней как к политической фигуре [3, с. 267].

2. Нарушение приватности

Создание реалистичных видео с участием людей не всегда сопровождается получением разрешения на использование образа человека. Это ведет к нарушению приватности, а также способствует распространению компрометирующей информации. В ряде стран уже вступили в силу законы, направленные на борьбу с подобным контентом [4, с. 382].

В 2019 г. в Интернет было выгружено более 15 000 deepfake-видео. Из них 96 % было порнографического содержания, а в 98 % использовались лица знаменитых женщин.

3. Инструмент для киберпреступлений

Еще один отрицательный аспект технологии связан с созданием аудио- или видеозаписей с целью обмана пользователей. Благодаря быстрой обучаемости искусственный интеллект может воспроизвести необходимый акцент и манеру речи.

Генеральный директор британской компании перевел почти 243 000 долларов на венгерский банковский счет после того, как мошенник использовал технологию deepfake, чтобы имитировать голос главы страховой компании Euler Hermes Group[2].

Deepfake является современной угрозой целостности информации и безопасности общества. С помощью разработанных программных решений, образовательных курсов для пользователей и применения законодательных мер можно значительно снизить риск негативных последствий [6, с. 61–62].

---

[2] Мошенник подделал голос CEO и украл $243 тыс. при помощи технологии deepfake // Inc. Russia. – URL: https://incrussia.ru/news/deepfake-moshennik-ukral-243-tys/ (дата обращения: 01.03.2024).

5. ПРОБЛЕМЫ МОЛОДЕЖИ, СВЯЗАННЫЕ С DEEPFAKE

Несмотря на то что молодые люди являются активными и продвинутыми пользователями цифровых приложений, именно эта группа находится в зоне повышенной опасности. Провести границу между реальным и фальшивым становится всё труднее из-за огромного количества поддельных текстов, аудио- и видеоматериалов. Ориентирование пользователя в цифровой среде усложняется, подрывается доверие к информации, возникают ситуации, когда нет достоверных фактов, на которые можно опереться при анализе действительности и формировании картины мира [5, с. 22].

Важно упомянуть, что большая часть инструментов для создания deepfake-контента общедоступна, достаточно проста в использовании и может быть освоена пользователями, не имеющими специальных навыков.

6. РАЗВИТИЕ И БУДУЩЕЕ DEEPFAKE

Развитие приложений привело к значительному улучшению качества технологии deepfake. Реалистичность фальшивого контента стала правдоподобной настолько, что даже опытные специалисты испытывают затруднения при распознавании фальсифицированного продукта.

Потенциал технологии deepfake заключается в том, что ее можно применять в различных областях. Однако наряду с потенциалом возникают угрозы, которые побуждают продолжать исследования в области обеспечения безопасности и этичности использования deepfake.

В настоящее время исследования в области deepfake продолжаются. Их цель — разработка надежных методов обнаружения синтетического контента, а также снижение риска злоупотребления использования deepfake. Инновации в рассматриваемой области включают в себя разработку алгоритмов искусственного интеллекта для проверки подлинности контента.

Алгоритмы искусственного интеллекта для распознавания deepfake могут состоять из анализа артефактов фотографии, видео или аудио, неестественных изменений в мимике лица или отклонений в речи. Обучение искусственного интеллекта проводится на большом объеме данных. Для повышения эффективности процесса обучения можно создавать репозитории, в которых будут храниться deepfake-данные.

Разработка систем для цифрового подтверждения подлинности контента тесно связана с известной технологией блокчейн. За счет высокого уровня безопасности блокчейна возможность фальсификации и изменения информации уменьшается, что положительно сказывается на доверии к цифровым данным.

Очевидным направлением развития сферы deepfake является совершенствование правовых мер. На данный момент в российском законодательстве нет понятия deepfake. В Гражданском кодексе РФ ст. 152.1 регулирует вопросы, связанные с охраной изображения гражданина. 23 января 2024 г. Государственная Дума РФ приняла в первом чтении законопроект об оборотных штрафах за утечку данных. К второму чтению предлагается добавить положения об использовании технологии deepfake (поправки в законодательство о персональных данных, которые введут ответственность за несанкционированное использование голоса и изображения людей)[3].

Изучив литературу и научные статьи, предлагаем способы минимизации отрицательных аспектов технологии deepfake:

[3] Протокол № 185 заседания Государственной Думы // Информация Совета Государственной Думы. – URL: https://sozd.duma.gov.ru/download/5CF63C7E-4FDF-4637-A409-2AD62FF9B032 (дата обращения: 01.03.2024).

- создание межведомственной экспертной группы по противодействию распространению дезинформации и другим противоправным действиям с использованием deepfake-технологии;
- разработка терминологии и общепринятой классификации deepfake;
- разработка системы маркировки синтетического контента в социальных сетях;
- введение дисциплин в образовательных учреждениях, а также открытие курсов по формированию рационального поведения в информационной среде;
- создание образовательной платформы для обучения пользователей разной степени подготовленности;
- повышение правовой грамотности граждан в сфере использования синтетического контента.

В заключение данного исследования можно сделать вывод, что технология deepfake обладает большим потенциалом и в положительных, и в отрицательных аспектах. Развитие и будущее этой технологии будут определяться сразу несколькими факторами: техническими достижениями, образованностью информационного общества и правовым регулированием использования технологии. На сегодняшний день стоят задачи по разработке защитных механизмов, направленных на минимизацию негативных последствий использования технологии deepfake.

Литература

1. *Getman A. G.*, *Ling Y.* The Deepfake Technology: Threats or Opportunities for Customs // Administrative consulting. – 2023. – No. 4 (172). – P. 30–36. EDN TWWJOY

2. *Дремлюга Р. И.*, *Моисейцев В. В.*, *Парин Д. В.*, *Романова Л. И.* Национальное правовое регулирование использования и распространения реалистичных аудиовизуальных поддельных материалов (deepfake): опыт Китая // Азиатско-Тихоокеанский регион: экономика, политика, право. – 2022. – № 4. – С. 91–104. – URL: https://cyberleninka.ru/article/n/natsionalnoe-pravovoe-regulirovanie-ispolzovaniya-i-rasprostraneniya-realistichnyh-audiovizualnyh-poddelnyh-materialov-deepfake (дата обращения: 13.09.2023).

3. *Желудков М. А.* Изучение влияния новых цифровых технологий на детерминацию мошеннических действий (технология Deepfake) // Развитие наук антикриминального цикла в свете глобальных вызовов обществу. – 2021. – С. 262–270. – URL: https://www.elibrary.ru/download/elibrary_46479126_99144112.pdf (дата обращения: 14.09.2023).

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. – URL: https://cyberleninka.ru/article/n/deepfakes-perspektivy-primeneniya-v-politike-i-ugrozy-dlya-lichnosti-i-natsionalnoy-bezopasnosti (дата обращения: 13.09.2023).

5. *Игнатьев А. Г.*, *Курбатова Т. А.* Дипфейки в цифровом пространстве: основные международные подходы к исследованию и регулированию // Центр глобальной ИТ-кооперации. – 2023. – 52 с. – URL: https://cgitc.ru/upload/iblock/23f/67eesk6aig8owuk7brnrgwospccquaec.pdf (дата обращения: 21.10.2023).

6. *Киселёв А. С.* О необходимости правового регулирования в сфере искусственного интеллекта: дипфейк как угроза национальной безопасности // Московский юридический журнал. – 2021. – № 3. – С. 54–64. – URL: https://cyberleninka.ru/article/n/o-neobhodimosti-pravovogo-regulirovaniya-v-sfere-iskusstvennogo-intellekta-dipfeyk-kak-ugroza-natsionalnoy-bezopasnosti (дата обращения: 14.09.2023).

7. *Уорр Кэти*. Надежность нейронных сетей: укрепляем устойчивость ИИ к обману. Серия: Бестселлеры O'Reilly. – СПб.: Питер, 2021. – 272 с. – URL: https://ibooks.ru/bookshelf/377030/reading (дата обращения: 13.09.2023). – ISBN 978-5-4461-1676-8.

Для цитирования:

Бычинская А. А., Иванова С. И. Deepfake: новая реальность // Новизна. Эксперимент. Традиции (Н.Экс.Т). – 2024. – Т. 10. – № 2 (26). – С. 32–39.