УДК 342

ИСТОРИЯ РАЗВИТИЯ ТЕХНОЛОГИИ ДИПФЕЙК

М. Д. ВОЛЫНЕЦ
(Представлено: К. Д. САВИЦКАЯ)

Данная статья посвящена истории развития технологии дипфейк, начиная с ее зарождения в сфере компьютерной графики и искусственного интеллекта до современных применений. В статье рассматриваются ключевые этапы развития технологии, включая появление первых алгоритмов генерации изображений и видео, развитие глубокого обучения и появление специализированных нейронных сетей для создания дипфейков.

Актуальность исследуемой темы обусловлена тем, что в современном мире технологии дипфейк стали широко распространенными и вызывают все больший интерес у исследователей, общественности и законодателей. Дипфейки могут использоваться как для развития современных технологий, так и для манипулирования общественным мнением, подрывая доверие к информации. Кроме того, дипфейки поднимают множество этических вопросов, поэтому необходимость в разработке правовых норм, регулирующих использование дипфейков, очевидна.

Технологии становятся все более сложными и все более взаимосвязанными. Автомобили, самолеты, медицинское оборудование, финансовые операции и другое, – основа их функционирования базируется на компьютерном программном обеспечении. Из-за чего кажется, что их труднее понять, а в некоторых случаях и сложнее контролировать. Изучение того, как сделать такие технологии, как искусственный интеллект или Интернет вещей, «объяснимыми», стало отдельной областью исследований.

За последние 10 лет самые эффективные системы искусственного интеллекта, такие как распознаватели речи на смартфонах или новейший автоматический переводчик Google, были созданы на основе метода, называемого «глубоким обучением».

«Глубокое обучение» – это новых подход к искусственному интеллекту, называемого нейронными сетями. Нейронные сети были впервые предложены в 1944 году Уорреном Маккалоу и Уолтером Питтсом, двумя исследователями из Чикагского университета, которые перешли в Массачусетский технологический институт в 1952 году в качестве основателей того, что иногда называют первым отделом когнитивных наук [1].

Нейронные сети были основной областью исследований как в нейробиологии, так и в информатике до 1969 года, когда, согласно знаниям, в области компьютерных наук, они были уничтожены математиками Массачусетского технологического института Марвином Мински и Сеймуром Папертом, которые год спустя стали содиректорами новая лаборатория искусственного интеллекта Массачусетского технологического института [1].

Затем эта технология возродилась в 1980-х годах во многом благодаря возросшей вычислительной мощности графических чипов.

«Существует мнение, что научные идеи немного похожи на эпидемии вирусов», – говорит Томазо Поджо, профессор Юджина МакДермотта кафедры мозга и когнитивных наук Массачусетского технологического института [1]. «Очевидно, существует пять или шесть основных штаммов вирусов гриппа, и, очевидно, каждый из них возвращается с периодом около 25 лет. Люди заражаются, у них развивается иммунный ответ, и поэтому они не заражаются в течение следующих 25 лет. А потом появляется новое поколение, готовое заразиться тем же штаммом вируса. В науке люди влюбляются в идею, воодушевляются ею, забивают до смерти, а затем получают прививку – им это надоедает. Значит, идеи должны иметь одинаковую периодичность!» [1].

И точно так же, как наше понимание управления технологиями развивается новыми и интересными способами, так же развивается и наше понимание социальных, культурных, экологических и политических аспектов новых технологий. Мы осознаем, как проблемы, так и важность определения всего спектра способов, которыми технологии меняют наше общество, того, как мы хотим, чтобы эти изменения выглядели, и какие инструменты мы должны попытаться повлиять на эти изменения и направить их.

Дипфейк – синтетические медиа, включая изображения, видео и аудио, созданные с помощью технологии искусственного интеллекта (ИИ), которые изображают нечто, чего не существует в реальности, или события, которые никогда не происходили [2].

В 1997 году Кристоф Бреглер, Мишель Ковелл и Малкольм Слейни опубликовали статью, в которой была разработана инновационная, поистине уникальная программа, которая по существу автоматизировала то, что могли делать некоторые киностудии. Программа Video Rewrite может синтезировать новую лицевую анимацию из аудиовыхода. Он основывался на более старых работах, которые интерпретировали лица, синтезировали звук из текста и моделировали губы в трехмерном пространстве, но был первым, кто объединил все это и убедительно анимировал [2].

Это одна из важнейших работ в разработке дипфейков. Фактически, многие из распространенных сегодня видеоэффектов, включенных в такие программы, как Premiere Pro или Final Cut, используют усовершенствованные алгоритмы из этой статьи.

Авторы отмечают, что эту систему «можно использовать для дубляжа фильмов, телеконференций и создания спецэффектов», хотя этого еще предстоит увидеть.

Компьютерное зрение все глубже проникло в мир распознавания лиц. Разработки в этой области привели к радикальным улучшениям в таких вещах, как отслеживание движений, которые делают современные дипфейки более убедительными.

Термин «дипфейк» объединяет в себе «глубокий», взятый из технологии глубокого обучения искусственного интеллекта (тип машинного обучения, включающий несколько уровней обработки), и «фейковый», указывающий на то, что контент не является реальным [2]. Этот термин стал использоваться для синтетических медиа в 2017 году, когда модератор Reddit создал субреддит под названием «deepfakes» и начал публиковать видеоролики, в которых использовалась технология подмены лиц для вставки изображений знаменитостей в существующие порнографические видеоролики.

В основе дипфейков лежит искусственный интеллект, а именно разновидность машинного обучения, называемая глубоким обучением. Глубокое обучение использует многоуровневые нейронные сети для анализа и интерпретации больших наборов данных [1]. Дипфейки – это гиперреалистичные цифровые произведения, обычно видео- или аудиозаписи, созданные с использованием методов искусственного интеллекта и машинного обучения. Они включают в себя наложение существующих изображений и видео на исходные изображения или видео с использованием техники глубокого обучения, в частности генеративно-состязательных сетей («GAN»).

GAN включают в себя две нейронные сети – генератор и дискриминатор. Генератор создает изображения/видео, а дискриминатор сравнивает их с реальными кадрами. Благодаря итеративным процессам генератор становится все лучше и лучше создает реалистичные подделки. Эта технология позволяет создавать убедительно реалистичный фейковый контент, в котором люди говорят или делают то, чего на самом деле никогда не делали.

Дипфейки создаются с использованием значительного объема данных, включая фотографии, звуковые фрагменты или видео целевого человека. Поскольку больше данных может создать более убедительные дипфейки, наиболее распространенными целями становятся знаменитости, политики и другие общественные деятели [3].

Появление приложений и онлайн-сервисов, где пользователи могут создавать дипфейки с минимальными техническими знаниями, еще больше снизило планку для входа.

ИИ тренируется, используя собранные данные, чтобы понимать тонкости и нюансы целевого человека, такие как черты лица, тон и интонации голоса, а также движения [4]. Затем обученная модель накладывает изображение целевого человека на исходную фотографию или видео. Видеоманипуляции выполняются покадрово для создания реалистичного видео.

Несмотря на кажущееся волшебство новейших инструментов искусственного интеллекта, почти все дипфейки по-прежнему легко и быстро разоблачаются.

Таким образом, исследование, проведенное в статье на основании изучения исторического развития технологии дипфейк, позволяет сделать вывод о подтверждении актуальности вопроса. Данные технологии охватывают масштабное количество сфер жизни деятельности, как правовых, так и социальных. Вышеприведенные примеры указывают на необходимость автоматизации технологии дипфейк.

ЛИТЕРАТУРА

1. Кирова Л. М., Макаревич М. Л. Правовые аспекты использования нейронных сетей // Инновационная экономика: перспективы развития и совершенствования. – 2023. – 50–75 с.
2. Пользовательский контроль визуального контента в интернете : сб. науч.ст. / Масленкова Н. А – Минск, 2024. – 180−196 с.
3. 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. – 2024. – 360−380 с.
4. Сорокин В. Н., Вьюгин В. В., Тананыкин А. А. Распознавание личности по голосу: аналитический обзор // Информационные процессы. – 2012. – № 1. − С. 1–30.