МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ

РОССИЙСКОЙ ФЕДЕРАЦИИ

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ

ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ

«ПЯТИГОРСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»

ИНСТИТУТ МЕЖДУНАРОДНЫХ ОТНОШЕНИЙ

Кафедра журналистики, медиакоммуникаций и связей с общественностью

Фоменкова Анастасия Алексеевна

**DEEPFAKE И ЖУРНАЛИСТИКА: УГРОЗА ДОВЕРИЮ ИЛИ НОВЫЙ ИНСТРУМЕНТ?**

Реферат, выполненный в рамках изучения дисциплины
«Системы искусственного интеллекта»

Направление подготовки: 42.03.02 – Журналистика

Направленность (профиль): «Международная журналистика»

Пятигорск – 2025

**Оглавление**

[**ВВЕДЕНИЕ** 3](#_Toc204806136)

[**1. DEEPFAKE: ТЕХНОЛОГИЧЕСКАЯ ОСНОВА И ЭВОЛЮЦИЯ** 4](#_Toc204806137)

[**2. DEEPFAKE КАК УГРОЗА ДОВЕРИЮ К ЖУРНАЛИСТИКЕ И ИНФОРМАЦИОННОЙ БЕЗОПАСНОСТИ** 5](#_Toc204806138)

[**3. DEEPFAKE КАК ПОТЕНЦИАЛЬНЫЙ ИНСТРУМЕНТ В ЖУРНАЛИСТСКОЙ ПРАКТИКЕ** 6](#_Toc204806139)

[**4. БОРЬБА С ВРЕДОНОСНЫМИ DEEPFAKE И ОБЕСПЕЧЕНИЕ ДОСТОВЕРНОСТИ** 8](#_Toc204806140)

[**5. ЭТИЧЕСКИЕ И ПРАВОВЫЕ ДИЛЕММЫ ИСПОЛЬЗОВАНИЯ DEEPFAKE В ЖУРНАЛИСТИКЕ** 9](#_Toc204806141)

[**ЗАКЛЮЧЕНИЕ** 11](#_Toc204806142)

[Список литературы 12](#_Toc204806143)

## **ВВЕДЕНИЕ**

Технологии глубоких подделок (deepfake) стремительно развиваются на стыке нейросетей и мультимедиа. Deepfake – это гиперреалистичный синтетический медиа-контент (изображения, аудио, видео или текст), созданный с помощью методов глубокого обучения и практически неотличимый от оригинала[[1]](#footnote-1)[[2]](#footnote-2). Появление масс-медиа и социальных сетей сделало возможным мгновенную публикацию подобных материалов в медиапространстве. Это породило дискуссию о том, подрывают ли дипфейки доверие к журналистике или могут стать её новым инструментом. Уже в 2022 году в украинских СМИ появилось фальшивое «обращение Зеленского» к гражданам, призывающее сдать оружие, – это оказался дипфейк, опровергнутый самим президентом. Хотя подделка была низкого качества, она создала волну недоверия и внимания общественности. В эпоху политической поляризации и «постправды» такие события подчеркивают, насколько важно различать реальные и ложные сообщения[[3]](#footnote-3). В данной статье рассматривается эволюция дипфейков, их угроза для доверия к медиа и информационной безопасности, а также возможные положительные применения в журналистской практике. Особое внимание уделяется мерам противодействия вредоносным дипфейкам и этико-правовым аспектам их использования в СМИ[[4]](#footnote-4).

## **1. DEEPFAKE: ТЕХНОЛОГИЧЕСКАЯ ОСНОВА И ЭВОЛЮЦИЯ**

Глубокие подделки базируются на разработках в области нейросетей и машинного обучения. Ключевую роль здесь сыграли генеративные состязательные сети (GAN, Generative Adversarial Networks), предложенные И. Гудфеллоу в 2014 году, которые научились создавать новые изображения и звуки на основе большого объема данных. Спустя три года, в 2017 г., анонимный пользователь под ником «deepfakes» выложил в Интернет первое широко известное видео-подделку – порнофрагмент с «заменой лиц» голливудских актеров1. Интернет-сообщество быстро подхватило технологию, и вскоре были созданы тысячи подобных роликов. Нейросети обучаются на образцах реальных лиц и голосов, что позволяет генерировать реалистичное аудио- и видеоизображение. Так, например, технология Adobe Voco способна проанализировать образец голоса человека и воспроизвести его в цифровой форме, говоря новые тексты с тем же тембром и интонацией4. При этом, по оценкам экспертов, 90–95 % ранних дипфейков были сексуального характера, использовались для порочащих видеоподделок1. Сегодня deepfake охватывает не только видео – существуют синтетические голоса, «оживление» неподвижных изображений, а также генерация текста (например, чат‑боты нового поколения). Как отмечает исследователь А.В. Шадрин, «технология дипфейка основана на использовании искусственного интеллекта для создания реалистичных видео- и аудиоподделок»4. В целом, развитие технологий генеративного ИИ делает процесс создания дипфейков все проще и доступнее даже непрофессионалам. При этом эволюция идёт от грубых монтажей к почти неотличимым фальшивкам – в скором времени программы глубокого обучения будут генерировать видео высокого разрешения, практически неотличимые от реальных материалов.

## **2. DEEPFAKE КАК УГРОЗА ДОВЕРИЮ К ЖУРНАЛИСТИКЕ И ИНФОРМАЦИОННОЙ БЕЗОПАСНОСТИ**

Дипфейки представляют серьёзную угрозу доверия к СМИ и информационной безопасности. Поддельные видео и аудио способны распространять дезинформацию намного быстрее: человек, увидев впечатляющий фейк, склонен принять увиденное за достоверное, особенно если сообщение подкрепляет его убеждения[[5]](#footnote-5). Исследование философа Д. Фаллиса показывает, что такие подделки не только ведут к ложным убеждениям, но и ослабляют обоснование истинных знаний, подрывая легитимность реальных новостей3. Даже разоблаченное видео оставляет сомнения в публикуемой информации: «после того как хотя бы один раз лживая информация была представлена как подлинная, общество начинает сомневаться в честности СМИ»[[6]](#footnote-6).

Также дипфейки могут служить инструментом политической и психологической манипуляции. Например, они позволяют показать известных политиков в ситуациях, которых на самом деле не было: это не только подрывает доверие к лидерам, но и может кардинально изменить общественное мнение или исход выборов4. Классическим примером является поддельное обращение политика, которое затем использовалось в пропагандистских целях. Также опасна практика создания порочащих материалов: широко известны случаи, когда лица женщин без их согласия «монтировали» в порнографические видео, создавая ложную компрометирующую базу. Г. Беркман (организация Social Media Safety) указывает, что такие ролики могут служить инструментом шантажа и клеветы3.

Кроме того, дипфейки усиливают общую проблему информационной безопасности. По оценкам, фейковые новости обходятся мировой экономике десятками миллиардов долларов в год, а компании и государствам приходится вкладываться в защиту репутации, медиаграмотность и борьбу с фальсификациями. При этом алгоритмы распространяют информацию в фильтр-пузырях: разные группы людей получают поляризованные версии «правды», что углубляет кризис доверия4. Чем более достоверным становится контент, тем сложнее распознавать мошенничество: согласно исследованию, «любой массовый синтетический контент лишь хочет казаться неотличимым от настоящего»1, поэтому невооруженным глазом подделки нередко незаметны. Совокупно эти факторы создают атмосферу сомнений: когда люди перестают верить в то, что видят или слышат, фундамент современной журналистики – ответственность и проверяемость информации – оказывается под угрозой.

**3. DEEPFAKE КАК ПОТЕНЦИАЛЬНЫЙ ИНСТРУМЕНТ В ЖУРНАЛИСТСКОЙ ПРАКТИКЕ**

Несмотря на риски, технологии дипфейка можно использовать и в созидательных целях. Уже существует множество примеров, когда синтетические медиа применялись для образовательных или культурных задач, не связанных с манипуляцией фактами12. Так, историческая реконструкция с помощью искусственного интеллекта позволяет «оживлять» утраченную или недоступную информацию. Известен случай, когда один из первых дипфейков XIX века создали неизвестные художники: они наложили голову президента Авраама Линкольна на гравюру с вице-президентом Джоном Кэлхуном, подставив документы с актуальными после Гражданской войны надписями (см. рисунок)1. Подобная методика может применяться журналистами при подготовке исторических материалов или документальных фильмов. 

*Рис. 1. Исторический пример “дипфейка” середины XIX в.: слева – портрет Авраама Линкольна, созданный наложением его головы на оригинальную гравюру с вице-президентом Джоном Кэлхуном (справа). Такая техника иллюстрирует возможность «воссоздания» визуальных материалов.*

Кроме того, синтетические медиа могут быть полезны для анонимизации источников и создания инсценировок. Так, исследователи разрабатывают системы для мгновенного преобразования голоса человека в нейросетевой речевой аватар – например, проект Rochester University «Voice Makeovers» позволяет менять голос по возрасту, полу и диалекту в режиме реального времени[[7]](#footnote-7). Для журналистов подобные инструменты могут использоваться, когда требуется скрыть личность собеседника: рассказ очевидца или жертвы можно озвучить чужим искусственным голосом, сохранив эмоциональную окраску речи. Аналогично, глубокие синтетические изображения лиц (аватаров) позволят монтировать сцены или диалоги с участием исторических или удаленных персонажей – например, выступление старого журналиста, которого уже нет в живых, – что откроет новые формы повествования. Учёные также подчеркивают потенциал дипфейков в образовании и здравоохранении: созданные с помощью ИИ виртуальные персонажи могут выступать в ролях учителей или терапевтов, адаптированных под конкретные нужды2. Это показывает: технологии дипфейка сами по себе не обладают моралью – всё зависит от того, кто и как их использует.

**4. БОРЬБА С ВРЕДОНОСНЫМИ DEEPFAKE И ОБЕСПЕЧЕНИЕ ДОСТОВЕРНОСТИ**

Для противодействия опасным дипфейкам развивается целый комплекс мер. Технические решения включают детекторы подделок, цифровые метки и методы аутентификации контента. Сейчас существуют алгоритмы, анализирующие видео на предмет артефактов – например, нестабильность освещения или несовпадение движения губ и звука. Модуль Pindrop Pulse Inspect способен автоматически определять синтетические голоса в аудиозаписях6. В дополнение к этому разрабатываются системы цифровой водяной маркировки (watermarking): например, в ЕС и США обсуждаются стандарты JPEG Trust и законы, обязывающие наносить невидимые метки на сгенерированное ИИ содержимое. Эти метки позволят при воспроизведении определить, было ли изображение или видео создано ИИ, и предупредить потребителя о возможной недостоверности.

Политические и нормативные меры включают в себя принятие законов и регламентов. Социальные сети уже ввели правила по удалению «явных» дипфейков с платформ. В ряде стран уже есть нормы, предусматривающие гражданскую ответственность авторов намеренно вредных дипфейков3. В России готовятся проекты поправок к Уголовному кодексу, предусматривающие наказание (вплоть до лишения свободы) за злоупотребление дипфейками при клевете и мошенничестве. Такие законодательные инициативы призваны сдерживать злоумышленников, распространяющих опасные фейки. Кроме того, отмечается роль международных стандартов: недавно Европарламент рекомендовал обязательное маркирование ИИ-контента и внедрение меток для контроля происхождения материалов[[8]](#footnote-8).

Медиаобразование и процедурные подходы также играют критическую роль. Издатели и редакции должны внедрять внутренние протоколы проверки контента: проверять источники, использовать инструменты fact-checking, обучать журналистов новым навыкам распознавания синтетики. Широкую роль играет повышение медиаграмотности аудитории: обучающие кампании учат людей скептически относиться к визуальной информации и искать подтверждение в нескольких независимых источниках6. В конечном счёте, комплекс мер – от технологий до правового регулирования и образования – позволит минимизировать вред от дипфейков и сохранить доверие к СМИ78.

**5. ЭТИЧЕСКИЕ И ПРАВОВЫЕ ДИЛЕММЫ ИСПОЛЬЗОВАНИЯ DEEPFAKE В ЖУРНАЛИСТИКЕ**

Использование дипфейков в журналистике сопряжено с серьёзными этическими вопросами. С одной стороны, применение синтетических медиа ради развлечения или анонимности кажется оправданным, но с другой – всегда возникает риск введения аудитории в заблуждение. Главная этическая дилемма – должны ли журналисты признавать, что контент создан ИИ, или выдавать его за «реальный»? Согласно кодексу журналистской этики, СМИ обязаны предотвращать распространение манипулированных материалов. Если фейк был опубликован без пометки, редакция рискует потерять доверие аудитории и подвергнуться общественному порицанию6. Публикация «наполовину правдивого» ролика может привести к юридической ответственности за клевету и другие правонарушения, поскольку в таких случаях злоумышленно искажается образ человека или событие.

К тому же использование дипфейков поднимает вопросы согласия и приватности. Авторами законопроектов предлагается разрешать запись голоса или изображения человека только при его явном согласии. В практическом плане это означает: трансляция интервью с имитацией голоса источника должна быть предварительно согласована, а иначе – рассматриваться как фальсификация. Известен пример из международной практики: канадская телекомпания CBC изучала возможность анонимизации жертв преступлений с помощью AI-сгенерированных лиц и голосов. Специальная комиссия заключила, что такие материалы могут неправильно восприняться публикой и нарушить приватность, и в итоге от идеи отказались[[9]](#footnote-9). Этот кейс демонстрирует: этические риски глубоких фейков высоки, и редакциям необходимо вырабатывать чёткие стандарты использования подобных технологий.

Наконец, следует отметить юридическую неопределённость. Дипфейк находится на стыке технологий и права, поэтому во многих странах еще нет готовых норм. Это создаёт «правовой вакуум»: авторы постановок пока не всегда знают, какое наказание им грозит. В то же время законодатели спешат заполнить эту пустоту – во многих юрисдикциях (США, ЕС, Россия и др.) разрабатываются специальные поправки против дипфейков. Одна из трудностей – отмыть законодательство от избыточного регулирования, чтобы не подавлять инновации. Ведь по сути технологии дипфейка имеют двоякий характер: с одной стороны, они могут применяться во вред (дезинформация, шантаж), с другой – приносить общественную пользу (образование, креатив). Поэтому этический подход часто сводится к принципу «добросовестного использования»: журналисты должны ясно обозначать, когда материал является синтетическим, и не использовать дипфейки в дезинформационных целях6.

**ЗАКЛЮЧЕНИЕ**

Дипфейк – сложный феномен на пересечении технологий и массовых коммуникаций. С одной стороны, он действительно может подрывать доверие к журналистике и информационной безопасности, усложняя проверку фактов и усиливая распространение фальшивых новостей. С другой – синтетические медиа обладают творческим потенциалом: они расширяют возможности повествования и анонимизации источников, обогащают образовательные и культурные практики. В реальности глубокие фейки – это не «враг» или «друг» сами по себе, а инструмент, который может быть и опасным, и полезным. Ключевым фактором остается осознанность: информированность общества, ответственный подход журналистов и адекватное правовое регулирование. Только при сочетании технических мер по обнаружению дипфейков, юридических санкций за вредоносное использование и этических стандартов в медиасреде возможно сохранить высокий уровень доверия к СМИ. В перспективе полезность или вред дипфейков в журналистике будет определяться тем, насколько быстро общество научится отделять «реальность от симуляции», поддерживая прозрачность и ответственность в информационной сфере.

## **СПИСОК ЛИТЕРАТУРЫ**

1. Капитанов А.В., Линдре Ю.М. *Дипфейк: невинная технология для развлечения или угроза современному обществу?* // Российский совет по международным делам. – 2022.
2. Lundberg, E., Mozelius, P. The potential effects of deepfakes on news media and entertainment. AI & Soc 40, 2159–2170 – 2025.
3. Deepfakes and Fake News Pose a Growing Threat to Democracy [Электронный ресурс]. URL : https://news.northeastern.edu/2022/04/01/deepfakes-fake-news-threat-democracy/
4. Кочеткова Н.П. *Влияние дипфейков и постправды на общественное сознание: феноменологический анализ* // Общество: философия, история, культура. – 2024. – № 7.
5. How Deepfakes Are Impacting Public Trust in Media [Электронный ресурс]. URL : https://www.pindrop.com/article/deepfakes-impacting-trust-media/
6. The Impact of Deepfakes on Journalism [Электронный ресурс]. URL : https://www.pindrop.com/article/impact-deepfakes-journalism/
7. Researchers developing tool to instantly conceal and anonymize voices [Электронный ресурс]. URL : https://www.rochester.edu/newscenter/voice-changer-in-real-time-anonymizer-software-630652/
8. Generative AI and watermarking [Электронный ресурс]. URL : https://www.europarl.europa.eu/RegData/etudes/BRIE/2023/757583/EPRS\_BRI(2023)757583\_EN.pdf
1. Капитанов А.В., Линдре Ю.М. Дипфейк: невинная технология для развлечения или угроза современному обществу? // Российский совет по международным делам. – 2022. [↑](#footnote-ref-1)
2. Lundberg, E., Mozelius, P. The potential effects of deepfakes on news media and entertainment. AI & Soc 40, 2159–2170 – 2025. [↑](#footnote-ref-2)
3. Электронный ресурс: https://news.northeastern.edu/2022/04/01/deepfakes-fake-news-threat-democracy/ [↑](#footnote-ref-3)
4. Наталья Павловна Кочеткова Влияние дипфейков и постправды на общественное сознание: феноменологический анализ // Общество: философия, история, культура. 2024. №7. [↑](#footnote-ref-4)
5. Электронный ресурс: https://www.pindrop.com/article/deepfakes-impacting-trust-media/ [↑](#footnote-ref-5)
6. Электронный ресурс: https://www.pindrop.com/article/impact-deepfakes-journalism/ [↑](#footnote-ref-6)
7. Электронный ресурс: https://www.rochester.edu/newscenter/voice-changer-in-real-time-anonymizer-software-630652/ [↑](#footnote-ref-7)
8. Электронный ресурс: https://www.europarl.europa.eu/RegData/etudes/BRIE/2023/757583/EPRS\_BRI(2023)757583\_EN.pdf [↑](#footnote-ref-8)
9. Электронный ресурс: https://partnershiponai.org/cbc-framework-case-study/ [↑](#footnote-ref-9)