УДК 004.8:004.056

DEEPFAKE ТЕХНОЛОГИИ: УГРОЗЫ ДЛЯ ОБЩЕСТВА И МЕТОДЫ ОБНАРУЖЕНИЯ

DEEPFAKE TECHNOLOGIES: THREATS TO SOCIETY AND DETECTION METHODS

К.Т. Луговая, студентка I курса гр. ПИб-251.3
А.В. Ионина, к.т.н., доцент кафедры ИТиЭД,
Филиал Федерального Государственного бюджетного образовательного учреждения высшего образования «Кузбасский государственный технический университет имени Т.Ф. Горбачева» в г. Новокузнецке, г. Новокузнецк

Аннотация: статья посвящена комплексному анализу феномена технологий глубоких подделок (deepfake), представляющих собой синтез медиаконтента на основе искусственного интеллекта. Рассматривается двойственная природа технологии: ее потенциальные возможности и критические угрозы для информационной безопасности, политической стабильности и психологического благополучия общества. Особое внимание уделяется обзору и классификации современных методов обнаружения дипфейков, включая анализ пространственно-временных несоответствий и использование передовых нейросетевых архитектур. Доказывается необходимость комплексного подхода, сочетающего технологические решения, правовое регулирование и повышение медиаграмотности для минимизации рисков, связанных с распространением синтетического контента.

Ключевые слова: глубокие подделки, искусственный интеллект, информационная безопасность, дезинформация, методы обнаружения, нейросети, цифровая аутентификация.

Abstract: the article is devoted to a comprehensive analysis of the deepfake technology phenomenon, which involves the synthesis of media content based on artificial intelligence. The dual nature of the technology is considered: its potential opportunities and critical threats to information security, political stability, and the psychological well-being of society. Special attention is paid to the review and classification of modern deepfake detection methods, including the analysis of spatiotemporal inconsistencies and the use of advanced neural network architectures. The necessity of an integrated approach combining technological solutions, legal regulation, and improved media literacy to minimize the risks associated with the spread of synthetic content is substantiated.

Keywords: deepfakes, artificial intelligence, information security, disinformation, detection methods, neural networks, digital authentication.

Стремительное развитие генеративных нейросетей привело к появлению и широкому распространению технологий deepfake. Эти алгоритмы позволяют создавать гиперреалистичные видео-, аудио- и фотоматериалы, в которых действия и слова реальных людей являются сгенерированными. Как отмечают исследователи, deepfake создают беспрецедентные вызовы в сфере информационной безопасности, порождая феномен «кризиса свидетельства», когда аудиовизуальная запись перестает быть неопровержимым доказательством события [1]. Проблема усугубляется высокой доступностью данных технологий и постоянным совершенствованием алгоритмов генерации, что требует разработки принципиально новых, автоматизированных подходов к верификации контента.

Особую остроту проблема приобретает в контексте социально-политических процессов. Многочисленные исследования фиксируют взрывной рост числа дипфейков, используемых для дезинформации и кибератак. Актуальность исследования подтверждается конкретными данными. Только в России, по данным АНО «Диалог Регионы», с января по сентябрь 2025 года было выявлено 342 уникальных дипфейка, что в 4,1 раза больше, чем за весь 2024 год. При этом 79% таких подделок нацелены на губернаторов и других государственных служащих [2]. Количество просмотров подобного контента в 2025 году достигло 122,5 млн, что в 3,1 раза превышает показатели предыдущего года [2]. Эти цифры наглядно демонстрируют масштаб угрозы и необходимость эффективного противодействия. Целью данной работы является систематизация угроз, исходящих от deepfake-технологий для общества, анализ современных методов их обнаружения, а также обоснование необходимости комплексной стратегии противодействия, сочетающей технологические, правовые и образовательные меры.

Исследования общественного мнения показывают неоднозначное отношение к технологиям глубоких подделок. Масштабное исследование в семи европейских странах (N=7083) продемонстрировало, что респонденты во всех странах оценивают риски deepfake выше, чем потенциальные выгоды, причем этот разрыв увеличивается с возрастом и снижением уровня дохода [3]. Восприятие рисков варьируется в зависимости от страны и коррелирует с уровнем цифровой грамотности и зрелостью регулирования [3].

В таблице 1 представлены данные опросов россиян относительно их осведомленности о технологиях deepfake и отношения к ним.

Таблица 1 – Отношение россиян к технологиям deepfake по данным опросов 2025 г.

№	Показатель	Источник	Значение, %
1	Никогда не слышали о дипфейках	МТС Линк	48
2	Хотя бы слышали о технологии	Контур.Толк	60
3	Уверены, что смогут распознать дипфейк	Контур.Толк	52
4	Сталкивались с мошенническими дипфейками на работе	МТС Линк	30
5	Считают технологию опасной для всех	Контур.Толк	46

Угрозы от применения deepfake-технологий можно разделить на несколько категорий. В первую очередь, это использование синтетического контента для кибермошенничества и хищений. Как показывают криминалистические исследования, методы социальной инженерии с применением дипфейков (подмена лица, синтез голоса по образцу) открывают новые возможности для злоумышленников, что требует совершенствования методик расследования таких преступлений [4].

Во-вторых, это политическая дезинформация и дискредитация публичных лиц. В-третьих, подрыв доверия к медиа и социальным институтам в целом. Ответом на эти угрозы является развитие методов обнаружения дипфейков, которые также стремительно эволюционируют.

Современные подходы к детекции дипфейков можно классифицировать по нескольким направлениям. Первое направление включает методы, анализирующие физиологические и пространственно-временные несоответствия (артефакты на границах подмены, несинхронность губ и речи). Однако они часто оказываются неэффективны против высококачественных генеративных моделей. Более перспективным является второе направление – использование глубоких нейронных сетей.

Комплексный обзор современных методов, представленный в работе Singh et al., показывает, что наибольшей эффективностью обладают модели на базе трансформеров и генеративно-состязательных сетей (GAN), которые обучаются выявлять характерные «цифровые отпечатки» подделок на уровне пикселей или в частотной области [1]. Важнейшей проблемой, однако, остается обобщающая способность детекторов – их эффективность при столкновении с новыми, незнакомыми типами подделок. Для решения этой проблемы предлагаются различные подходы. Например, метод NormCura основан на отборе обучающих образцов, демонстрирующих стабильность предсказаний при изменении параметров нормализации, что позволяет фильтровать артефакты предобработки и обучать модели на более надежных признаках подделок [5].

Еще одним инновационным подходом является использование архитектуры на базе сетей Колмогорова–Арнольда (KAN). Метод LAKAN использует ключевые точки лица (ландмарки) в качестве структурного приоритета для динамической генерации внутренних параметров сети, что позволяет направлять фокус общего энкодера изображений на наиболее информативные области лица, содержащие признаки подделки [6]. Эксперименты показывают, что LAKAN достигает высоких результатов (AUC 99,71% на FF++), превосходя многие современные методы [6].

Третье перспективное направление – анализ кросс-модальных несоответствий, например, между аудио- и видеорядом. Метод AuViRe реконструирует речевые представления из одной модальности на основе другой; расхождения в этой реконструкции служат надежным индикатором манипуляции, позволяя точно локализовать поддельные фрагменты во времени [7].

Таким образом, технологии deepfake представляют собой серьезный вызов для современного общества, однако ответ на этот вызов уже формируется в виде постоянно совершенствующихся методов обнаружения. Эффективное противодействие должно строиться на комплексном подходе, включающем три ключевых блока:

1. Технологический блок – разработка более устойчивых (robust) алгоритмов детекции, обобщающих новые типы генерации, устойчивых к шумам, а также создание систем активной защиты, таких как цифровые водяные знаки.
2. Правовой блок – регулирование, аналогичное европейскому Digital Services Act и Акту об ИИ, которые устанавливают требования к платформам по выявлению и маркировке синтетического контента [8].
3. Образовательный блок – повышение цифровой и медиаграмотности населения, обучение навыкам критического мышления и фактчекинга.

Только синергия этих усилий позволит минимизировать риски и адаптироваться к новым реалиям информационного пространства.

Список литературы

1. Singh S. Unmasking digital deceptions: An integrative review of deepfake detection, multimedia forensics, and cybersecurity challenges / S. Singh [et al.] // MethodsX. – 2025. – Vol. 15. – P. 103632.
2. В России зафиксировали исторический максимум распространения дипфейков // ТАСС: сайт. – 2025. – 6 октября. – URL: https://tass.ru/obschestvo/25262115 (дата обращения: 10.03.2026).
3. Hynek N. Risks and benefits of artificial intelligence deepfakes: Systematic review and comparison of public attitudes in seven European Countries / N.

Hynek, B. Gavurova, M. Kubak // Journal of Innovation & Knowledge. – 2025. – Vol. 10, No. 5. – P. 100782.
4. Бражников Д. А. Criminalistic forecasting of thefts committed by using "deepfake"-technologies / Д. А. Бражников // Российский девиантологический журнал. – 2023. – № 2. – С. 187-195. – URL: https://sciup.org/140301937 (дата обращения: 10.03.2026).
5. Hou S. Normalization-consistent data curation for generalizable deepfake detection / S. Hou, X. Jiang, K. Xu [et al.] // Neurocomputing. – 2026. – Vol. 661. – P. 131838.
6. Jiang J. LAKAN: Landmark-assisted Adaptive Kolmogorov-Arnold Network for Face Forgery Detection / J. Jiang, X. Li, Y. Zhang [et al.] // arXiv preprint arXiv:2510.00634v3. – 2026.
7. Koutlis C. AuViRe: Audio-visual Speech Representation Reconstruction for Deepfake Temporal Localization / C. Koutlis, S. Papadopoulos // arXiv preprint arXiv:2511.18993v1. – 2026.
8. Answer given by Executive Vice-President Virkkunen on behalf of the European Commission to parliamentary question E-000617/25 // European Parliament: сайт. – 2025. – 11 June. – URL: https://www.europarl.europa.eu/doceo/document/E-10-2025-000617-ASW_EN.html (дата обращения: 10.03.2026).