ISSN: 2582-4686 SJIF 2021-3.261,SJIF 2022-2.889, 2024-6.875 ResearchBib IF: 9.948 / 2024

THE MULTIDISCIPLINARY JOURNAL OF SCIENCE AND TECHNOLOGY

VOLUME-6, ISSUE-1

# ЦИФРОВАЯ КРИМИНАЛИСТИКА В ЭПОХУ ДИПФЕЙКОВ: ВЫЗОВЫ И МЕТОДИКИ ВЫЯВЛЕНИЯ ИИ-ГЕНЕРИРОВАННЫХ АУДИО- И ВИДЕОЗАПИСЕЙ.

*Главный экспертно-криминалистический центр МВД РУ*
*гл. эксперт* ***Куприянова И.М.***

**Аннотация.** Стремительное развитие генеративных моделей искусственного интеллекта, включая генеративно-состязательные сети (GAN), диффузионные модели и системы синтеза речи, привело к широкому распространению дипфейков — синтетических фото-, видео- и аудиоматериалов, практически неотличимых от реальных. Данное явление ставит под угрозу достоверность цифровых доказательств и существенно осложняет деятельность правоохранительных органов и судебно-экспертных учреждений.

В статье рассматриваются ключевые угрозы, связанные с использованием дипфейков в преступной деятельности, анализируются пассивные и активные методы их выявления, а также предлагается практико-ориентированная методика судебно-криминалистического исследования подозрительных аудио- и видеозаписей.

**Ключевые слова:** цифровая криминалистика, дипфейк, искусственный интеллект, PRNU, цифровые водяные знаки, C2PA, аудио-дипфейк, видео-дипфейк, судебная экспертиза.

## ВВЕДЕНИЕ

Под дипфейками принято понимать изображения, видеозаписи и аудиоматериалы, созданные либо существенно модифицированные с применением алгоритмов машинного обучения и нейросетевых моделей. Современные технологии позволяют с высокой степенью правдоподобия имитировать внешний облик и голос конкретного человека, что уже активно используется в целях шантажа, финансового мошенничества, политической дезинформации и дискредитации должностных лиц.

Для цифровой криминалистики это означает резкое снижение доверия к ранее считавшимся «золотым стандартом» доказательств — фотографиям, видеозаписям и аудиозаписям. Судебные системы многих государств не успевают адаптировать процессуальные механизмы под новую технологическую реальность, а дефицит подготовленных специалистов в области анализа цифровых доказательств становится критическим.

В этих условиях задача судебного эксперта заключается не только в выявлении факта монтажа, но и в дифференциации традиционных методов фальсификации от синтетического контента, созданного с применением искусственного интеллекта, а также, по возможности, в установлении типа использованной генеративной технологии.

### Технологические основы создания дипфейков

Современные исследования позволяют выделить несколько основных классов дипфейков, каждый из которых обладает собственными техническими особенностями и криминалистически значимыми признаками.

К первой группе относятся манипуляции лицом в видеозаписях. Они включают технологии *face swapping*, предполагающие замену лица одного человека лицом другого, *face reenactment* (или *puppeteering*), при которых мимика и движения лица управляются по образцу лица «актёра», а также *lip-sync* — синхронизацию движений губ с подменённой аудиодорожкой.

ISSN: 2582-4686 SJIF 2021-3.261,SJIF 2022-2.889, 2024-6.875 ResearchBib IF: 9.948 / 2024

THE MULTIDISCIPLINARY JOURNAL OF SCIENCE AND TECHNOLOGY

VOLUME-6, ISSUE-1

Вторая группа представлена полной генерацией визуального образа. С использованием GAN и диффузионных моделей создаются реалистичные, но фактически не существующие люди, сцены и события, не имеющие реального прототипа.

Отдельную категорию составляют аудио-дипфейки. К ним относятся системы синтеза речи (*text-to-speech*), позволяющие генерировать речь по тексту с голосом конкретного человека, а также технологии *voice conversion*, преобразующие голос одного говорящего в голос другого при сохранении исходного лингвистического содержания.

Каждый из перечисленных классов оставляет характерные цифровые следы, которые могут быть использованы в рамках судебно-криминалистического анализа.

**Риски для правоохранительной деятельности и судебной экспертизы**

Использование дипфейков формирует комплекс угроз для правоохранительных органов и судебной системы. Прежде всего речь идёт о фальсификации доказательств, включая создание фиктивных «видеопризнаний», поддельных аудиозаписей переговоров, а также фабрикацию алиби посредством синтетических записей с камер видеонаблюдения.

Существенную опасность представляют атаки на биометрические системы. Дипфейки голоса и лица всё чаще используются для обхода систем дистанционной идентификации, банковских сервисов и механизмов аутентификации.

Не менее значимым является риск массовой дезинформации и подрыва доверия к официальным сообщениям. В условиях информационных операций дипфейки могут быть направлены как против государственных институтов, так и против конкретных должностных лиц.

Отдельного внимания заслуживает так называемый «эффект всеобщего отрицания». Чем более известны технологии дипфейков, тем проще реальному правонарушителю заявлять, что подлинная запись также является подделкой. Это требует от экспертов применения комплексных методик аутентификации и строгого соблюдения процессуальных правил обращения с цифровыми доказательствами.

**Пассивные методы выявления дипфейков в видеозаписях**

Пассивные методы детекции не предполагают предварительного внедрения специальных меток в контент и основаны на анализе самой записи и сопутствующих метаданных.

К данной группе относятся алгоритмы, выявляющие локальные артефакты генерации, такие как размытые границы лица, неестественная текстура кожи, некорректные отражения в глазах, несогласованность освещения и теней. Анализируются также геометрические несоответствия, включая аномальное положение глаз и рта, а также нарушения биологических паттернов — частоты моргания, микродвижений глаз и мимических реакций.

Как правило, эти признаки извлекаются с использованием нейросетевых моделей и применяются для классификации видеоматериала как подлинного либо синтетического.

Генеративные модели часто оставляют характерные следы в частотном спектре и динамике видео. К ним относятся неестественная структура высокочастотного шума, следы интерполяции кадров, а также рассогласование движений лица и фона. Для выявления подобных аномалий используются методы фильтрации, вейвлет- и DCT-преобразования, а также анализ временной согласованности последовательных кадров.

В традиционной цифровой криминалистике широко используется анализ Photo Response Non-Uniformity (PRNU) — уникального шумового «отпечатка» матрицы камеры. Для дипфейков

ISSN: 2582-4686 SJIF 2021-3.261,SJIF 2022-2.889, 2024-6.875 ResearchBib IF: 9.948 / 2024

THE MULTIDISCIPLINARY JOURNAL OF SCIENCE AND TECHNOLOGY

VOLUME-6, ISSUE-1

характерно либо полное отсутствие PRNU, либо его существенное искажение, а также несоответствие заявленному устройству съёмки. Часто наблюдается выпадение PRNU именно в области подменённого лица.

Аудио-дипфейки, создаваемые системами синтеза речи и преобразования голоса, также оставляют характерные цифровые следы.

Анализ акустических и спектральных признаков позволяет выявить избыточно сглаженный либо, напротив, фрагментированный спектр, повторяющиеся паттерны шума и аномалии высокочастотной составляющей. Существенное значение имеет исследование просодических характеристик, включая вариативность высоты тона, ритм и интонацию речи.

В практической деятельности всё шире применяются глубокие нейросетевые детекторы, обученные на специализированных наборах данных. Вместе с тем для судебной экспертизы принципиально важно сочетать алгоритмические методы с классическими приёмами фонетической и акустической экспертизы.

**Активные методы аутентификации и противодействия дипфейкам**

Активные методы предполагают предварительное внедрение в контент специальных маркеров, упрощающих последующую проверку его подлинности.

Цифровые водяные знаки и скрытые маркеры позволяют подтверждать происхождение записи, выявлять существенные изменения и обеспечивать прослеживаемость копий. Существенным ограничением является необходимость их массового внедрения в аппаратно-программную инфраструктуру.

На международном уровне развивается стандарт C2PA, предусматривающий использование криптографически защищённых метаданных (*Content Credentials*). Такой подход позволяет формировать цепочку провенанса, отражающую историю создания и обработки контента.

Регуляторные инициативы, включая положения Европейского акта об искусственном интеллекте, вводят обязательства по маркировке синтетического контента и формируют правовую основу для его идентификации.

**Проект методики судебно-криминалистического исследования аудио- и видеозаписей**

Для экспертно-криминалистической практики целесообразно применение многоуровневой методики, включающей обеспечение сохранности исходных данных, анализ контейнера и метаданных, кадрово-покадровый и спектральный анализ видео, исследование аудиодорожки, проверку активных меток и контент-провенанса, а также контекстуальную и OSINT-проверку. Формирование выводов должно сопровождаться чётким разграничением установленных фактов, вероятностных оценок и вопросов, выходящих за пределы компетенции эксперта.

**Проблемы и перспективы развития**

Несмотря на активное развитие методов детекции, цифровая криминалистика в сфере дипфейков сталкивается с рядом нерешённых проблем, включая быструю эволюцию генеративных моделей, развитие контр-форонзики, низкую распространённость активных стандартов и процессуальную неопределённость. Перспективными направлениями остаются интеграция детекторов в платформы распространения контента, развитие стандартов и совершенствование подготовки экспертов.

Эпоха дипфейков требует перехода от точечного анализа отдельных файлов к системному подходу, сочетающему пассивные и активные методы выявления, строгие процессуальные правила и контекстуальную проверку цифровых доказательств. Только при тесном

ISSN: 2582-4686 SJIF 2021-3.261,SJIF 2022-2.889, 2024-6.875 ResearchBib IF: 9.948 / 2024



взаимодействии разработчиков ИИ-систем, законодателей и экспертно-криминалистических подразделений возможно сохранить доверие общества к цифровым доказательствам и обеспечить эффективное расследование преступлений в условиях стремительного развития генеративных технологий.

**Список литературы:**

1. Amerini I. et al. Deepfake Media Forensics: Status and Future Challenges. Journal of Imaging, 2025.
2. Malik A. DeepFake Detection for Human Face Images and Videos. IEEE Transactions on Information Forensics and Security.
3. Zhang Y. et al. A Review of Generated Images and Deepfake Detection Technologies. Visual Computing for Industry, Biomedicine, and Art, 2025.
4. Lugstein F. et al. PRNU-based Deepfake Detection. Proceedings of the ACM Conference, 2021.
5. Zhang B. et al. Audio Deepfake Detection: What Has Been Achieved and Where We Are Going. Sensors, 2025.
6. Warren K. et al. Pitch Imperfect: Detecting Audio Deepfakes Through Acoustic Prosodic Analysis. arXiv, 2025.
7. Lai Z. et al. Enhancing Deepfake Detection: Proactive Forensics Techniques Using Digital Watermarking. Digital Investigation, 2025.
8. Coalition for Content Provenance and Authenticity (C2PA). Content Credentials Specification, v.2.2.
9. EU Artificial Intelligence Act, Article 50.
10. Axios. Courts aren’t ready for AI-generated evidence, 2025.