Дата поступления рукописи в редакцию: 21.05.2026 г.
Дата принятия рукописи в печать: 11.07.2026 г.

DOI: 10.24412/2782-3849-2026-7-320-325

СКВОРЦОВ Константин Васильевич,

кандидат технических наук, доцент, доцент кафедры специальной техники и информационных технологий юридического факультета ВЮИ ФСИН России, Владимир, Россия, ORCID 0000-0001-8611-3353,

e-mail: k-skv@yandex.ru

АЛЯКРИНСКАЯ Наталия Олеговна,

старший преподаватель кафедры специальной техники и информационных технологий юридического факультета ВЮИ ФСИН России, Владимир, Россия,

e-mail: mail@law-books.ru

АКТУАЛЬНЫЕ ПРОБЛЕМЫ ДЕТЕКЦИИ АУДИОВИЗУАЛЬНЫХ ДИПФЕЙКОВ В ПРАВООХРАНИТЕЛЬНОЙ И СУДЕБНО-ЭКСПЕРТНОЙ ДЕЯТЕЛЬНОСТИ

Аннотация. *Развитие генеративных нейросетей привело к появлению высокореалистичных дипфейков, создающих угрозы достоверности цифровых доказательств. В работе анализируются пассивные (частотный, PRNU, физиологический, аудио-спектральный) и активные (C2PA, цифровые водяные знаки) методы детекции синтетического контента. Показано, что эффективность детекторов резко снижается на реальных («in-the-wild») данных, что требует комплексной экспертизы. Для уголовно-исполнительной системы дипфейки создают риски дискредитации сотрудников и фальсификации доказательств. Обоснована необходимость разработки ведомственной криминалистической методики и подготовки кадров.*

Ключевые слова: *цифровая криминалистика, дипфейк, генеративно-состязательные сети, диффузионные модели, аудио-дипфейк, уголовно-исполнительная система, судебная экспертиза.*

SKVORTSOV Konstantin Vasilyevich,

candidate of Technical Sciences (PhD), Associate Professor, associate Professor of the Department of Special Technology and Information Technologies of the Law Faculty of Vladimir Law Institute of the Federal Penitentiary Service of Russia, Vladimir, Russia

ALYAKRINSKAYA Nataliya Olegovna,

senior Lecturer of the Department of Special Technology and Information Technologies of the Law Faculty of Vladimir Law Institute of the Federal Penitentiary Service of Russia, Vladimir, Russia

CURRENT ISSUES IN AUDIOVISUAL DEEPFAKE DETECTION IN LAW ENFORCEMENT AND FORENSIC EXPERT PRACTICE

Annotation. *The development of generative neural networks has led to the emergence of highly realistic deepfakes, posing threats to the reliability of digital evidence. This paper analyzes passive (frequency-domain, PRNU, physiological, audio-spectral) and active (C2PA, digital watermarks) methods for synthetic content detection. It is shown that the performance of detectors drops sharply on real-world («in-the-wild») data, necessitating comprehensive forensic examination. For the penal enforcement system, deepfakes create risks of staff discreditation and evidence falsification. The need to develop a departmental forensic methodology and to provide relevant training is substantiated.*

Key words: *digital forensics, deepfake, generative adversarial networks, diffusion models, audio deepfake, penal enforcement system, forensic examination*.

ВВЕДЕНИЕ

Цифровизация общественных отношений является ключевым фактором трансформации современного социума и правоприменительной практики, одновременно создавая новые формы криминальных посягательств в информационной среде. Одним из наиболее значимых феноменов последних лет стали технологии глубокого синтеза – дипфейки, понимаемые как реалистичная манипуляция аудио-, фото- и видеоматериалами с использованием алгоритмов глубокого обучения для достижения максимального сходства с реальными объектами [1; 2]. По оценкам зарубежных аналитических центров, количество дипфейк-материалов в сети возросло с порядка 500 тыс. объектов в 2023 г. до нескольких миллионов к 2025 г., при этом число мошеннических атак с их использованием увеличилось в десятки раз [3]. Международные структуры предупреждают, что в ближайшие годы синтетически сгенерированный контент может составлять значительную долю всей цифровой информации, циркулирующей в интернете [4]. В Российской Федерации аналогичные тенденции фиксируются в сегменте киберпреступности. По данным МВД России, общее количество преступлений в сфере компьютерной информации к 2024 г. достигло рекордных значений, а их удельный вес в общей преступности продолжает расти [5; 6]. Исследователи отмечают, что киберпреступления всё чаще затрагивают информационную инфраструктуру учреждений и органов уголовно-исполнительной системы России (далее – УИС), становясь фактором дестабилизации их деятельности и источником существенных материальных и репутационных рисков [5; 7]. Особую опасность для цифровой криминалистики представляют дипфейки, способные имитировать поведение участников правоотношений: спецконтингент, сотрудников УИС, иных должностных лиц. Такие материалы могут использоваться для фальсификации цифровых доказательств, дискредитации органов правопорядка. В этих условиях возникает объективная потребность в формировании целостной технико-криминалистической парадигмы выявления и исследования дипфейк-контента, опирающейся на современные достижения судебной экспертизы, направленной на сбор, анализ, сохранение и представление цифровых доказательств, полученных из электронных устройств, сетей и информационных систем (англ. Digital Forensics, далее – цифровой форензики) и учитывающей специфику информационной безопасности УИС, включая предложенные ранее меры по профилактике киберпреступлений и кибератак в учреждениях ФСИН России [5; 6].

Цель исследования - разработка и обосновании технико-криминалистической парадигмы выявления и исследования дипфейк-контента в системе цифровой криминалистики. Задачи исследования – определить, какие технические и криминалистические средства позволяют обеспечить достоверную идентификацию дипфейков, какова их эмпирическая результативность и каким образом они могут быть интегрированы в практику ведомственных экспертных подразделений правоохранительных органов России.

Криминалистический и технико-криминалистический анализ служит основой для систематизации методов детекции дипфейков, выявления их сильных и слабых сторон, а также определения областей применимости в судебно-экспертной и ведомственной практике [8; 9]. Обзорно-эмпирический подход реализуется при обобщении результатов международных и отечественных исследований, посвящённых эффективности детекторов дипфейков на стандартных бенчмарках (FaceForensics++, DFDC и др.) и реальном контенте («in-the-wild») [2; 10]. Системный подход позволяет интегрировать технико-криминалистические решения с организационными и правовыми мерами, ранее предложенными для повышения информационной безопасности учреждений и органов ФСИН России, включая меры по противодействию кибератакам, организацию многоуровневой аутентификации и совершенствование подготовки персонала.

Анализ эволюции генеративных технологий показывает, что переход от классических генеративно-состязательных сетей (Generative Adversarial Networks, далее – GAN) к диффузионным моделям и мультимодальным системам синтеза привёл к качественному усложнению криминалистической задачи выявления дипфейков. Генеративно-состязательные нейронные сети реализуют состязательное обучение генератора и дискриминатора, где первый стремится создавать изображения и видео, неотличимые от реальных, а второй – распознавать подделки. Диффузионные модели, в отличие от GAN, опираются на стадийный процесс зашумления и восстановления изображения, обеспечивая более стабильное обучение и более высокий уровень фотореализма, что существенно снижает эффективность методов, ориентированных на классические артефакты GAN-генерации [2; 9].

С криминалистической точки зрения целесообразно выделять несколько видов дипфейк-контента. К ним относятся подмена лица (face-swap), управление мимикой

(face-reenactment), полностью синтетический контент, аудио-дипфейки и мультимодальные конструкции, совмещающие искажение визуального и звукового ряда [2; 4; 8]. Для уголовно-исполнительной системы наиболее опасными являются мультимодальные дипфейки, способные воспроизводить якобы подлинные высказывания и действия осуждённых или сотрудников ФСИН России, что создаёт риск информационного давления и дискредитации органов УИС [5; 6].

Пассивные методы технико-криминалистического исследования дипфейков основаны на выявлении признаков искусственного происхождения в анализируемом контенте без предварительного встраивания защитной информации. К числу таких методов относится визуально-артефактный анализ, ориентированный на фиксацию пиксельных аномалий, несогласованности освещения, артефактов по контуру лица и иных следов генерации, которые могут быть выявлены как экспертным путём, так и с использованием свёрточных нейронных сетей и визуальных трансформеров [2; 9]. Частотно-доменный анализ (FFT, DCT и др.) позволяет обнаруживать специфические спектральные распределения мощности сигнала по частотным диапазонам, которые характерны для синтетически сгенерированных изображений и отличаются от спектральных свойств реальных записей, возникающие при генерации и последующей обработке изображений и видеозаписей, и отличается относительной устойчивостью к повторному сжатию и пересылке, что важно для условий циркуляции контента через социальные сети и мессенджеры.

PRNU-анализ исследует статистически устойчивый «отпечаток» матрицы цифровой камеры и позволяет выявлять несоответствие между зазванным источником записи и фактическими характеристиками изображения [8; 9]. Методы, основанные на анализе физиологических и поведенческих признаков, включают дистанционную фотоплетизмографию, оценку частоты моргания, микродвижений лицевых мышц и синхронизации артикуляции губ с аудиорядом, что позволяет отличать живое изображение от синтетического на основе биологически обусловленных закономерностей. В сфере аудио-форензики активно развиваются подходы, использующие мел-спектрограммы и другие акустические признаки, анализируемые моделями CNN/LSTM; показана высокая точность идентификации синтетически сгенерированного голоса на ряде специализированных и кросс-датасетных выборок [11; 12]. Для УИС это имеет принципиальное значение при проверке аудиозаписей, представляемых в качестве доказательств противоправных действий, угроз, вымогательства либо мошеннических действий.

Активные методы аутентификации контента предполагают предварительное встраивание верифицируемой информации в медиафайлы и использование стандартов происхождения. Цифровые водяные знаки и стандарт C2PA обеспечивают формирование машиночитаемых «паспортов происхождения» для изображений, видео и аудио, включая сведения об авторе, средстве съёмки, цепочке обработки и криптографической подписи [13]. Интеграция C2PA со средствами распределённого реестра (блокчейн-верификацией) позволяет создать устойчивую цепочку доверия от момента формирования медиафайла, что особенно важно для видеоконференцсвязи и электронного документооборота в учреждениях и органах УИС. Внедрение таких решений логично продолжает предложенный ранее комплекс организационно-технических мер по противодействию кибератакам в УИС, ориентированных на многоуровневую аутентификацию, сегментацию сетей и использование сертифицированных средств защиты информации.

Эмпирическая оценка эффективности детекторов дипфейков показывает, что многие модели демонстрируют высокие показатели точности на стандартных бенчмарках (FaceForensics++, DFDC и др.), однако их результативность резко снижается при применении к реальному («in-the-wild») контенту, полученному с использованием иных генеративных средств и прошедшему дополнительную обработку в цифровой среде. При этом человеческая способность распознавать высококачественные дипфейки также ограничена, что подтверждается экспериментальными данными о низкой доле правильных идентификаций даже у подготовленных респондентов [2]. В совокупности эти факторы свидетельствуют о недопустимости опоры исключительно на автоматизированные средства анализа без комплексной криминалистической оценки и без учёта контекста получения и использования цифровых материалов.

Угрозы дипфейков для УИС следует рассматривать в тесной связи с уже исследованными угрозами кибератак. Они могут примениться как инструмент дискредитации сотрудников и органов УИС, с целью давления на руководство учреждений, совершения мошеннических действий в отношении осуждённых и их родственников, а также манипулирования общественным мнением в связи

с инцидентами в местах лишения свободы [5; 7]. Как подчёркивается в историко-криминологическом анализе киберпреступности в Российской Федерации [14], за период 2000–2025 гг. число IT-преступлений выросло с единичных случаев до 765,4 тыс. в 2024 году, а их доля в общей структуре преступности достигла 40 %, что свидетельствует о системном характере угроз, включая использование дипфейков в криминальных целях. Это обуславливает необходимость включения проблематики дипфейков в систему ведомственного планирования мер по обеспечению информационной безопасности УИС.

Особое значение приобретают методические и кадровые аспекты. Существующие методики судебной видеотехнической, фонографической и компьютерно-технической экспертизы только частично учитывают специфику генеративного ИИ и требуют переработки в направлении формирования частной криминалистической методики исследования дипфейк-контента [15; 16]. Такая методика должна предусматривать использование нескольких независимых методов обнаружения, оценку совокупной вероятности синтетического происхождения контента, стандартизацию документирования применяемых программных средств и алгоритмов, а также адаптацию процедур к особенностям видеонаблюдения, видеоконференцсвязи в органах УИС. Одновременно исследования, посвящённые кибератакам на учреждения ФСИН России, подчёркивают необходимость систематической подготовки персонала как в части соблюдения требований информационной безопасности, так и в части базового понимания современных цифровых угроз, включая дипфейки. Это предполагает включение соответствующих модулей в программы подготовки и повышения квалификации сотрудников ФСИН России, прежде всего экспертов и сотрудников оперативных подразделений, что позволит обеспечить адекватную реакцию на новые криминальные практики, основанные на использовании технологий глубокого синтеза.

Проведённое исследование позволяет сформулировать ряд обобщающих выводов.

Технологии глубокого синтеза (GAN, диффузионные модели, нейросинтез речи) привели к качественному изменению криминальной ситуации в цифровой среде, создав класс высокореалистичных дипфейков, способных подрывать доверие к цифровым доказательствам и использоваться в преступных целях.

Современные пассивные (визуально-артефактные, частотно-доменные, PRNU-, физиологические, аудио-форензика) и активные (цифровые водяные знаки, C2PA, блокчейн-верификация) методы обнаружения дипфейков обладают существенным потенциалом, однако страдают от «кризиса обобщения» при применении к реальному контенту «in-the-wild», что требует их использования в рамках комплексной криминалистической экспертизы и при строгом соблюдении принципов научной обоснованности и верифицируемости.

Отечественная судебно-экспертная практика по исследованию дипфейк-контента находится на этапе становления; необходима разработка частной криминалистической методики экспертизы дипфейков с учётом специфики задач ФСИН России и требований Федерального закона от 25 июня 2002 года № 73-ФЗ.

Перспективными направлениями дальнейших исследований являются разработка адаптивных детекторов дипфейков с механизмами непрерывного обучения, создание отечественных эталонных датасетов дипфейк-контента, в том числе на основе материалов, релевантных для УИС, а также формирование межведомственных протоколов экспертного взаимодействия при расследовании преступлений и инцидентов, связанных с использованием технологий глубокого синтеза.

Список литературы:

[1] Бодров, Н. Ф. Понятие дипфейка в российском праве, классификация дипфейков и вопросы их правового регулирования / Н. Ф. Бодров, А. К. Лебедева // Юридические исследования. – 2023. – № 11. – С. 26-41. – DOI 10.25136/2409-7136.2023.11.69014. – EDN DYIHIR.

[2] Tolosana, R., et al. «Deepfake Media Forensics: Status and Future Challenges» // Journal of Imaging (спецвыпуск по теме Media Forensics). – 2025. – Vol. 11, No. 3. – Art. 73.

[3] Deepfake Statistics 2025: The Data Behind the AI Fraud Wave [Электронный ресурс] // DeepStrike : website. – URL: https://deepstrike.io/blog/deepfake-statistics-2025 (дата обращения: 12.05.2026)

[4] Basu R., Chakraborty S., Datta R., Zhang S. Generative Artificial Intelligence and the Evolving Challenge of Deepfake Detection: A Systematic Analysis // Journal of Sensor and Actuator Networks. 2025. Vol. 14, no. 1. P. 17. URL: https://www.mdpi.com/2224-2708/14/1/17 (дата обращения: 12.05.2026)

[5] Скворцов, К. В. О мерах профилактики киберпреступлений в уголовно-исполнительной системе Российской Федерации / К. В. Скворцов, В. И. Силенков // Профессиональное юридическое образование и наука. – 2024. – № 3(15). – С. 61-63.

– EDN USZQFG

[6] Скворцов, К. В. Кибератаки как угроза информационной безопасности учреждениям и органам уголовно-исполнительной системы Российской Федерации / К. В. Скворцов, В. И. Силенков, А. В. Печенин // Профессиональное юридическое образование и наука. – 2025. – № 1(17). – С. 48-51. – EDN EPQAVA.

[7] Цимбал, В. Н. Анализ деятельности государственных органов Российской Федерации по противодействию киберпреступности / В. Н. Цимбал // Вестник Московского университета МВД России. – 2024. – № 4. – С. 183-191. – DOI 10.24412/2073-0454-2024-4-183-191. – EDN JUBWVA.

[8] Korus, P., & Huang, J. «Multi-scale Analysis Strategies in PRNU-based Tampering Localization» // IEEE Transactions on Information Forensics and Security. – 2017. – Vol. 12, No. 4. – pp. 809–824.

[9] Verdoliva, L. «Media Forensics in the Age of Generative AI: A Survey of Recent Advances and Challenges» // IEEE Journal of Selected Topics in Signal Processing. – 2024. – Vol. 18, No. 2. – pp. 120-145.

[10] Deepfake-Eval-2024: A Multi-Modal In-the-Wild Benchmark of Deepfakes Circulated in 2024. arXiv:2503.02857. – 2025.

[11] Alnaqbi M., Ikuesan R. A. A systematic review of audio deepfake detection techniques for digital investigation // Discover Computing. — 2026. — Vol. 29, iss. 1. — DOI: 10.1007/s10791-026-10077-1

[12] Левшин, Д. В. Исследование задачи автоматизированного сопоставления аудиофайлов / Д. В. Левшин, Д. В. Быстряков, А. В. Зубков // Моделирование, оптимизация и информационные технологии. – 2025. – Т. 13, № 4(51). – DOI 10.26102/2310-6018/2025.51.4.004. – EDN LJEWOD.

[13] Coalition for Content Provenance and Authenticity (C2PA). C2PA Technical Specification // C2PA : [сайт]. — URL: https://c2pa.org (дата обращения: 12.05.2026)

[14] Скворцов К.В., Силенков В.И. Киберпреступность в Российской Федерации: историко-криминологический анализ (2000–2025 гг.) // Профессиональное юридическое образование и наука. – 2026. – № 4 – С. 867-876.

[15] Зайцев В. В. Дипфейки: криминалистические риски и возможности использования в доказывании // Судебная экспертиза. – 2023. – № 3 (75). – С. 34–42.

[16] Богатырев, К. М. Формирование частной теории использования специальных знаний в целях обеспечения информационно-мировоззренческой безопасности в цифровой среде / К. М. Богатырев // Актуальные проблемы российского права. – 2021. – Т. 16, № 10(131). – С. 124-134. – DOI 10.17803/1994-1471.2021.131.10.124-134. – EDN TNTMEW.

References:

[1] Bodrov, N.F. The concept of deepfake in Russian law, the classification of deepfakes and issues of their legal regulation/N.F. Bodrov, A.K. Lebedeva//Legal research. – 2023. – № 11. - S. 26-41. – DOI 10.25136/2409-7136.2023.11.69014. – EDN DYIHIR.

[2] Tolosana, R., et al. «Deepfake Media Forensics: Status and Future Challenges» // Journal of Imaging (спецвыпуск по теме Media Forensics). – 2025. – Vol. 11, No. 3. – Art. 73.

[3] Deepfake Statistics 2025: The Data Behind the AI Fraud Wave [Электронный ресурс] // DeepStrike : website. - URL: https://deepstrike.io/blog/deepfake-statistics-2025 (accessed on: 12.05.2026)

[4] Basu R., Chakraborty S., Datta R., Zhang S. Generative Artificial Intelligence and the Evolving Challenge of Deepfake Detection: A Systematic Analysis // Journal of Sensor and Actuator Networks. 2025. Vol. 14, no. 1. P. 17. URL: https://www.mdpi.com/2224-2708/14/1/17 (accessed on: 12.05.2026)

[5] Skvortsov, K.V. On measures to prevent cybercrime in the penal system of the Russian Federation/K.V. Skvortsov, V.I. Silenkov//Professional legal education and science. – 2024. – № 3(15). - S. 61-63. – EDN USZQFG

[6] Skvortsov, K.V. Cyberattacks as a threat to information security to institutions and bodies of the penal system of the Russian Federation/K.V. Skvortsov, V.I. Silenkov, A.V. Pechenin//Professional legal education and science. – 2025. – № 1(17). - S. 48-51. – EDN EPQAVA.

[7] Tsimbal, V. N. Analysis of the activities of state bodies of the Russian Federation to counter cybercrime/V. N. Tsimbal//Bulletin of the Moscow University of the Ministry of Internal Affairs of Russia. – 2024. – № 4. - S. 183-191. – DOI 10.24412/2073-0454-2024-4-183-191. – EDN JUBWVA.

[8] Korus, P., & Huang, J. «Multi-scale Analysis Strategies in PRNU-based Tampering Localization» // IEEE Transactions on Information Forensics and Security. – 2017. – Vol. 12, No. 4. – pp. 809–824.

[9] Verdoliva, L. «Media Forensics in the Age of Generative AI: A Survey of Recent Advances and Challenges» // IEEE Journal of Selected Topics in Signal Processing. – 2024. – Vol. 18, No. 2. – pp. 120-145.

[10] Deepfake-Eval-2024: A Multi-Modal In-the-Wild Benchmark of Deepfakes Circulated in 2024. arXiv:2503.02857. – 2025.

[11] Alnaqbi M., Ikuesan R. A. A systematic review of audio deepfake detection techniques for digital investigation // Discover Computing. — 2026. — Vol. 29, iss. 1. — DOI: 10.1007/s10791-026-

10077-1

[12] Levshin, D.V. Study of the task of automated comparison of audio files/D.V. Levshin, D.V. Bystryakov, A.V. Zubkov//Modeling, optimization and information technologies. – 2025. - T. 13, NO. 4 (51). – DOI 10.26102/2310-6018/2025.51.4.004. – EDN LJEWOD.

[13] Coalition for Content Provenance and Authenticity (C2PA). C2PA Technical Specification // C2PA : [сайт]. - URL: https://c2pa.org (accessed on: 12.05.2026)

[14] Skvortsov K.V., Silenkov V.I. Cybercrime in the Russian Federation: historical and criminological analysis (2000-2025)//Professional legal education and science. – 2026. - NO. 4 - S. 867-876.

[15] Zaitsev V.V. Deepfakes: forensic risks and possibilities of use in evidence//Forensic examination. – 2023. – № 3 (75). - S. 34-42.

[16] Bogatyrev, K. M. Formation of a private theory of the use of special knowledge in order to ensure information and ideological security in the digital environment/K. M. Bogatyrev//Actual problems of Russian law. – 2021. - T. 16, NO. 10 (131). - S. 124-134. – DOI 10.17803/1994-1471.2021.131.10.124-134. – EDN TNTMEW.