

Аналитическая статья

УДК 316.77; DOI: 10.61260/2074-1618-2026-1-88-92

## ЭТИЧЕСКИЕ ВЫЗОВЫ ДИПФЕЙКОВ

✉ Шляпников Виктор Валерьевич.

Санкт-Петербургский университет ГПС МЧС России, Санкт-Петербург, Россия

✉ [shlyapnikovv@mail.ru](mailto:shlyapnikovv@mail.ru)

*Аннотация.* Анализируются дипфейки, или созданные с помощью искусственного интеллекта фальшивые видео- и/или аудиоматериалы, стирающие грань между правдой и обманом и поднимающие важные этические вопросы, связанные с дезинформацией, ущербом репутации и вмешательством в частную жизнь. Констатируется, что дипфейки могут оказывать влияние на общественное мнение и доверие к средствам массовой информации, а также представлять серьезную проблему для правовых систем. Подчеркивается важность осознания разработчиками и пользователями своей этической ответственности при использовании современных технологий для обеспечения максимизации их положительных аспектов и минимизации негативных последствий. Делается вывод, что обучение критическому мышлению и медиаграмотности может способствовать повышению осведомленности о существовании и потенциальных опасностях дипфейков и принятию обоснованных решений относительно распространяемого контента.

*Ключевые слова:* дипфейки, искусственный интеллект, медиаграмотность, этика

**Для цитирования:** Шляпников В.В. Этические вызовы дипфейков // Психолого-педагогические проблемы безопасности человека и общества. 2026. № 1 (70). С. 88–92. DOI: 10.61260/2074-1618-2026-1-88-92

Analytical article

## ETHICAL CHALLENGES OF DEEPFAKES

✉ Shlyapnikov Viktor V.

Saint-Petersburg university of State fire service of EMERCOM of Russia, Saint-Petersburg, Russia

✉ [shlyapnikovv@mail.ru](mailto:shlyapnikovv@mail.ru)

*Abstract.* This article analyzes deepfakes, or artificially generated video and/or audio materials that blur the line between truth and deception and raise important ethical questions related to disinformation, reputational damage, and invasion of privacy. It notes that deepfakes can influence public opinion and trust in media, and pose a serious challenge to legal systems. It emphasizes the importance of developers and users understanding their ethical responsibilities when using modern technologies to maximize their positive aspects and minimize their negative consequences. It concludes that training in critical thinking and media literacy can help raise awareness of the existence and potential dangers of deepfakes and enable informed decision-making regarding the content they distribute.

*Keywords:* deepfakes, artificial intelligence, media literacy, ethics

**For citation:** Shlyapnikov V.V. Ethical challenges of deepfakes // Psihologo-pedagogicheskie problemy bezopasnosti cheloveka i obshchestva = Psychological and pedagogical problems of human and social security. 2026. № 1 (70). P. 88–92. DOI: 10.61260/2074-1618-2026-1-88-92

## Введение

Дипфейки (англ. deepfake от deep learning «глубокое обучение» + fake «подделка») – это видео- или аудиофайлы, созданные с помощью искусственного интеллекта, в которых люди говорят или делают то, чего они на самом деле никогда не говорили и не делали [1]. Технология дипфейков одновременно вызывает огромный интерес и серьезные этические проблемы. Крайне важно решать эти проблемы и продвигать ответственное использование этой технологии, поскольку возможность создания реалистичных дипфейков может нанести серьезный ущерб при использовании злоумышленниками.

## Основная часть

Технология, лежащая в основе дипфейков, использует машинное обучение и нейронные сети для замены лиц или голосов людей в видео- или аудиофайлах [2, 3]. Обычно это делается настолько убедительно, что часто бывает трудно отличить подлинный материал от поддельного. Дипфейки создаются с помощью процессов обучения, в ходе которых нейронная сеть получает большие объемы данных для изучения определенных черт лица или голосовых шаблонов и может применять их к целевой видеопоследовательности или аудиофайлу. Например, дипфейком может быть видеозапись с участием политика, произносящего речь, которую он на самом деле никогда не произносил, или изображение известного человека в скандальном контексте, которого никогда не было.

Дипфейки создают множество этических проблем, начиная от дезинформации и заканчивая ущербом репутации и вмешательством в личную жизнь [4]. Создавая и распространяя поддельные видео- или аудиозаписи, можно манипулировать событиями или заявлениями, в реальности не существовавшими. Это не только подрывает доверие к средствам массовой информации, но и влияет на политическую ситуацию и общественное восприятие. Более того, сложность различения дипфейков и подлинного контента может привести к общему недоверию ко всему медиаконтенту, что, в свою очередь, может усугубить социальную поляризацию [5].

Помимо воздействия на общество в целом, дипфейки также представляют прямую угрозу для частных лиц. Они могут использоваться для уничтожения репутации, шантажа или в качестве инструмента киберзапугивания. Особую обеспокоенность вызывает использование дипфейков для создания несогласованных порнографических материалов, что оказывает серьезное эмоциональное и психологическое воздействие на жертв [6, 7].

Развитие дипфейк-технологий также представляет собой огромную проблему для правовых систем по всему миру. Многие существующие законы, касающиеся конфиденциальности, клеветы и авторского права, не готовы к конкретным проблемам, которые создают дипфейки [8]. Особый интерес представляет конфиденциальность, поскольку для создания дипфейков персональные данные часто используются без согласия, что поднимает вопросы об этическом использовании данных и необходимости принятия законов или руководящих принципов, регулирующих создание и распространение дипфейков.

Во многих странах предпринимаются усилия по разработке как технических, так и юридических решений для борьбы с проблемами, вызванными дипфейками [9, 10]. При этом важно найти баланс между продвижением инновационных технологий и защитой индивидуальных прав и социальной целостности. Сотрудничество между разработчиками, специалистами по этике, правительствами и общественностью имеет решающее значение для разработки политики, обеспечивающей этическое использование технологии дипфейков.

Разработка этических руководств и кодексов является важным шагом на пути к ответственному использованию технологии дипфейков. Руководящие принципы призваны побудить разработчиков и пользователей учитывать потенциальное влияние этой технологии на отдельных лиц и общество в целом. Они должны включать такие этические принципы,

как прозрачность, справедливость и уважение к конфиденциальности, которые предназначены для контроля за разработкой и использованием дипфейков таким образом, чтобы избежать манипуляций, дезинформации и причинения вреда [11].

Принятие ответственности означает выход за рамки технических аспектов и рассмотрение этических, социальных, психологических и политических последствий создания и распространения дипфейков. Для разработчиков это включает реализацию мер безопасности, которые предотвращают несанкционированное использование их технологий. Для пользователей это означает критический анализ и проверку перед тем, как делиться контентом [12].

Технологии обнаружения дипфейков постоянно развиваются, чтобы идти в ногу с усовершенствованными методами их создания [13]. Для обнаружения дипфейков используются различные инструменты и методы, основанные на искусственном интеллекте [14]. Например, модели глубокого обучения, которые учатся отличать подлинник от подделки на основе больших наборов данных, или технологии блокчейн, которые можно использовать для проверки подлинности цифровых носителей. Эти методы анализируют видео на предмет несоответствий в качестве изображения, движениях лица или синтезе речи, чтобы найти признаки манипулирования [15]. Будущие разработки могут включать усовершенствованные модели искусственного интеллекта и новые подходы к обнаружению дипфейков в реальном времени.

Образование и осведомленность являются значимыми компонентами в борьбе с негативными последствиями дипфейков. Как широкая общественность, так и конкретные целевые группы должны быть информированы о существовании и потенциале дипфейков [16, 17]. Семинары, онлайн-курсы и информационные кампании могут помочь повысить осведомленность и вооружить людей навыками распознавания и правильного обращения с дипфейк-контентом. Обучение критическому мышлению и медиаграмотности имеет важное значение для развития способности распознавать обман и принимать обоснованные решения относительно распространяемого контента.

### Заключение

Центральным этическим аспектом дипфейков является их способность изменять восприятие реальности и серьезно подрывать доверие к информации. Они могут использоваться для распространения ложных заявлений, продвижения политических программ или нарушения прав личности, когда люди могут быть изображены в спорных или компрометирующих ситуациях.

Для эффективного решения проблемы дипфейков необходимо сочетание технологических, правовых и образовательных мер. Передовые методы обнаружения дипфейков и маркировка медиафайлов дадут возможность отличить подлинный контент от поддельного. Принятие законов и правил создаст правовую защиту от злоупотреблений. Информационные кампании и целенаправленное обучение позволят просветить общественность и помогут пользователям ответственно обращаться с медиаконтентом.

### Список источников

1. Воронин И.А., Гавра Д.П. Дипфейки: современное понимание, подходы к определению, характеристики, проблемы и перспективы // Российская школа связей с общественностью. 2024. № 33. С. 28–47. DOI: 10.24412/2949-2513-2023-33-28-47
2. DeepFake and its Enabling Techniques: A Review / B. Rachael [et al.] // APSIPA Transactions on Signal and Information Processing. 2022. Vol. 11. № 2, e40. DOI: 10.1561/116.00000024
3. Westerlund M. The Emergence of Deepfake Technology: A Review // Technology Innovation Management Review. 2019. Vol. 9. № 11. P. 39–52. DOI: 10.22215/timreview/1282
4. Sure T. Ethical Implications of Deepfake Technology in User Interfaces // International Journal of Computer Techniques. 2023. Vol. 10. № 2. P. 1–5.

5. Veljkovic S.Z., Curcic M.T., Gavrilovic I.P. Dark sides of deepfake technology // *Military Technical Courier*. 2024. Vol. 72. № 3. P. 1441–1463. DOI: 10.5937/vojtehg72-49630
6. Порнографический дипфейк: вымысел или виртуальная реальность? / И.Н. Архипцев [и др.] // *Социально-политические науки*. 2021. Т. 11. № 1. С. 69–74. DOI: 10.33693/2223-0093-2021-11-1-69-74
7. Литовченко А.И. Смоделированные изображения порнографического характера как предмет преступлений, связанных с обращением порнографии: компаративистское исследование // *Вестник Южно-Уральского государственного университета*. Сер.: Право. 2024. Т. 24. № 2. С. 27–35. DOI: 10.14529/law240205
8. Добробаба М.Б. Дипфейки как угроза правам человека // *Lex Russica (Русский закон)*. 2022. Т. 75. № 11 (192). С. 112–119. DOI: 10.17803/1729-5920.2022.192.11.112-119
9. Виноградов В.А., Кузнецова Д.В. Зарубежный опыт правового регулирования технологии «дипфейк» // *Право. Журнал Высшей школы экономики*. 2024. № 2. С. 215–240. DOI: 10.17323/2072-8166.2024.2.215.240
10. Regulating deep fakes: legal and ethical considerations / E. Meskys [et al.] // *Journal of Intellectual Property Law & Practice*. 2020. Vol. 15. № 1. P. 24–31. DOI: 10.1093/jiplp/jpz167
11. Kaan Tuysuz M., Kılıç A. Analyzing the Legal and Ethical Considerations of Deepfake Technology // *Interdisciplinary Studies in Society, Law, and Politics*. 2023. Vol. 2. № 2. P. 4–10. DOI: 10.61838/kman.isslp.2.2.2
12. Шляпников В.В. Проблемы навигации в цифровом пространстве: информационная перегрузка и фейковые новости // *Психолого-педагогические проблемы безопасности человека и общества*. 2024. № 4 (65). С. 43–47. DOI: 10.61260/2074-1618-2024-4-43-47
13. Localization and detection of deepfake videos based on self-blending method / J. Xu [et al.] // *Scientific Reports*. 2025. Vol. 15. Art. 3927. DOI: 10.1038/s41598-025-88523-1
14. A systematic literature review on the effectiveness of deepfake detection techniques / L. Stroebel [et al.] // *Journal of Cyber Security Technology*. 2023. Vol. 7. № 2. P. 83–113. DOI: 10.1080/23742917.2023.2192888
15. Alanazi S., Asif S. Exploring deepfake technology: creation, consequences and countermeasures // *Human-Intelligent Systems Integration*. 2024. Vol. 6. P. 49–60. DOI: 10.1007/s42454-024-00054-8
16. Шляпников В.В. Цифровое образование, этика данных и будущее гражданского общества // *Общество. Среда. Развитие*. 2024. № 1 (70). С. 125–128. DOI: 10.53115/19975996\_2024\_01\_125\_128
17. The challenges of media and information literacy in the artificial intelligence ecology: deepfakes and misinformation / A. Sanchez-Acedo [et al.] // *Communication & Society*. 2024. Vol. 37. № 4. P. 223–239. DOI: 10.15581/003.37.4.223-239

## References

1. Voronin I.A., Gavra D.P. Dipfejki: sovremennoe ponimanie, podhody k opredeleniyu, harakteristiki, problemy i perspektivy // *Rossijskaya shkola svyazej s obshchestvennost'yu*. 2024. № 33. S. 28–47. DOI: 10.24412/2949-2513-2023-33-28-47
2. DeepFake and its Enabling Techniques: A Review / B. Rachael [et al.] // *APSIPA Transactions on Signal and Information Processing*. 2022. Vol. 11. № 2, e40. DOI: 10.1561/116.00000024
3. Westerlund M. The Emergence of Deepfake Technology: A Review // *Technology Innovation Management Review*. 2019. Vol. 9. № 11. P. 39–52. DOI: 10.22215/timreview/1282
4. Sure T. Ethical Implications of Deepfake Technology in User Interfaces // *International Journal of Computer Techniques*. 2023. Vol. 10. № 2. P. 1–5.
5. Veljkovic S.Z., Curcic M.T., Gavrilovic I.P. Dark sides of deepfake technology // *Military Technical Courier*. 2024. Vol. 72. № 3. P. 1441–1463. DOI: 10.5937/vojtehg72-49630
6. Pornograficheskij dipfejki: vymysel ili virtual'naya real'nost'? / I.N. Arhipcev [i dr.] // *Social'no-politicheskie nauki*. 2021. Т. 11. № 1. С. 69–74. DOI: 10.33693/2223-0093-2021-11-1-69-74

7. Litovchenko A.I. Smodelirovannye izobrazheniya pornograficheskogo haraktera kak predmet prestuplenij, svyazannyh s obrashcheniem pornografii: komparativistskoe issledovanie // Vestnik Yuzhno-Ural'skogo gosudarstvennogo universiteta. Ser.: Pravo. 2024. T. 24. № 2. S. 27–35. DOI: 10.14529/law240205
8. Dobrobaba M.B. Dipfejki kak ugroza pravam cheloveka // Lex Russica (Russkij zakon). 2022. T. 75. № 11 (192). S. 112–119. DOI: 10.17803/1729-5920.2022.192.11.112-119
9. Vinogradov V.A., Kuznecova D.V. Zarubezhnyj opyt pravovogo regulirovaniya tekhnologii «dipfejk» // Pravo. Zhurnal Vysshej shkoly ekonomiki. 2024. № 2. S. 215–240. DOI: 10.17323/2072-8166.2024.2.215.240
10. Regulating deep fakes: legal and ethical considerations / E. Meskys [et al.] // Journal of Intellectual Property Law & Practice. 2020. Vol. 15. № 1. P. 24–31. DOI: 10.1093/jiplp/jpz167
11. Kaan Tuysuz M., Kılıç A. Analyzing the Legal and Ethical Considerations of Deepfake Technology // Interdisciplinary Studies in Society, Law, and Politics. 2023. Vol. 2. № 2. P. 4–10. DOI: 10.61838/kman.isslp.2.2.2
12. Shlyapnikov V.V. Problemy navigacii v cifrovom prostranstve: informacionnaya peregruzka i fejkovyje novosti // Psihologo-pedagogicheskie problemy bezopasnosti cheloveka i obshchestva. 2024. № 4 (65). S. 43–47. DOI: 10.61260/2074-1618-2024-4-43-47
13. Localization and detection of deepfake videos based on self-blending method / J. Xu [et al.] // Scientific Reports. 2025. Vol. 15. Art. 3927. DOI: 10.1038/s41598-025-88523-1
14. A systematic literature review on the effectiveness of deepfake detection techniques / L. Stroebel [et al.] // Journal of Cyber Security Technology. 2023. Vol. 7. № 2. P. 83–113. DOI: 10.1080/23742917.2023.2192888
15. Alanazi S., Asif S. Exploring deepfake technology: creation, consequences and countermeasures // Human-Intelligent Systems Integration. 2024. Vol. 6. P. 49–60. DOI: 10.1007/s42454-024-00054-8
16. Shlyapnikov V.V. Cifrovoe obrazovanie, etika dannyh i budushchee grazhdanskogo obshchestva // Obshchestvo. Sreda. Razvitie. 2024. № 1 (70). S. 125–128. DOI: 10.53115/19975996\_2024\_01\_125\_128
17. The challenges of media and information literacy in the artificial intelligence ecology: deepfakes and misinformation / A. Sanchez-Acedo [et al.] // Communication & Society. 2024. Vol. 37. № 4. P. 223–239. DOI: 10.15581/003.37.4.223-239

**Информация о статье:** статья поступила в редакцию: 29.11.2025; принята к публикации: 12.12.2025  
**Information about the article:** the article was submitted to the editorial office: 29.11.2025; accepted for publication: 12.12.2025

*Информация об авторе:*

**Шляпников Виктор Валерьевич**, доцент философии и социальных наук Санкт-Петербургского университета ГПС МЧС России (196105, Санкт-Петербург, Московский пр., д. 149), кандидат философских наук, доцент, e-mail: [shlyapnikovv@mail.ru](mailto:shlyapnikovv@mail.ru), <https://orcid.org/0000-0002-6502-5810>, SPIN-код: 4210-5441

*Information about the author:*

**Shlyapnikov Viktor V.**, associate professor of the department of philosophy and social sciences of Saint-Petersburg university of State fire service of EMERCOM of Russia (196105, Saint-Petersburg, Moskovsky ave., 149), candidate of philosophical sciences, associate professor, e-mail: [shlyapnikovv@mail.ru](mailto:shlyapnikovv@mail.ru), <https://orcid.org/0000-0002-6502-5810>, SPIN: 4210-5441