

CASE STUDIES

УДК 1:004:8

DOI: 10.32326/2618-9267-2025-8-1-69-77

О НЕКОТОРЫХ ЭТИЧЕСКИХ АСПЕКТАХ ПРИМЕНЕНИЯ НЕЙРОСЕТЕВОЙ ТЕХНОЛОГИИ DEEPFAKE

Андреева Мария Алексеевна – аспирант кафедры философии и методологии науки философского факультета, Московский государственный университет им. М.В. Ломоносова. Российская Федерация, 119991, г. Москва, Ломоносовский проспект, д. 27, корп. 4; e-mail: aleksing938@mail.ru.

Нейросетевые технологии искусственного интеллекта (далее – ИИ) применяются во всех сферах жизни, от медицины до искусства. Несмотря на широкое распространение этих технологий, с этической точки зрения постоянно возникают новые, до сих пор не имеющие однозначной трактовки вопросы. В последние годы идёт процесс формализации этических норм применения технологий ИИ на уровне международных организаций («Проект рекомендации об этических аспектах искусственного интеллекта» Юнеско), а также на национальных и отраслевых уровнях, не исключены и спорные случаи. Учёные и специалисты по этике из *DeepMind* применяют междисциплинарный подход для разработки искусственного общего интеллекта, подчёркивая важность этики и безопасности в использовании ИИ. Также рассматриваются технологии *DeepFake* (далее – дипфейк), позволяющие создавать реалистичные визуализации, в том числе изображения покойных людей. Несмотря на их потенциал для социально значимого контента, дипфейк представляет угрозу мошенничества и манипуляций. Кроме того, исследуется этичность использования дипфейков в культуре. Этические аспекты использования образов покойных артистов обсуждаются в контексте согласия на использование изображений и уважения к наследию. Автор подчёркивает, что сложно найти баланс между инновациями и этическими нормами в искусстве и технологиях. Новизна исследования связана с актуальной темой этического регулирования технологий дипфейк посредством его положительного использования в социокультурной сфере.

Ключевые слова: этика, искусственный интеллект, нейросети, культура, искусство, дипфейк

Цитирование: Андреева М.А. О некоторых этических аспектах применения нейросетевой технологии DeepFake // Цифровой учёный: лаборатория философа. 2025. Т. 8. № 1. С. 69-77. DOI: 10.32326/2618-9267-2025-8-1-69-77

Рукопись получена: 11 марта 2025

Пересмотрена: 10 апреля 2025

Принята: 12 апреля 2025

ON SOME ETHICAL ASPECTS OF USING DEEPAKE NEURAL NETWORK TECHNOLOGY

Mariia A. Andreeva – post-graduate student of the Department of Philosophy and Methodology of Science at the Faculty of Philosophy, Lomonosov Moscow State University. 27 bld. 4 Lomonosovsky Ave., Moscow 119991, Russian Federation;
e-mail: aleksing938@mail.ru

Artificial intelligence (hereinafter referred to as AI) and neural technologies are used in all spheres of life, from medicine to art. Despite the widespread use of these technologies, new ethical issues constantly arise that have not yet been clearly interpreted. Despite the fact that in recent years there has been a process of formalizing ethical standards for the use of AI technologies at the level of international organizations (UNESCO's Draft Recommendation on the Ethical Aspects of Artificial Intelligence), as well as national and industry levels, controversial cases may still occur. Scientists and ethicists from DeepMind take an interdisciplinary approach to developing artificial general intelligence, emphasizing the importance of ethics and safety in the use of AI. This paper also considers DeepFake technologies, which allow the creation of realistic visualizations, including images of deceased people. Despite their potential for socially significant content, deepfakes pose a threat of fraud and manipulation. In addition, this study examined the ethics of using deepfakes in culture. The ethical aspects of using images of deceased artists are discussed through the prism of consent to use of images and respect for heritage. The article highlights the difficulty of finding a balance between innovation and ethics in art and technology. The novelty of the research is related to the updated topic of ethical regulation of DeepFake technologies through its positive use in the sociocultural sphere.

Keywords: ethics, artificial intelligence, neural networks, culture, art, deepfake

How to cite: Andreeva, M.A. (2025). On some ethical aspects of using DeepFake neural network technology. *The Digital Scholar: Philosopher's Lab*, 8 (1): 69-77. DOI: 10.32326/2618-9267-2025-8-1-69-77 (In Russian).

Received: 11 March 2025

Revised: 10 April 2025

Accepted: 12 April 2025

Этическая сторона применения нейросетевых технологий искусственного интеллекта (далее – ИИ) находится в поле внимания как академических исследований, так и технологических компаний. Так, этическими проблемами ИИ занимается коллектив *DeepMind* (часть корпорации *Google*) – учёные, инженеры, специа-

листы по этике, которые используют междисциплинарный подход, объединяющий новые идеи и достижения в области машинного обучения, нейронауки, инженерии, математики, моделирования и вычислительной инфраструктуры, а также новые способы организации научной деятельности. Цель объединения состоит в том, чтобы разрабатывать более общие и эффективные системы решения проблем, известные как искусственный общий интеллект (*AGI* — *artificial general intelligence*) (Bergman, Stryker, 2024, web). Именно эта команда создала программу *AlphaGo*, обыгравшую гроссмейстера Ли Седоля в игру го. Также её сотрудники добились первых успехов в компьютерных играх, которые исследователи часто используют для тестирования ИИ. Одна из программ научилась играть в 49 различных игр *Atari* с нуля, просто воспринимая пиксели и счёт на экране. Объединение придаёт большое значение этическим вопросам ИИ и безопасности. По их мнению, ИИ должен использоваться в общественно полезных целях и всегда оставаться под человеческим контролем (Iason, Manzini, 2024). Этический отдел *DeepMind* решает вопросы, включающие возможность гарантировать то, что ценности, разработанные для систем ИИ, не противоречат ценностям самого общества. Системы ИИ, возможно, могут сделать общество более справедливым (*AI principles*, 2024, web). Вместе с тем необходимо учитывать, что разные группы людей придерживаются разных ценностей, а это означает, что договориться об универсальных принципах будет сложно. Согласно рекомендациям ЮНЕСКО (*Draft Recommendations*, 2021, web), необходимо признавать воздействие культурного плюрализма, учитывать различие систем ценностей при обсуждении вопросов, связанных как с позитивным, так и негативным воздействием ИИ.

Осмысление этического влияния технологических инноваций на человека в академическом поле началось ещё в прошлом веке – в 1985 г. Один из первых теоретиков исследований в этой сфере Д. Мур пишет о том, что компьютерная этика должна включать «анализ природы и социального влияния компьютерных технологий, а также заниматься формулировкой и обоснованием политики этического использования этих технологий» (Moog, 1985, p. 266). Часто технологическое развитие и применение новых технологий происходят существенно быстрее, чем осуществляется философская рефлексия проблематики, что приводит к риску их неэтичного использования. Философ и исследователь технологий ИИ Н.Ю. Клюева отмечает риск, связанный с злонамеренным использованием технологии *DeepFake* (далее – дипфейк), – за счёт качественно сгенерированной речи и голоса можно достаточно легко обмануть человека, и более глобально – манипулировать общественным мнением, распространяя дипфейки посредством сети Интернет (Клюева, 2021, с. 54).

Дипфейк считается одним из самых опасных способов использования ИИ. Большинство его реальных применений имеют либо дискредитирующие, либо мошеннические намерения. Один из пер-

вых случаев мошенничества с использованием дипфейка произошёл в Великобритании. Мошенники позвонили генеральному директору британской энергетической компании и, подделав голос руководителя, приказали ему перевести 220 тыс. евро на банковский счёт третьего лица (Острожно, DeepFake, 2024, web). Такие случаи учащаются, так как результат технологии может быть неотличимым от реальности.

Рассмотрим ряд примеров использования технологии. На её основе можно создавать детализированные визуализации и голограммы, в том числе умерших людей. Благодаря этому вновь «выступают» и «гастролируют» такие артисты, как рэп-исполнитель Тупак, певец Майкл Джексон и Михаил Горшенев (солист российской рок-группы «Король и Шут»), создаются рекламные ролики, затрагивающие социальные темы. Термин «дипфейк» представляет собой комбинацию двух слов: «глубокий» и «фальшивый». Под словом «глубокий» здесь подразумевается технология ИИ, известная как машинное обучение. Технология дипфейк используется в синтетических медиа для создания фальсифицированного контента, замены или синтеза лиц и речи, манипулирования эмоциями. Дипфейк используется для цифровой имитации действий человека, которые тот не совершал, и работает на предварительно загруженном массиве данных. Кроме того, часто используется технология GAN (Science & Tech, 2020). GAN (генеративно-состязательная сеть) представляет собой алгоритмы машинного обучения, построенные на комбинации двух нейронных сетей. Одна генерирует образцы, а другая стремится отличить правильные образцы от неправильных (Overview of Gan, 2025, web). GAN способна улучшать уже существующие изображения и создавать новые, достаточно натуралистичные.

Дипфейки используются также в сфере культуры. Они, помимо создания основ для голограмм, применяются в изобразительном искусстве, видео-арте, кинематографе и даже в пространстве социальных сетей. В этом случае дипфейк мы представляем как цифровую персону, являющуюся «набором информации, который когнитивная и сенсорная механика человека преобразует в индивидуального актёра, персонажа» (Käsk, 2023, web). Одно из популярных применений дипфейка в искусстве – заставить говорить знаменитые портреты. В музее Дали в Сент-Питерсберге штата Флорида, в трёх киосках, демонстрировалась цифровая голограмма Сальвадора Дали, которая потрясает своим внешним видом и эксцентричной манерой поведения. Цифровой Дали кратко рассуждает на 125 убедительно озвученных тем (Ask Dali, 2024, web): о погоде, о своём искусстве, о бессмертии – «Я не верю в свою смерть. А вы?». Директор музея Хэнк Хайн заявил, что все, кто имел дело с голограммами мёртвых, сталкиваются с этическим вопросом: «Как мы на самом деле репрезентируем этого человека?». Всё, что Дали говорит об искусстве и о себе самом, взято дословно из интервью и книг. Но важное отличие состоит в том, что музей не пытался

раскрыть личность Дали целиком, а лишь стремился познакомить нас с образом мышления гения. При этом большинство голограммных туров имитируют полноценный концертный опыт с артистом, что представляет особую сложность, так как важно передать особенности мимики, голоса, движения, тем самым создавая образ, который должен быть пугающе неотличимым от живого артиста. Голографические изображения используются в парках развлечений, рекламе, кинематографе, а теперь и в музыкальной индустрии. Выступления мёртвых музыкантов вызывают крайне неоднозначную реакцию.

В музыкальной индустрии практически все используют термин «голограмма», что не совсем верно. Настоящая голограмма – это пересечение света и материи, которое создаёт трёхмерный объект, предназначенный для просмотра со всех сторон. Сегодняшние музыкальные голограммы – это цифровые творения, проецируемые на экран или сетку и больше похожие на видео. Главный директор компании *Facebank Group* Джон Текстор занимался созданием голограмм артистов, в том числе и голографическим изображением рэп-исполнителя Тупака, которое создавалось путём объединения компьютерной графики (*CGI*) с двойником и театральным трюком XIX в., известным как «Призрак Пеппера». Голограмма Тупака стала открытием и сенсацией, но после фестиваля Коачелла так и не отправилась в тур. Гастрольное производство голограмм без правильной творческой и этической концепции может быть нелогичным и бесполезным использованием технологии. После нескольких песен зрителю становится скучно. Уже известны случаи отмены туров с артистами-голограммами из-за негативной реакции публики. Компания, разрабатывающая голографический тур Эми Уайнхаус, отложила проект (Snapes, 2019, web). В заявлении *Vase Hologram* говорится, что компания столкнулась с «некоторыми этическими проблемами», несмотря на желание уважительно и тактично использовать образ и наследие певицы. Основательница платформы виртуальной реальности *Limina Immersive*, эксперт по виртуальной реальности и её этике Кэтрин Аллен заявила о том, что теперь согласие на использование голограмм становится особенной темой. При согласии живого артиста юридические трудности не возникают, в то время как для использования голограмм усопших деятелей искусства требуется разрешение от их наследников. Современные зарубежные артисты испытывают противоречивые чувства, когда речь заходит о возможности воссоздания их образа и личности в виде голограммы. Актёр и сценарист Робин Уильямс рассматривал идею посмертного возрождения, но перед смертью решил передать права фонду, базирующемуся в Калифорнии, и запретил коммерческое использование своего изображения до 2039 г. (Setty, 2023, web). В то же время музыкант Мик Джаггер из *Rolling Stones* открыт для идеи посмертного голографического тура (Shah, 2023, web).

Таким образом, использование голограммы вместо живых артистов является неоднозначным. С одной стороны, когда мы отправляемся на концерт, то желаем услышать настоящую музыку, а не запись. Иначе не было бы частых скандалов по поводу выступления живых артистов под фонограмму. С другой стороны, голограмма – это один из способов актуализировать воспоминания о прошлом, о том, как приятно было находиться на выступлении любимого артиста, и пережить эти моменты вновь.

С социальной точки зрения, при грамотном использовании, дипфейки могут принести пользу. Так, в 2020 г. в США главным героем рекламного ролика, посвящённого стрельбе в школах и запрету свободного владения огнестрельным оружием, стал 17-летний подросток Хоакин Оливер, убитый во время расстрела учеников школы в Паркленде в 2018 г. С помощью технологий ИИ удалось создать убедительную визуализацию – были реконструированы в том числе движения и мимика усопшего. Голографическое изображение школьника призывает зрителей голосовать против массового распространения огнестрельного оружия в США. Создание видеоролика стало возможным благодаря согласию родителей мальчика – они надеялись, что таким образом смогут привлечь внимание к проблеме, которая распространилась во всём мире. Объединение *Change the Ref*, создавшее это видео, использует искусство как способ борьбы с катастрофическими последствиями подобных трагедий (*Change the Ref*, 2021, web). Изображение ученика весьма резкое и угловатое, движения рваные – очевидно, информации для реконструкции при создании было немного. Тем не менее подобные ролики не оставляют равнодушными общественность и оказывают сильное эмоциональное воздействие на эмпатичных зрителей.

Относительно применения голограмм существует широкий спектр мнений, которые часто зависят от контекста и условий использования технологии. Например, если видеопроекция применяется временно для акцентирования внимания на социальной теме или для дополнения художественного образа, то она может эффективно пробуждать интерес зрителей. Однако в случае длительных концертов и гастролей однообразные механистические движения и повторяющаяся фоновая запись могут вызывать утомление и скуку у аудитории. В результате сегодня такие подходы к использованию голограмм часто оказываются неудачными и не способными удерживать внимание зрителей продолжительное время. Кроме того, некоторые зрители выражают опасения, что такие технологии могут негативно сказаться на психическом здоровье, и ставят под сомнение их потенциальную финансовую выгоду. Психологи предупреждают о рисках, связанных с созданием иллюзии присутствия умерших людей в реальной жизни, что может привести к сложным эмоциональным реакциям. Например, возвращение покойных актёров на экраны часто вызывает у поклонников не восторг, а подсознательное отторжение. Это свидетель-

ствует о глубоком внутреннем конфликте, сложности принятия утраты и создаёт новую травму. Надеемся, что этичное использование дипфейков будет сопровождать создание видеоконтента и поможет усилить авторское высказывание, а также появятся юридические нормы публичного применения этой технологии.

Информация о конфликте интересов

Автор заявляет об отсутствии конфликта интересов

Declaration of Conflicting Interests

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Список литературы

Клюева, 2021 – Клюева Н.Ю. Этико-прикладные аспекты применения технологий искусственного интеллекта // Вестник Московского университета. Серия 7. Философия. 2021. № 5. С. 52-66.

Осторожно, deepfake, 2024, web – Осторожно, deepfake. Как распознать мошенника на видео // Сайт СберБанк. URL: <https://www.sberbank.ru/ru/person/kibrary/articles/ostorozhno-dipejfk-kak-uznat-moshennika-na-video> (дата обращения: 11.03.2025).

AI principles, 2024, web – AI principles // AI Google. 2024. URL: <https://ai.google/responsibility/principles/> (дата обращения: 11.03.2025).

Ask Dali, 2024, web – Ask Dali (via artificial intelligence) // Сайт Dali Home. URL: <https://thedali.org/exhibit/ask-dali/> (дата обращения: 11.03.2025).

Bergman, Stryker, 2024, web – *Bergman D., Stryker C.* What is artificial general intelligence (AGI)? // IBM. 16 September 2024. URL: <https://www.ibm.com/think/topics/artificial-general-intelligence> (дата обращения: 11.03.2025).

Change, 2021, web – Change the ref – the unfinished votes // Campaigns of the World. 26 December 2021. URL: <https://campaignsoftheworld.com/tech-innovations/the-unfinished-votes/> (дата обращения: 11.03.2025).

Draft Recommendation, 2021, web – Draft recommendation of the ethics of artificial intelligence // UNESCO. 2021. URL: <https://unesdoc.unesco.org/ark:/48223/pf0000378931> (дата обращения: 11.03.2025).

Iason, Manzini, 2024, web – *Iason G., Manzini A.* The ethics of advanced AI assistants // Google DeepMind. 19 March 2024. URL: <https://deepmind.google/discover/blog/the-ethics-of-advanced-ai-assistants/> (дата обращения: 11.03.2025).

Käsk, 2023, web – *Käsk H.* Understanding our digital existence: digital identity versus digital persona // Agrello. 18 May 2023. URL: <https://www.agrello.io/post/understanding-our-digital-existence-digital-identity-versus-digital-persona> (дата обращения: 11.03.2025).

Moor, 1985 – *Moor J.H.* What is computer ethics? // *Metaphilosophy*. 1985. Vol. 16. No. 4. Pp. 266-275.

Overview of GAN, 2025, web – Overview of GAN structure // Generative Adversarial Networks. 2025. URL: https://developers.google.com/machine-learning/gan/gan_structure (дата обращения 08.04.2025).

Science & Tech, 2020, web – Science & tech spotlight: deepfakes // U.S. Government Accountability Office. 20 February 2020. URL: <https://www.gao.gov/products/gao-20-379sp> (дата обращения: 11.03.2025).

Setty, 2023, web – Setty R. John Lennon AI revival adds to debate over IP rights of the dead // Bloomberg Law. 15 September 2023. URL: <https://news.bloomberglaw.com/ip-law/john-lennon-ai-revival-adds-to-debate-over-ip-rights-of-the-dead> (дата обращения: 11.03.2025).

Shal, 2023, web – Shal N. How Mick Jagger has kept the Rolling Stones in business for six decades // WALLST.J. 26 September 2023. URL: <https://www.wsj.com/style/mick-jagger-rolling-stones-hackney-diamonds-7d4f193b> (дата обращения: 11.03.2025).

Snapes, 2019, web – Snapes L. Amy Winehouse hologram tour postponed due to “unique sensitivities” // The Guardian. 28 March 2019. URL: <https://www.theguardian.com/music/2019/feb/22/amy-winehouse-hologram-tour-postponed#:~:text=The%20company%20developing%20an%20Amy,celebratory%20and%20respectful%20way%20possible> (дата обращения: 11.03.2025).

References

Ask Dali (via artificial intelligence) (2024). Dali Home. Retrieved March 11, 2025, from <https://thedali.org/exhibit/ask-dali/>

AI principles (2024). AI Google. Retrieved March 11, 2025, from <https://ai.google/responsibility/principles/>

Bergman, D., Stryker, C. (2024). What is artificial general intelligence (AGI)? IBM, 16 September. Retrieved March 11, 2025, from <https://www.ibm.com/think/topics/artificial-general-intelligence>

Careful, deepfake. How to recognize a scammer in a video (2024). SberBank. Retrieved March 11, 2025, from <https://www.sberbank.ru/ru/person/kibrary/articles/ostorozhno-dipfejk-kak-uznat-moshennika-na-video> (In Russian)

Draft Recommendation of the Ethics of Artificial Intelligence (2021). UNESCO. Retrieved March 11, 2025, from <https://unesdoc.unesco.org/ark:/48223/pf0000378931>

Iason, G, Manzini, A. (2024). The ethics of advanced AI assistants. Google DeepMind, 19 April. Retrieved March 11, 2025, from <https://deepmind.google/discover/blog/the-ethics-of-advanced-ai-assistants/>

Käsk, H. (2023). Understanding our digital existence: digital identity versus digital persona. Agrello, 18 May. Retrieved March 11, 2025, from <https://www.agrello.io/post/understanding-our-digital-existence-digital-identity-versus-digital-persona>

Klyueva, N. Yu. (2021). Applied ethical issues of artificial intelligence. *Bulletin of Lomonosov Moscow State University. Series 7. Philosophy*, 5, 52-66. (In Russian)

Moor, J. H. (1985). What is computer ethics? *Metaphilosophy*, 16 (4), 266-275.

Overview of GAN structure (2025). *Generative Adversarial Networks*. Retrieved April 9, 2025, from https://developers.google.com/machine-learning/gan/gan_structure

Science & tech spotlight: deepfakes (2020). U.S. Government Accountability Office, 20 February. Retrieved March 11, 2025, from <https://www.gao.gov/products/gao-20-379sp>

Setty, R. (2023). John Lennon AI revival adds to debate over IP rights of the dead. *Bloomberg Law*, 15 June. Retrieved March 11, 2025, from <https://news.bloomberglaw.com/ip-law/john-lennon-ai-revival-adds-to-debate-over-ip-rights-of-the-dead>

Shah, N. (2023). How Mick Jagger has kept the Rolling Stones in business for six decades. *Wall Street Journal*, 26 September. Retrieved March 11, 2025, from <https://www.wsj.com/style/mick-jagger-rolling-stones-hackney-diamonds-7d4f193b>

Snapes, L. (2019). Amy Winehouse hologram tour postponed due to “unique sensitivities”. *The Guardian*, 28 March. Retrieved March 11, 2025, from <https://www.theguardian.com/music/2019/feb/22/amy-winehouse-hologram-tour-postponed#:~:text=The%20company%20developing%20an%20Amy,celebratory%20and%20respectful%20way%20possible>”

Change the ref – the unfinished votes (2021). *Campaigns of the World*, 26 December. Retrieved March 11, 2025, from <https://campaignsoftheworld.com/tech-innovations/the-unfinished-votes/>