

СОЦИАЛЬНАЯ ФИЛОСОФИЯ

DOI: 10.17212/2075-0862-2026-18.2.1-109-125

УДК 316.77

Дипфейки как социально-политический феномен

Шляпников Виктор Валерьевич,

кандидат философских наук, доцент,

доцент кафедры философии и социальных наук

Санкт-Петербургского университета ГПС МЧС России,

Россия, 196105, Санкт-Петербург, Московский пр., 149

ORCID: 0000-0002-6502-5810

shlyapnikovv@mail.ru

Аннотация

В настоящее время аудиовизуальным контентом можно быстро и убедительно манипулировать с помощью генеративного искусственного интеллекта. Создаются так называемые дипфейки, т. е. изображения, видео и звуковые дорожки, которые выглядят или звучат так, как будто люди сделали или сказали то, чего они никогда не делали или не говорили. Автор констатирует, что дипфейки вызывают серьезные проблемы с политическим и общественным дискурсом, подрывают доверие к информационному контенту и политическим институтам, способствуя тем самым поляризации общества и ослаблению социальной сплоченности. В то же время дипфейк-технологии предлагают увлекательные перспективы для новых форм историко-политического образования и творческого самовыражения. Индустрия развлечений экспериментирует с дипфейками для снижения производственных затрат и создания инновационных эффектов. В рекламной отрасли применение дипфейк-технологий позволяет создавать виртуальных инфлюенсеров и открывает новые возможности для персонализированного, в том числе социально-политического, взаимодействия. Автор приходит к выводу, что существует необходимость реагирования на вызовы, связанные с дипфейками, но при этом важно использовать возможности, которые они открывают. Для достижения положительного результата необходимо объединение усилий различных социальных, политических и институциональных субъектов. Важную роль могут сыграть политики, разрабатывая соответствующие нормативные предложения на национальном и межгосударственном уровнях, направленные, в частности, на регулирование деятельности операторов крупных социальных сетей. Однако и сами сети посредством саморегулирования могут взять на себя ответственность и способствовать пресечению использования дипфейков в преступных или неэтичных целях. Разработчики могут совершенствовать технологии обнаружения дипфейков и маркировки контента, сгенерированного искусственным интеллектом, а средства массовой информации – использовать эффективные инструменты фактчекинга. Кроме того, необходимо грамотное гражданское общество, которое, учитывая медиареальность,

измененную дипфейками, особенно критично относится к медиаконтенту, происхождение и достоверность которого не были проверены заслуживающими доверия медиаорганизациями.

Ключевые слова: искусственный интеллект, нейронные сети, дипфейки, медиаконтент, социальные сети, социально-политический дискурс.

Deepfakes as a Socio-Political Phenomenon

Viktor Shlyapnikov,

Cand. Sc. (Philosophy), Associate Professor,

Associate Professor at the Department of Philosophy and Social Sciences,

Saint-Petersburg University of State Fire Service of EMERCOM of Russia,

149, Moskovskiy Ave., St. Petersburg, 196105, Russian Federation

ORCID: 0000-0002-6502-5810

shlyapnikovv@mail.ru

Abstract

Nowadays, audiovisual content can be manipulated quickly and convincingly using generative artificial intelligence. So-called deepfakes are created, i.e. images, videos, and audio tracks that look or sound as if people did or said something they never did or said. The author notes that deepfakes cause serious problems for political and public discourse, undermine trust in information content and political institutions, thereby contributing to the polarization of society and the weakening of social cohesion. At the same time, deepfake technologies offer exciting prospects for new forms of historical and political education and creative expression. The entertainment industry is experimenting with deepfakes to reduce production costs and create innovative effects. In the advertising industry, the use of deepfake technologies allows the creation of virtual influencers and opens up new possibilities for personalized, including socio-political, interaction. The author concludes that there is a need to respond to the challenges posed by deepfakes, but also to seize the opportunities they offer. Achieving a positive outcome requires the combined efforts of various social, political, and institutional actors. Policymakers can play an important role by developing appropriate regulatory proposals at the national and international levels, in particular to regulate the activities of large social media operators. However, the networks themselves, through self-regulation, can also take responsibility and help prevent the use of deepfakes for criminal or unethical purposes. Developers can improve deepfake detection and AI-generated content labeling technologies, and media outlets can use effective fact-checking tools. In addition, a literate civil society is needed that, given the media reality altered by deepfakes, is particularly critical of media content whose origin and veracity have not been verified by credible media organizations.

Keywords: artificial intelligence, neural networks, deepfakes, media content, social media, socio-political discourse.

Библиографическое описание для цитирования:

Шляпников В.В. Дипфейки как социально-политический феномен // Идеи и идеалы. – 2026. – Т. 18, № 2, ч. 1. – С. 109–125. – DOI: 10.17212/2075-0862-2026-18.2.1-109-125.

Shlyapnikov V. Deepfakes as a Socio-Political Phenomenon. *Idei i idealy = Ideas and Ideals*, 2026, vol. 18, iss. 2, pt. 1, pp. 109–125. DOI: 10.17212/2075-0862-2026-18.2.1-109-125.

Введение

Технология, лежащая в основе дипфейков, стремительно совершенствовалась с тех пор, как она впервые привлекла внимание общественности в 2017 году. С момента появления генераторов аудио и видеоизображений пользователи, даже не имеющие технических знаний, получили доступ к алгоритмам и инструментам, которые позволяют им с минимальными затратами создавать реалистичные изображения людей, звуковые дорожки человеческих голосов и приемлемые видеоролики. Таким образом, качество и доступность дипфейков растут, а следовательно, растут и их количество. Это в сочетании с динамикой распространения дезинформации в медиапространстве и особой силой воздействия аудиовизуальных произведений все больше затрудняет для зрителей и слушателей возможность отличить подлинный контент от поддельного, что, в свою очередь, открывает новые возможности для манипулирования общественным сознанием.

Технологические основы и тенденции развития дипфейков

Искусственный интеллект (ИИ) – это раздел компьютерных наук, который занимается разработкой алгоритмов, имитирующих когнитивные способности человека, такие как логическое мышление, обучение, планирование и творчество. ИИ может анализировать большие объемы данных, например текст, изображения или аудиосигналы, распознавать в них закономерности и извлекать из них информацию. В 1950-е годы были заложены теоретические основы современных компьютерных технологий, а также появились первые соображения о том, как создавать алгоритмы, которые учатся самостоятельно решать проблемы, например, посредством повторяющегося процесса «проб и ошибок» [6].

Для понимания ИИ важно понятие алгоритма. Алгоритм описывает четко определенную последовательность конечного числа шагов или инструкций для решения конкретной проблемы или выполнения задачи. Алгоритмы получают входные данные, обрабатывают их в соответствии с определенными правилами и возвращают выходные данные. Особенность ИИ заключается в том, что он представляет собой группу алгорит-

мов, которые могут в определенной степени автоматически создавать дополнительные алгоритмы и таким образом генерировать заданные последовательности действий. Если некоторое время назад ИИ позволял решать только очень специфические задачи в конкретной области, то сегодня алгоритмы ИИ становятся все более адаптируемыми и могут использоваться в разных областях. Теперь один и тот же ИИ может, например, планировать поездки по городу, программировать, решать домашние задания. За последние 75 лет появилось несколько типов искусственного интеллекта и определенные подходы стали общепринятыми. Два технических термина, которые в последние годы привлекли большое внимание, – это «машинное обучение» и «глубокое обучение» [9, 10].

Машинное обучение – это подраздел ИИ. В 1950-е годы идея ИИ была скорее абстрактной и теоретической моделью, но в 1960-е годы она впервые была реализована на практике с помощью концепций машинного обучения, которые используют алгоритмы, обучающиеся на основе набора данных и делающие выводы о статистических вероятностях.

Глубокое обучение – это подраздел машинного обучения, представляющий собой процесс обучения глубоких искусственных нейронных сетей. Слово «глубокий» в этом контексте означает, что множество искусственных нейронов взаимосвязано. Их функционирование смоделировано по образцу человеческого мозга и было теоретически описано еще в 1940–50-е годы. В области машинного обучения нейронные сети долгое время были одним из многих инструментов, но результаты исследований, особенно после 2016 года, доказали их потенциал и привели к широкому применению. Прорывом стала возможность масштабировать сети до любого размера с помощью аппаратного и программного обеспечения. Современные системы глубокого обучения (т. е. нейронные сети) превосходят предыдущие подходы из области искусственного интеллекта по точности высказываний и скорости генерации ответов. Нейронные сети могут распознавать изображения, писать тексты, генерировать изображения или видео и многое другое. Когда сегодня говорят о генеративном ИИ, чаще всего имеют в виду нейронные сети.

Можно сказать, что ИИ, и в частности нейронные сети, предоставляет широкий спектр возможностей. ИИ способен обнаруживать закономерности в больших объемах данных, классифицировать их, синтезировать новые данные. Примечательно, что тип данных, обрабатываемых нейронными сетями, не имеет значения. Текст, изображение, звук или видео – все в конечном итоге кодируется в битах и байтах, и архитектура нейронной сети просто должна быть способна обрабатывать соответствующий объем данных. Таким образом, сеть, генерирующая текст, может быть адаптирована для генерации изображений. Для ИИ одно слово в тексте – это

то же самое, что один пиксель в изображении. ИИ вычисляет вероятности того, где в тексте или изображении определенное слово или пиксель лучше всего способствуют решению проблемы. Эта аналогия может быть произвольно перенесена на другие типы данных, что делает ИИ высокоадаптивным и является причиной того, что эта технология уже присутствует во многих областях нашей повседневной жизни. Генерация текста, синтез изображений, автономное вождение и дипфейки – это лишь некоторые из сегодняшних применений ИИ.

Под дипфейками (англ. deepfake от deep learning «глубокое обучение» + fake «подделка») понимаются реалистично выглядящие или звучащие, но поддельные изображения, видео- или аудиозаписи. Дипфейки впервые появились на платформе Reddit, где в конце 2017 года пользователь под псевдонимом deepfakes опубликовал сфальсифицированные порнографические видео с лицами известных актрис [17]. В 2020 году было выпущено программное обеспечение с открытым исходным кодом, которое позволяет производить высококачественные дипфейк-видео бесплатно. В настоящее время существует несколько компаний, которые предлагают производство видео- и аудиодипфейков за плату. Большинство профессиональных поставщиков имеют политики безопасности, согласно которым не каждый человек, голос или изображение могут быть синтезированы без согласия автора, но в этом механизме есть пробелы.

В связи с достижениями в развитии нейронных сетей наблюдается растущая тенденция к постоянному повышению качества систем синтеза аудио-, фото- или видеодипфейков, когда и людям, и даже алгоритмам становится все труднее идентифицировать этот контент как поддельный. Кроме того, независимые разработчики, которые часто известны только под псевдонимами, выпускают системы без каких-либо барьеров безопасности. Еще одна тенденция – сокращение времени генерации аудио- и видеоконтента, вплоть до возможности работы в режиме реального времени, что значительно упрощает проведение телефонных разговоров или видеоконференций с использованием дипфейков.

Социально-политические аспекты использования дипфейков

Общественные, политические, медийные и научные дебаты о дипфейках в основном сосредоточены на потенциале их использования для политической манипуляции и дезинформации [16]. Эта проблема не является абсолютно новой. Манипулирование средствами массовой информации так же старо, как и сами СМИ. Фотографии обрабатывались в пропагандистских целях задолго до появления программы Photoshop. Однако ИИ-технологии сделали манипулирование средствами массовой информации гораздо более быстрым, простым, доступным и убедительным. Раньше об-

работка аудио- и видеоматериалов требовала много времени и ресурсов. Теперь это можно сделать одним нажатием кнопки, что приводит к значительному росту количества политических дипфейков. В то же время люди по-прежнему с большим доверием относятся к аудиовизуальным средствам массовой информации, а дипфейки быстро охватывают огромную аудиторию через социальные сети, знаменуя тем самым не только новое количество, но и новое качество политической манипуляции.

Для эффективного функционирования политического дискурса крайне важно, чтобы люди могли договариваться, общаться и обсуждать фундаментальные вещи. Однако дипфейки сеют сомнения в правдивости политических новостей, так что дискуссии часто вязнут в обсуждении фактической точности контента вместо того, чтобы способствовать формированию общественного мнения и принятию социальных решений [12]. Но для чего именно используются дипфейки в политической сфере, насколько они опасны и насколько опасными могут стать?

Одно из самых серьезных опасений заключается в том, что качественные дипфейки могут исказить результаты выборов, дискредитируя кандидатов или ослабляя доверие к выборам в целом. В последние годы во всем мире растет число примеров использования дипфейков для манипулирования электоральным поведением [19]. Однако представляется, что дипфейки на сегодняшний день пока не оказали решающего влияния ни на одни выборы. Дипфейки в отношении влиятельных политиков, как правило, быстро разоблачаются. Тем не менее они могут нанести ущерб репутации, а разоблачения часто оказываются малоэффективными. Дипфейки ослабляют веру избирателей в достоверность информации и честность выборов, потенциально снижая их участие в политической жизни.

Но даже и вне выборного процесса дипфейки могут использоваться для атак на политических оппонентов с целью заставить замолчать отдельных лиц или группы, расколоть общество или разжечь ожесточенный конфликт. Кроме того, сами кандидаты, партии и предвыборные команды все чаще используют дипфейки для собственной рекламы, охвата большей или новой аудитории, распространения содержания своих предвыборных программ, представления себя в благоприятном свете и улучшения имиджа [13].

Дипфейки также могут использоваться для военной пропаганды и контрпропаганды. Созданные ИИ изображения, а также аудиоподделки с выступлениями различных знаменитостей могут служить интересам соответствующих воюющих сторон, способствуя росту недоверия к свидетельствам СМИ о войне [2]. Дипфейки, связанные с иностранным влиянием, могут подорвать национальную безопасность и суверенитет государств. Они также могут служить, особенно в контексте выборов или войн, для

распространения неопределенности, ослабления доверия населения к политике и демотивации союзников.

Потеря доверия, вызванная дипфейками, и растущая неуверенность населения в правдивости аудиовизуальных медиа позволяют получить так называемые «дивиденды лжеца»: публичные деятели, подвергающиеся критике, могут заявить, что компрометирующие их доказательства являются поддельными [8]. Ссылаясь на само существование дипфейков, можно поставить под сомнение достоверность подлинных материалов. Политические деятели могут пользоваться этим, чтобы уйти от обвинений и ответственности. «Дивиденды лжеца» создают растущие проблемы не только для граждан, но и для СМИ как органов общественного контроля. Независимо от конкретных дипфейков, они подрывают авторитет и подотчетность избранных представителей, а также качество политических дебатов.

Необходимо отметить, что сегодня много говорится о политических опасностях и потенциальном вреде дипфейков, при этом различные другие возможности этой технологии часто игнорируются. Например, образовательные учреждения и музеи могут использовать дипфейки, чтобы сделать учебный контент более увлекательным или даже захватывающим, а также открыть новые возможности для изучения исторических фактов. Дипфейки, особенно для историко-политического образования, обладают огромным потенциалом для раскрытия важных исторических свидетельств и материалов. В музеях дипфейки известных личностей или очевидцев событий могут не только говорить, но и отвечать на вопросы посетителей. Дипфейки также могут усилить ощущение погружения в историю [15].

Таким образом, исторические дипфейки позволяют по-новому взглянуть на историю и могут вдохновить больше людей. Они также могут побудить задуматься о значении исторических событий и опыта для настоящего, поставить под сомнение текущие реалии и принять участие в политической жизни. Однако крайне важно, чтобы дипфейки были четко обозначены как таковые в рамках историко-политического образования, чтобы избежать манипуляций и дезинформации. Кроме того, дипфейки давно ушедших людей поднимают ряд этических и правовых вопросов, например в какой степени они нарушают достоинство и личные права умерших или их родственников и какое значение имеет их собственное согласие на такие дипфейки при жизни или согласие родственников после их смерти.

Интерактивные и иммерсивные дипфейковые опыты всегда представляют собой лишь одну из интерпретаций истории и могут упускать из виду существенные нюансы и различные точки зрения. В соответствии

с этим важно обеспечить, чтобы опыт работы с дипфейками в рамках историко-политического образования позволял критически и сбалансированно рассматривать историю. Также важно, чтобы дипфейки всегда правильно классифицировались, особенно если они нацелены на детей и молодежь.

Использование дипфейков в образовательном контексте может не ограничиваться только историко-политическим образованием. Их использование также возможно и в сфере собственно политического образования, например для иллюстрации эффектов и механизмов дезинформации и манипулирования.

Дипфейки открывают новые формы создания контента в искусстве. Художники все чаще используют дипфейки и особенно генераторы изображений на основе искусственного интеллекта в качестве нового средства самовыражения [23]. Некоторые художники также используют дипфейки, чтобы привлечь внимание к социальным проблемам. В политическом искусстве и сатире дипфейки могут осуждать, оспаривать и подвергать сомнению существующие властные отношения, предлагать альтернативные варианты действий, а также усиливать голос и влияние социально маргинализированных групп.

Неправительственные организации, государственные органы, политические группы и движения могут использовать дипфейки для рекламных и лоббистских целей, а также для информирования и просвещения [24]. Вполне возможно, что в будущем просвещающие и информирующие правительственные и неправительственные дипфейки станут «гиперперсонализированными». Рекламная индустрия уже использует дипфейки для адаптации рекламных видеороликов к конкретным регионам или даже отдельным клиентам. Дипфейки также позволяют представлять новый контент для политической и социальной мобилизации. Активистские дипфейки могут оказать мощное воздействие, привлечь больше внимания к определенным политическим вопросам и группам.

Дипфейки уже некоторое время служат и развлекательным целям, таким как пародии, мемы или спецэффекты в фильмах и сериалах. Индустрия развлечений осознала потенциал этой технологии и активно ее использует, так что целое поколение вырастает на этом неоднозначном жанре медиа. Для подростков и молодых людей дипфейки не обязательно являются чем-то вроде пугающего призрака. Первый контакт с дипфейками часто происходит в социальных сетях. Молодые люди, с одной стороны, очень активны на цифровых платформах, а с другой стороны, создание и распространение дипфейков не является чем-то сложным.

В киноиндустрии крупные студии уже несколько лет используют дипфейк-технологии, например, для «омоложения» стареющих актеров.

Но подобные события вызывают не только энтузиазм. «Гильдия сценаристов Америки» (WGA) и «Гильдия киноактеров – Американская федерация артистов телевидения и радио» (SAG-AFTRA) отреагировали на растущее использование генеративного ИИ и дипфейков, призвав к забастовке, которая продлилась несколько месяцев. Ключевой задачей забастовки была защита рабочих мест в творческих отраслях от угрозы того, что ИИ и дипфейк-технологии могут постепенно заменить людей [20].

Все бóльшую значимость в связи с растущим использованием основанного на ИИ контента приобретает феномен «виртуальных инфлюенсеров». Они впервые появились в 2016 году и быстро были идентифицированы как «инфлюенсеры» благодаря их активному участию в маркетинговых кампаниях [14]. Технологические достижения сделали возможным создание виртуальных инфлюенсеров с реалистичными характеристиками и поведением.

Виртуальные инфлюенсеры представляют себя так же, как и люди, с помощью фотографий, видео и историй из своей «жизни». Они могут принимать самые разные формы: от мультипликационных фигур до гиперреалистичных персонажей, которые практически неотличимы от реальных людей. Их цель – оказать влияние на пользователей социальных сетей, делясь контентом и взаимодействуя со своей аудиторией. Дискуссии вокруг них являются противоречивыми, особенно в отношении границ между человеческим и нечеловеческим, а также изображения нереалистичных идеалов красоты. Их подлинность и прозрачность также являются предметом споров, поскольку не все аккаунты виртуальных инфлюенсеров указывают на то, что они являются цифровыми персонажами.

Виртуальные инфлюенсеры также могут играть определенную роль в политическом дискурсе. Они могут распространять политические сообщения среди целевых групп, мало интересующихся политическими вопросами, тем самым способствуя их участию в политической жизни. Возможность целенаправленного распространения дезинформации или экстремистской пропаганды, в свою очередь, может поставить под угрозу целостность политических процессов. Кроме того, виртуальные инфлюенсеры могут мобилизоваться для участия в выборах и поддерживать политические кампании «инновационными» способами, целенаправленно продвигая политические партии и кандидатов.

Регулирование создания и распространения дипфейков

Обсуждение регулирования дипфейков, как правило, сосредоточено на предотвращении или, по крайней мере, сдерживании их негативных последствий при сохранении потенциала. Отправной точкой для дискуссий по вопросам регулирования является осознание того, что, например, опе-

раторы социальных сетей не предпринимают достаточных мер по борьбе с дипфейками. Поэтому необходимы обязательные правовые требования, обеспечивающие высокий уровень защиты пользователей [4].

Поскольку дипфейки часто публикуются и распространяются в социальных медиа, регулирование различных платформ считается ключевым элементом в борьбе с дипфейками. А поскольку для создания дипфейков регулярно используются видео, фотографии и аудиозаписи людей, важную роль играют также законы о защите данных и об авторском праве [5].

Некоторые страны принимают специальные законы о дипфейках. Например, в Китае такой закон вступил в силу в январе 2023 года [18]. В действующем российском законодательстве пока не содержится специальных норм, регулирующих правовой статус дипфейков, хотя в парламенте уже обсуждалось несколько законопроектов, направленных на регулирование в этой области [3].

Из-за множества возможных сценариев злоупотреблений, связанных с дипфейками, и одновременного быстрого развития технологий ИИ активно разрабатываются контрмеры, которые могут помочь обнаруживать дипфейки. В связи с актуальностью этой темы в настоящее время разработчики и операторы платформ ведут активные исследования.

Использование для обнаружения дипфейков ИИ является многообещающим подходом к решению этой проблемы. Однако, поскольку дипфейки существуют только с 2017 года, как их создание, так и обнаружение являются очень молодой областью исследований. Это означает, что с каждым новым методом обнаружения дипфейков разрабатываются новые, более сложные методы создания дипфейков, которые затрудняют их обнаружение. Подобно обнаружению вирусов и вредоносных программ антивирусным программным обеспечением, методы обнаружения дипфейков должны постоянно совершенствоваться, чтобы идти в ногу со временем [22].

Надежность методов обнаружения дипфейков на основе ИИ сильно зависит от качества и разнообразия обучающих данных. Часто имеющиеся данные недостаточно полны для надежной идентификации всех форм дипфейков, что может привести к ошибкам в их обнаружении. Таким образом, обнаружение дипфейков с помощью ИИ является важным, но не единственным компонентом в борьбе с дезинформацией. В последние годы на рынок все чаще выходят стартапы, предлагающие услуги обнаружения дипфейков. Одним из возможных вариантов в будущем может стать автоматизированный просмотр контента социальных сетей для различения реального и поддельного контента.

В дополнение к подходу «обнаружения дипфейков» существует возможность «проверки происхождения», которая подразумевает защищен-

ность медиаданных с помощью цифровых подписей, доказывающих их происхождение [7]. Например, нажав на изображение, можно увидеть, что оно было снято в определенный день определенной камерой. Однако проблема в том, что эти подписи технически очень легко удалить и они подтверждают только происхождение, а не подлинность.

Политическое медиаобразование является важным элементом грамотного обращения с дипфейками. Общество должно создать защитную структуру для обеспечения безопасности граждан. Социальные сети, смартфоны и ИИ массово используются «новыми медиа», и критически важные вопросы о медиа-опосредованном знании о мире в целом и политическом мире в частности должны быть переосмыслены [21].

Формирование политического мнения и политические действия происходят в динамично меняющихся социальных и медиатехнологических условиях. В то время как в мире аналоговых медиа несколько вещателей общались с массами получателей в одном направлении, а журналистские редакционные группы доставляли одну и ту же программу всем получателям в определенное время и в определенных форматах, сегодня платформы предлагают алгоритмически скомпилированный и индивидуализированный контент и одновременно производят персонализированные потоки данных.

Технология сама по себе не определяет социальные изменения, но ее социальное воздействие выходит далеко за рамки использования только в качестве инструмента. Медиаобразованию необходима позиция, которая предполагает взаимодействие между индивидуальным, социальным и медиатехнологическим мирами. Необходимо фундаментальное понимание того, как дипфейки меняют коммуникативные рамки и политические события. Поэтому формирование политического мнения, участие и расширение прав и возможностей должны поддерживаться конвергенцией политического образования и критического медиаобразования.

Помимо ориентации на компетентность, следует особо подчеркнуть образовательную перспективу: при понимании цифровой грамотности как цели обучения и концепции образования как взаимного проникновения себя и мира критическое мышление и развитие личности выходят на первый план. Пользователи должны понимать, что образование в цифровом мире не исчерпывается приобретением новых знаний и новых количественно измеримых компетенций; важны рефлексивные, качественные изменения в мировоззрении и самовосприятии.

В свою очередь, средства массовой информации сегодня должны не только предоставлять информацию, но и разоблачать фейковые новости, дипфейки и дезинформацию. Это делает очевидной огромную проблему, стоящую перед журналистами. Дилемма между необходимостью быстро-

го освещения событий, проверкой информации и поддержанием качества журналистики является особенно сложной. Проверка фактов и контекстуализация информации сейчас являются важнейшими задачами в профессиональной журналистике [1, 11].

Средства массовой информации должны завоевывать авторитет посредством прозрачности и открытости. На практике это означает, что они должны рассказывать о своих методах работы. Кроме того, быстрое распространение дезинформации требует адаптации традиционных журналистских форматов для эффективного и точного информирования аудитории. В сложном информационном ландшафте просветительская работа важнее, чем когда-либо, поскольку она помогает поддерживать доверие общественности к СМИ.

Заключение

Медийные, общественные и политические дебаты в основном вращаются вокруг потенциального злоупотребления дипфейками в политических целях: злоумышленники могут использовать их для фальсификации результатов выборов, нападений на политических оппонентов или попыток иностранного вмешательства во внутривнутриполитические события. Дипфейки подрывают доверие к средствам массовой информации и политическому процессу.

Однако в то же время эта технология предлагает и множество возможностей. С социально-политической точки зрения дипфейки позволяют использовать новые формы историко-политического образования и политической мобилизации. Все больше людей создают дипфейки в развлекательных целях и делятся ими в социальных сетях, которые часто становятся первой точкой контакта с этой технологией, особенно для молодежи. Киноиндустрия и рекламная индустрия, а также другие отрасли экспериментируют с коммерческим использованием дипфейков. Сюда также входят виртуальные инфлюенсеры – виртуальные «личности», созданные с помощью дипфейков, которые продвигают определенные бренды и продукты в социальных сетях. Однако такое коммерческое, развлекательное, сатирическое и активистское использование дипфейков также таит в себе множество потенциальных рисков, связанных с нарушением прав личности, потерей работы или обманом.

Чтобы минимизировать отрицательные последствия дипфейков и в то же время обеспечить реализацию их «положительного» потенциала, необходимо взаимодействие различных заинтересованных сторон и инструментов регулирования. Операторы социальных сетей обязаны принимать жесткие меры против дезинформирующих дипфейков. Аналогичным образом разработчики и поставщики должны взять на себя от-

ветственность и препятствовать злонамеренному использованию соответствующих технологий. Однако, поскольку мер саморегулирования недостаточно, необходимо одновременное развитие правового регулирования дипфейков и эффективное правоприменение. Помимо улучшенного регулирования и саморегулирования, необходимы технические решения, позволяющие обнаруживать дипфейки, маркировать подлинные медиафайлы или делать данные непригодными для использования генераторами дипфейков.

В связи с распространением дипфейков журналисты сталкиваются с новыми вызовами и задачами, которые включают в себя более тщательную проверку фактов и контекстуализацию информации, чтобы средства массовой информации продолжали выполнять свою центральную роль в качестве источника достоверной информации. Каждый отдельный пользователь социальных сетей также может внести свой вклад в минимизацию рисков, связанных с дипфейками, приобретя базовые знания об этой технологии и ее влиянии, а также критически осмыслив ее применение.

Литература

1. *Аббуд А.М.В.* Инструменты AI-журналистики в противодействии фейковой реальности // Знак: проблемное поле медиаобразования. – 2024. – № 4. – С. 31–38. – DOI: 10.47475/2070-0695-2024-54-4-31-38.
2. *Аль-Сахель С., Смирнов А.И.* Технологии искусственного интеллекта в палестинно-израильском конфликте // Международная жизнь. – 2023. – URL: <https://interaffairs.ru/news/show/43910> (дата обращения: 20.10.2025).
3. *Бодров Н.Ф., Лебедева А.К.* Понятие дипфейка в российском праве, классификация дипфейков и вопросы их правового регулирования // Юридические исследования. – 2023. – № 11. – С. 26–41. – DOI: 10.25136/2409-7136.2023.11.69014.
4. *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.
5. *Добробаба М.Б.* Дипфейки как угроза правам человека // Lex Russica. – 2022. – Т. 75, № 11. – С. 112–119. – DOI: 10.17803/1729-5920.2022.192.11.112-119.
6. *Евсеев С.М.* Этапы развития технологий искусственного интеллекта и уточнение терминологии // Инновации. – 2021. – № 4. – С. 39–48. – DOI: 10.26310/2071-3010.2021.270.4.005.
7. *Лемайкина С.В.* Методы обнаружения фальсификации информационного контента // Философия права. – 2023. – № 3. – С. 119–124.
8. *Макаров Е.Б.* Глубинная подделка: кто получает дивиденды лжеца // Военное обозрение. – 2021. – 12 мая. – URL: <https://topwar.ru/182814-glubinnaja-poddelka-kto-poluchaet-dividendy-lzheca.html> (дата обращения: 20.10.2025).

9. *Невоструев К.Н.* Обзор литературы по методам машинного обучения (Machine Learning) // Компьютерные инструменты в образовании. – 2014. – № 4. – С. 19–26.
10. *Созыкин А.В.* Обзор методов обучения глубоких нейронных сетей // Вестник Южно-Уральского государственного университета. Серия: Вычислительная математика и информатика. – 2017. – Т. 6, № 3. – С. 28–59. – DOI: 10.14529/cmse170303.
11. *Соколов А.В.* Типология инструментов фактчекинга: на пути к устойчивой информационной среде // Вопросы медиабизнеса. – 2024. – Т. 3, № 3. – С. 68–76. – DOI: 10.24412/3034-1930-2024-0230.
12. *Chesney B., Citron D.* Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security // California Law Review. – 2019. – Vol. 107, iss. 6. – P. 1753–1820. – DOI: 10.2139/ssrn.3213954.
13. *Chowdhury R.* AI-fuelled Election Campaigns are Here – Where are the Rules? // Nature. – 2024. – Vol. 628, iss. 8007. – P. 237. – DOI: 10.1038/d41586-024-00995-9.
14. *Conti M., Gathani J., Tricomi P.P.* Virtual Influencers in Online Social Media // IEEE Communications Magazine. – 2022. – Vol. 60, iss. 8. – P. 86–91. – DOI: 10.1109/MCOM.001.2100786.
15. Dalí Lives (via Artificial Intelligence) / Salvador Dalí Museum. – URL: <https://thedali.org/exhibit/dali-lives/> (accessed: 20.10.2025).
16. *Gosse C., Burkell J.* Politics and Porn: How News Media Characterizes Problems Presented by Deepfakes // Critical Studies in Media Communication. – 2020. – Vol. 37, iss. 5. – P. 497–511. – DOI: 10.1080/15295036.2020.1832697.
17. *Harris D.* Deepfakes: False Pornography Is Here and the Law Cannot Protect You // Duke Law & Technology Review. – 2019. – Vol. 17, iss. 1. – P. 99–128.
18. *Interesse G.* China to Regulate Deep Synthesis (Deepfake) Technology Starting 2023 // China Briefing. – 2022. – URL: <https://www.china-briefing.com/news/china-to-regulate-deep-synthesis-deep-fake-technology-starting-january-2023/> (accessed: 20.10.2025).
19. *Pavelec M.* Deepfakes and Democracy (Theory): How Synthetic Audio-Visual Media for Disinformation and Hate Speech Threaten Core Democratic Functions // Digital Society. – 2022. – Vol. 1. – Art. 19. – P. 1–37. – DOI: 10.1007/s44206-022-00010-6.
20. *Pfefferkorn J.* Computer-written Scripts and Deepfake Actors: What’s at the Heart of the Hollywood Strikes against Generative AI // The Conversation. – 2023. – URL: <https://theconversation.com/computer-written-scripts-and-deepfake-actors-whats-at-the-heart-of-the-hollywood-strikes-against-generative-ai-210191> (accessed: 20.10.2025).
21. The challenges of media and information literacy in the artificial intelligence ecology: deepfakes and misinformation / A. Sanchez-Acedo, A. Carbonell-Alcocer, M. Gertrudix, J.-L. Rubio-Tamayo // Communication & Society. – 2024. – Vol. 37, iss. 4. – P. 223–239. – DOI: 10.15581/003.37.4.223-239.

22. A systematic literature review on the effectiveness of deepfake detection techniques / L. Stroebel, M. Llewellyn, T. Hartley, T.S. Ip, M. Ahmed // *Journal of Cyber Security Technology*. – 2023. – Vol. 7, iss. 2. – P. 83–113. – DOI: 10.1080/23742917.2023.2192888.

23. *White E.* Positive Implications of Deepfake Technology in the Arts and Culture // *Arts Management & Technology Laboratory*. – 2021. – URL: <https://amt-lab.org/blog/2021/8/positive-implications-of-deepfake-technology-in-the-arts-and-culture> (accessed: 20.10.2025).

24. *Whiteside P.* David Beckham «Speaks Nine Languages» in Malaria Campaign's New Video // *Sky News*. – 2019. – URL: <https://news.sky.com/story/david-beckham-speaks-nine-languages-in-malaria-campaigns-new-video-11688600> (accessed: 20.10.2025).

References

1. Abbood A.M.W. Instrumenty AI-zhurnalistsiki v protivodeistvii feikovoi real'nosti [AI Journalism Tools in Counteracting Fake Realities]. *Znak: problemnoe pole mediaobrazovaniya = Sign: Problematic Field in Mediaeducation*, 2024, no. 4, pp. 31–38. DOI: 10.47475/2070-0695-2024-54-4-31-38.

2. Al-Sakhel S., Smirnov A.I. Tekhnologii iskusstvennogo intellekta v palestino-izrail'skom konflikte [Artificial Intelligence Technologies in the Israeli-Palestinian Conflict]. *Mezhdunarodnaya zhizn' = The International Affairs*, 2023. (In Russian). Available at: <https://interaffairs.ru/news/show/43910> (accessed 20.10.2025).

3. Bodrov N.F., Lebedeva A.K. Ponyatie dipfeika v rossiiskom prave, klassifikatsiya dipfeikov i voprosy ikh pravovogo regulirovaniya [The Concept of Deepfake in Russian Law, Classification of Deepfake and Issues of Their Legal Regulation]. *Yuridicheskie issledovaniya = Legal Research*, 2023, no. 11, pp. 26–41. DOI: 10.25136/2409-7136.2023.11.69014.

4. Vinogradov V.A., Kuznetsova D.V. Zarubezhnyi opyt pravovogo regulirovaniya tekhnologii «dipfeik» [Foreign Experience in Legal Regulating Deepfake Technology]. *Pravo. Zhurnal Vyshej shkoly ekonomiki = Law Journal of the Higher School of Economics*, 2024, no. 2, pp. 215–240. DOI: 10.17323/2072-8166.2024.2.215.240.

5. Dobrobaba M.B. Dipfeiki kak ugroza pravam cheloveka [Deepfakes as a Threat to Human Rights]. *Lex Russica*, 2022, vol. 75, no. 11, pp. 112–119. DOI: 10.17803/1729-5920.2022.192.11.112-119. (In Russian).

6. Evseenko S.M. Etapy razvitiya tekhnologii iskusstvennogo intellekta i utocnenie terminologii [Stages of Artificial Intelligence Technologies Development and Clarification of Terminology]. *Innovatsii = Innovations*, 2021, no. 4, pp. 39–48. DOI: 10.26310/2071-3010.2021.270.4.005.

7. Lemaikina S.V. Metody obnaruzheniya fal'sifikatsii informatsionnogo kontenta [Methods for Detecting Falsification of Information Content]. *Filosofiya prava = Philosophy of Law*, 2023, no. 3, pp. 119–124. (In Russian).

8. Makarov E.B. Glubinnaya poddelka: kto poluchaet dividendy lzhetca [Deep Forgery: Who Gets the Liar's Dividends]. *Voennoe obozrenie* [Military Review], 2021, May 12. Available at: <https://topwar.ru/182814-glubinnaja-poddelka-kto-poluchaet-dividendy-lzheca.html> (accessed 20.10.2025).
9. Nevostruev K.N. Obzor literatury po metodam mashinnogo obucheniya (Machine Learning) [Review of the literature on the methods of machine learning]. *Komp'yuternye instrumenty v obrazovanii = Computer Tools in Education*, 2014, no. 4, pp. 19–26.
10. Sozykin A.V. Obzor metodov obucheniya glubokikh neironnykh setei [An Overview of Methods for Deep Learning in Neural Networks]. *Vestnik Yuzhno-Ural'skogo gosudarstvennogo universiteta. Seriya: Vychislitel'naya matematika i informatika = Bulletin of the South Ural State University. Series: Computational Mathematics and Software Engineering*, 2017, vol. 6, no. 3, pp. 28–59. DOI: 10.14529/cmse170303.
11. Sokolov A.V. Tipologiya instrumentov faktchekinga: na puti k ustoichivoi informatsionnoi srede [Typology of Fact-Checking Tools: Towards a Sustainable Information Environment]. *Voprosy mediabiznesa = Issues of Media Business*, 2024, vol. 3, no. 3, pp. 68–76. DOI: 10.24412/3034-1930-2024-0230.
12. Chesney B., Citron D. Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security. *California Law Review*, 2019, vol. 107, iss. 6, pp. 1753–1820. DOI: 10.2139/ssrn.3213954.
13. Chowdhury R. AI-fuelled Election Campaigns are Here – Where are the Rules? *Nature*, 2024, vol. 628, iss. 8007, p. 237. DOI: 10.1038/d41586-024-00995-9.
14. Conti M., Gathani J., Tricomi P.P. Virtual Influencers in Online Social Media. *IEEE Communications Magazine*, 2022, vol. 60, iss. 8, pp. 86–91. DOI: 10.1109/MCOM.001.2100786.
15. Salvador Dalí Museum. *Dalí Lives (via Artificial Intelligence)*. Available at: <https://thedali.org/exhibit/dali-lives/> (accessed 20.10.2025).
16. Gosse C., Burkell J. Politics and Porn: How News Media Characterizes Problems Presented by Deepfakes. *Critical Studies in Media Communication*, 2020, vol. 37, iss. 5, pp. 497–511. DOI: 10.1080/15295036.2020.1832697.
17. Harris D. Deepfakes: False Pornography Is Here and the Law Cannot Protect You. *Duke Law & Technology Review*, 2019, vol. 17, iss. 1, pp. 99–128.
18. Interesse G. China to Regulate Deep Synthesis (Deepfake) Technology Starting 2023. *China Briefing*, 2022. Available at: <https://www.china-briefing.com/news/china-to-regulate-deep-synthesis-deep-fake-technology-starting-january-2023/> (accessed 20.10.2025).
19. Pawelec M. Deepfakes and Democracy (Theory): How Synthetic Audio-Visual Media for Disinformation and Hate Speech Threaten Core Democratic Functions. *Digital Society*, 2022, vol. 1, art. 19, pp. 1–37. DOI: 10.1007/s44206-022-00010-6.
20. Pfefferkorn J. Computer-written Scripts and Deepfake Actors: What's at the Heart of the Hollywood Strikes Against Generative AI. *The Conversation*, 2023. Available at: <https://theconversation.com/computer-written-scripts-and-deepfake-actors-whats-at-the-heart-of-the-hollywood-strikes-against-generative-ai-210191> (accessed 20.10.2025).

21. Sanchez-Acedo A., Carbonell-Alcocer A., Gertrudix M., Rubio-Tamayo J.L. The challenges of media and information literacy in the artificial intelligence ecology: deep-fakes and misinformation. *Communication & Society*, 2024, vol. 37, iss. 4, pp. 223–239. DOI: 10.15581/003.37.4.223-239.

22. Stroebel L., Llewellyn M., Hartley T., Ip T.S., Ahmed M. A systematic literature review on the effectiveness of deepfake detection techniques. *Journal of Cyber Security Technology*, 2023, vol. 7, iss. 2, pp. 83–113. DOI: 10.1080/23742917.2023.2192888.

23. White E. Positive Implications of Deepfake Technology in the Arts and Culture. *Arts Management & Technology Laboratory*, 2021. Available at: <https://amt-lab.org/blog/2021/8/positive-implications-of-deepfake-technology-in-the-arts-and-culture> (accessed 20.10.2025).

24. Whiteside P. David Beckham «Speaks Nine Languages» in Malaria Campaign's New Video. *Sky News*, 2019. Available at: <https://news.sky.com/story/david-beckham-speaks-nine-languages-in-malaria-campaigns-new-video-11688600> (accessed 20.10.2025).

Статья поступила в редакцию 21.10.2025.

Статья прошла рецензирование 17.12.2025.

The article was received on 21.10.2025.

The article was reviewed on 17.12.2025.