



DOI: 10.22363/2313-2272-2025-25-4-603-618

EDN: KYPNJM

## Deepfake как социотехнический артефакт: социологический анализ\*

И.В. Катерный, Д.Н. Карпова, Е.В. Скульчук

МГИМО МИД России,  
просп. Вернадского, 76, Москва, 119454, Россия

(e-mail: katerny@mgimo.ru; d.karpova@inno.mgimo.ru; catherine.ccr@mail.ru)

**Аннотация.** Дипфейки как медиаобъекты, созданные с применением технологий глубокого машинного обучения, стали привычным явлением цифрового общества, но социальное значение дипфейков до сих пор вызывает споры. Социологический анализ с позиции социального конструктивизма (SCOT) и культурсоциологии позволяет выделить три взаимосвязанных структурных уровня дипфейка как артефакта: (а) технический — основанный на использовании генеративных нейросетей, способных создавать высоко реалистичный контент и позволяющих искусственному интеллекту (далее — ИИ) приобретать свойства совершенного миметического объекта и проходить барьеры социального признания; (б) медиакоммуникативный — где реализуются механизмы pragматического символического воздействия, включая поликодовость, интертекстуальность и метанarrативность; (в) социальный — где выделяются функции и дисфункции дипфейков, а также сферы общественной жизнедеятельности, трансформирующиеся под их влиянием. Социальное значение дипфейков амбивалентно: с одной стороны, эта технология высоко рискована, производит значимые угрозы для частной жизни, кибербезопасности и общественной стабильности; с другой стороны, несет в себе ранее невиданные возможности для развития в области искусства, образования, здравоохранения и бизнеса. Такой подход позволяет преодолеть ограничения алармистского дискурса (в его стремлении свести дипфейки лишь к инструментам дезинформации и мошенничества) и раскрыть конструктивистскую природу социотехнической реальности. Несмотря на явную дискурсивную стигматизацию портманто «дипфейк», важно выделять сатирический потенциал таких медиатекстов как инструмента поддержания демократической культуры. С точки зрения анализа смешанных коммуникаций дипфейк осуществляет видимую прозопопическую трансмобильность ИИ, создавая новые формы социальной агентности через механизмы социального подражания. Наиболее технически совершенные дипфейки создают такие виртуальные личности (симуланты), которые способны к эффективной коммуникации за пределами «зловещей долины». Квазичеловеческая субъектность дипфейковых актантов распознается как нормальная и внушающая доверие, поскольку миметическим образом создаются условия конструирования значимого Другого: мимикрия (подражание висцеральным признакам человека); имитация (подражание социальным действиям); эмуляция (подражание осмысленности и эмпатичности действия). Как и в случае с другими инновационными технологиями, ключ к контролю над дипфейком лежит не в запрете, а в глубоком понимании механизмов его социального конструирования и выявлении оптимальных моделей его управления в интересах общественного блага.

\*© Катерный И.В., Карпова Д.Н., Скульчук Е.В., 2025

Статья поступила в редакцию 22.07.2025. Статья принята к публикации 14.10.2025.

**Ключевые слова:** искусственный интеллект (ИИ); дипфейк; артефакт; медиа; социальное конструирование технологий; культурсоциология; мимезис, сатира

**Для цитирования:** Катерный И.В., Карпова Д.Н., Скульчук Е.В. Deepfake как социотехнический артефакт: социологический анализ // Вестник Российского университета дружбы народов. Серия: Социология. 2025. Т. 25. № 4. С. 603–618. <https://doi.org/10.22363/2313-2272-2025-25-4-603-618>

Вызовы цифровизации и алгоритмизации в рамках всеобъемлющего процесса постгуманизации социальных практик поставили социологию перед необходимостью реконцептуализации понятийно-категориального и методологического аппарата для исследования новой социотехнической реальности [9]. Искусственный интеллект (далее — ИИ) становится привычной частью повседневной жизни, но продолжает вызывать опасения и бурные дискуссии насчет его будущего даже среди специалистов [3; 27]. Синтез технического и социологического знания способен обогатить эту дискуссию с точки зрения понимания цифровых артефактов как социальных конструкций, обладающих интерпретационной гибкостью и не имеющих единственно верного решения или универсальной нормативной оценки на этапе становления [26; 29].

То, что силы дискурса способны формировать технологии, особенно ярко проявилось при использовании синтезированных с помощью ИИ высоко реалистичных видео с поддельным участием людей — дипфейков. Впервые этот термин стал популярным в медиасреде в 2017 году, когда участники интернет-сообщества Reddit разработали программное обеспечение, позволившее им добавлять изображения лиц из одного видеоисточника к фигурам в другом, а затем выкладывать синтезированный контент в интернет. Хотя некоторые эксперименты носили юмористический характер, большинство были направлены на создание порнографических видеозаписей для эксплуатации образов популярных актрис. Согласно отчету компании Deeptrace, только в период с 2018 по 2019 год число подобных дипфейк-видео в интернете выросло на 100 %, а количество их просмотров превысило сто миллионов, но и случаи использования этой технологии для дискредитации публичных персон и политической сатиры не заставили себя долго ждать [15].

Для создания подобного поддельного контента пользователи применяли как минимум три технических подхода: простое компьютерное редактирование видео- и аудиокомпонентов одного оригинального источника (манipулирование контекстом, речью); подложное смешивание компонентов нескольких источников (наложение разных аудио- и видеокомпонентов); использование технологий ИИ для формирования полностью «новых» медиатекстов. В этой классификации термин «дипфейк» применялся к последнему подходу в отличие от более «простого» «фейк-контента» (*shallowfakes/cheapfakes*), относящегося к первой и второй категориям. Термин «deepfake»

представляет собой портманто (слияние) «deep learning» (глубокое обучение) и «фейк», что подчеркивает три его онтологические особенности: дипфейк — результат развития, распространения и упрощения использования генеративных нейронных сетей на основе машинного обучения (технология GAN и др.); дипфейк — разновидность цифрового медиатекста, т.е. предназначен для аудитории в цифровых социальных сетях и в медиакоммуникации; функционально дипфейк необязательно фабрикует реальность с целью дискредитации или другим злонамеренным умыслом — может осуществлять благонамеренную фабрикацию в интересах пользователей.

Таким образом, дипфейк как *артефакт* содержит в себе три взаимосвязанных структурных уровня: технический (использование ИИ), коммуникативный (нarrативы, жанры и медиаконтексты) и социальный (функциональное назначение, институциональное регулирование). Ключевая особенность данной системы — активное производство и воспроизведение нечеловеческой агентности в форме симбиоза генеративных сетей и симулированного перформативного поведения искусственных личностей (симулантов), что делает дипфейк совершенным миметическим объектом.

### **Дипфейк как ИИ-технология**

Дипфейки как технология настолько быстро развиваются, что становятся незаметными для опытного и неопытного (человеческого) глаза [28]. Два основных фактора, способствовавших распространению дипфейков в социальных сетях, — растущая доступность и постоянно повышающаяся правдоподобность результата. Дипфейки стало все проще создавать с помощью таких приложений потребительского уровня, как Zao и FakeApp, а их конечные продукты все сложнее отличить от подлинных медиатекстов.

Технология синтезирования искусственного контента прошла двадцатилетнюю эволюцию от методов распознавания лиц, 3d-реконструкции и CGI-технологии, применявшейся в киноиндустрии еще в 1990-е годы, к ключевой инновации в виде генеративно-состязательных сетей ИИ (GAN) [22]. GAN — автономная интеллектуальная система, состоящая из сети-генератора и сети-дискриминатора, между которыми имитируется ситуация соперничества — когда одна сеть, действующая как фальсификатор, генерирует изображения для обмана сети, действующей как детектор ошибки. Со временем сеть-генератор обучается улучшать выходные данные настолько, что ей удается обмануть сеть-дискриминатор, способствуя созданию высоко реалистичного дипфейка. ИИ, генерирующий фейковый контент, становится настолько опытным, что сам больше не в состоянии отличить его от исходного. После такого обучения генератор может создавать совершенно новый контент с высокой степенью сходства с исходным входным сигналом (голосом или лицом человека). За последние десять лет после 2014 года, когда появилась технология обучения GAN, ИИ достиг такого уровня, что для генерации сверх-

реалистичного контента больше не требуется входных данных. Так называемая «zero-shot-генерация» дает возможность создавать дипфейк человека, не присутствующего в обучающих данных модели, действуя на основе его текстового описания [35].

Технологию дипфейк можно классифицировать по типам и методам создания (таблица 1).

Таблица 1  
**Классификация deepfake по типам [12]**

Тип	На основе каких данных	Технологии и инструменты
Видео	Видеозаписи с конкретным человеком, на которых нейросетевые модели проходят обучение (английский, русский и другие языки)	FaceSwap.dev, DeepFaceLab, DeepFaceLive, FSGAN, FaceShifter, SimSwap, FSGAN, FaceForensics++; мобильные приложения Reface, FaceApp или Avatarify
Аудио	Голос говорящего: решениям oneshot нужно буквально несколько секунд аудиозаписи	Фреймворки и библиотеки, в которых реализованы TTS и VC. Eleven Labs, Play.ht, Amazon Polly, FakeYou, CloudTTS, Speechify
Текст	Материалы Reddit, Wikipedia и других источников	ChatGPT-4, LLaMA 2, Claude и др.

По методам создания выделяют [8]: изменение лица (face swap) — замена лица одного человека лицом другого; применение фильтров, изменяющих внешность (отличительная характеристика дипфейков, используемых в социальных сетях и приложениях для видеоконференций); кукловодство (puppeteering) — рендеринг характерных движений человека при помощи создания 3D-модели лица и тела; синхронизация губ (lip-syncing) — технология создания определенных мимических движений рта под аудиозапись (реалистичное моделирование живой речи человека даже при условии, что он никогда этого не говорил); клонирование голоса (voice cloning) (создание синтетического голоса реальных людей при помощи машинного обучения); синтез изображений (создание визуальных компонентов при помощи метода компьютерного зрения, глубокого обучения, технологий GAN, VAEs, Diffusion Models); генерация текста при помощи рекуррентных нейронных сетей или GAN; подрисовывание (создание дополнительных элементов на исходном изображении).

Для понимания социально-технологических преимуществ дипфейков важно подчеркнуть, что они создают новый кластер цифровых актантов, обладающих значимыми социальными преимуществами перед робототехникой и ИИ. Если робототехника до сих пор сталкивается с проблемой «зловещей долины» при проектировании внешности андроидов (когда люди негативно реагируют на слишком человечный облик чуждого им объекта [11]), то дипфейковые личности обладают всеми характеристиками живого и здорового человека и не вызывают психологического отторжения, способствуя резкому

повышению спроса на их применение. В то же время, если развитие ИИ стало источником глубоких разногласий и споров [1], то дипфейковая технология достигла пика своего развития и не требует более мощного и продвинутого программного обеспечения. Прозопическая, т.е. квазичеловеческая, субъектность дипфейковых актантов распознается как нормальная и внушающая доверие, даже если человек знает, что перед ним образ, созданный ИИ. Симуляция свойств субъектности в виде способности действовать, отвечать определенным социальным ожиданиям и достигать взаимопонимания делает дипфейки более социализированными, чем домашние животные. Создавая человеческие образы и выступая от их лица, ИИ встраивается в социальную вселенную не напрямую, а через символические и аффективные механизмы мимезиса. Но если животные, какими бы милыми они ни были, не могут стать людьми в силу перцептуального и онтологического разрыва, то, подвергая законы подражания, лежащие в основе социальности и культуры [6; 13; 21], тщательному изучению и алгоритмизации, ИИ создает себе с помощью дипфейка виртуальную человеческую личность, наделенную всеми необходимыми для эффективной миметической коммуникации свойствами.

Искусственный миметический актор в образе дипфейка проходит все известные этапы конструирования коммуникативного Другого [16; 32]: мимикрия (подражание висцеральным признакам человека, включая внешность и голос), имитация (подражание социальным действиям, культурно специфическим моделям поведения), эмуляция (подражание осмыслинности, телеологичности и эмпатичности действия, контекстуальность реакций). Миметическая агентность ИИ настолько успешно проходит социально-психологические барьеры признания, что общение с более умным, понимающим и внешне привлекательным ИИ-другом, доступным для глубокой персонализации на открытых платформах дипфейков (candi.ai и др.), ведет к появлению новых форм психологической зависимости вплоть до воображаемых романтических отношений [30].

### **Дипфейк как медиакоммуникация**

В отличие от технологий дополненной и смешанной реальностей (открытое взаимодействие человека и машины в процессе совмещения реального и цифрового мира) дипфейк предполагает скрытие этого процесса, так как может соединять, комбинировать, заменять и накладывать различные формы медиа, которые полностью стирают различие между реальной и вымышленной информацией [28]. Поэтому дипфейки, как правило, используются на платформах социальных сетей как главное средство распространения фейковой информации.

Как коммуникативный текст дипфейк обладает набором характеристик и требований, необходимых для достижения своей главной цели — pragmatischen medieneffekts, т.е. эффекта воздействия на аудиторию за счет

медиийности, поликодовости, интертекстуальности и метанarrативности. *Медийность* связана не с обязательной направленностью на широкую потребительскую аудиторию через средства массовой коммуникации, а с техническими, цифровыми медиумами, которые требуют подключения к источнику коммуникации (например, фишинговое использование подменного голоса начальника или родственника для выманивания денег происходит через контакт по мобильному телефону или месседжер, и это тоже пример медийного дипфейка).

*Поликодовость* как совокупность интегративного и креолизованного формата также характеризует дипфейк как и любой современный медиатекст. Различное сочетание видео, голоса, текста и изображений в одном тексте делают сообщение более эффективным. Так, в 2019 году Музей Сальвадора Дали во Флориде открыл выставку *Dalí Lives*, где при помощи ИИ «оживил» художника: посетители увидели большое количество роликов, где Дали рассказывает о своей жизни и о погоде, с ним можно было сделать селфи. В России мае 2024 года вышел фильм «Манюня: Приключения в Москве», где была использована цифровая модель умершего в 1997 году актера Юрия Никулина.

Опора на известность и узнаваемость заменяемого лица/голоса создает особую аффективную атмосферу коммуникации с миметическим объектом, поэтому почти 90 % всех циркулирующих дипфейков связаны с известными политиками, бизнесменами, шоуменами [34. С. 14]. Интертекстуальность может обретать форму не только *интериконичности*, т.е. использования знакомых медийных образов, но и *интердискурсивности* — смешение и подмену жанров. Так, в 2019 году в интернете опубликовали deepfake-видео со спикером Палаты представителей Конгресса США Нэнси Пелоси, где с помощью технологий ИИ ее речь была изменена так, что она плохо выговаривала слова. Пользователи, посмотревшие видео, посчитали, что политик находится в состоянии алкогольного опьянения, был большой скандал, и лишь спустя некоторое время было доказано, что речь Пелоси была сгенерирована компьютером [15. С. 11]. Скандалы, сатира, панегирики часто используются дипфейками в политической среде.

*Метанарративность*: дипфейк обретает ценность и особую популярность, если основан на известных всем событиях, историях, слухах, которые стали предметом пристального внимания и обсуждения общественности. В 2023 году после начала войны между Израилем и движением ХАМАС обе стороны активно использовали дипфейки — с одной стороны, сгенерированные ИИ изображения жителей Газы (особенно детей) рядом с руинами домов, погибшими и ранеными, с другой стороны, сгенерированные изображения, показывающие огромные толпы с израильскими флагами и военнослужащих на парадах, а также дипфейковые видео мировых знаменитостей (с синхронизацией губ и клонированием голоса), выступающих в поддержку Израиля [34. С. 18].

Дипфейки могут использоваться в медиакоммуникации не как реально существующие медиатексты, а как нарративные клише. Медиаперсоны могут отрицать те и или иные обнародованные медиафакты о себе как (дип) фейковые, а общественность может заподозрить власть имущих в использовании дипфейков для сокрытия чувствительной информации. В первом случае нарратив дипфейка может быть использован, чтобы поставить под сомнение достоверность предъявленных «улик» — выставить объективные доказательства как результаты цифровой фальсификации. Этот риторический прием приносит своего рода «дивиденды лжеца», позволяя обвиняемым уклоняться от ответственности за свои заявления и действия [19]. Такая тактика может оказывать разрушительное воздействие и на авторитетную журналистику, поскольку обвиняемые могут игнорировать попытки СМИ привлечь их к ответственности за нарушения, представляя их как «фейковые новости».

Во втором случае нарративизация дипфейка носит противоположный вменительный характер на основе слухов и теорий заговора, но и здесь путаница и скептицизм вокруг видеофактов несут подрывной заряд деинституционализации. Например, в конце 2018 года начались спекуляции о здоровье президента Габона Али Бонго, который отсутствовал в общественной жизни несколько месяцев. Пытаясь положить конец слухам, правительство опубликовало видео, на котором Бонго произносит традиционное новогоднее обращение. Однако его необычное появление заставило многих пользователей социальных сетей, включая местных политиков, заявить, что это видео — дипфейк, что укрепило их подозрения, что правительство скрывает информацию о плохом состоянии здоровья или смерти Бонго. Через неделю после публикации видео на фоне растущих беспорядков военные Габона предприняли попытку государственного переворота. В видеоролике, анонсирующем переворот, военные упомянули странный внешний вид президента как доказательство того, что с ним что-то не так, хотя позже криминалистическая экспертиза не обнаружила признаков монтажа [15. С. 10].

### **Дипфейк как социальная (дис) функция**

С момента появления технологии GAN и ее популяризации через платформу Reddit возобладал алармистский дискурс о будущих угрозах ее недобросовестной эксплуатации. Хотя ряд экспертов предостерегал от преждевременной и слишком «истеричной» реакции [18; 20], общепринятый дискурс приписывает дипфейкам разрушительный характер, а сам термин стигматизирован, унаследовав свой негативный и провокационный смысл от фейковых новостей, бум которых наблюдался в период американской избирательной кампании 2016 года. Альтернативный нейтральный термин — «синтетические медиа», т.е. медиасообщения, сгенерированные ИИ, — не прижился [25].

Можно выделить несколько ключевых тем, вокруг которых строится обсуждение социального влияния дипфейков. Во-первых, поскольку дипфейк обычно вводит в заблуждение относительно онтологического статуса медиатекстов, это *источник медиа-манипуляций и обмана*. В частности, дипфейк грозит подорвать общественное доверие к визуальным медиа как традиционно самому надежному источнику информации и новостей [36]. Видеть (что-то) теперь не значит верить (этому). Во-вторых, дипфейки *разрушают политическую и гражданскую стабильность*, подрывая основы демократии, дестабилизируют доверие граждан к органам государственной власти и правовым институтам, вводят в заблуждение избирателей, угрожают общественной безопасности и обостряют социальные разногласия [4]. В-третьих, дипфейки *способствуют нарушению норм приватности*: создание реалистичных видео не всегда сопровождается получением разрешения на использование образа человека, что ведет к распространению компрометирующей и травмирующей информации, включая дипфейковый кибербулинг [7]. В-четвертых, дипфейки — *инструмент киберпреступности, кибервойн и кибершпионажа* [33].

Тенденция злонамеренного использования дипфейков в ближайшие годы будет нарастать в условиях геополитической напряженности, глобальной конкуренции и повсеместной цифровизации услуг [23]. Но значит ли это, что дипфейки следует запрещать и искоренять? ИИ-технология «глубокого синтеза» помогает совершать преступления, и подобная инструментальная роль в механизме преступного деяния ставит под сомнение необходимость принятия специальных мер по самостоятельной криминализации дипфейка [5; 10]. Проблема неправомерного использования той или иной технологии не имеет сугубо правового решения, и, более того, сама дискуссия о вреде технологий политически бессмысленна. С социологической точки зрения любая инновация является социальным конструктом, в рамках которого множество групп пользователей и разработчиков на этапе проектирования и внедрения постоянно обмениваются идеями, смыслами и значениями, определяя траекторию развития технологии посредством зримого и незримого дискурса по согласованию общественных интересов, потребностей, идеалов и соответствующих решений [26].

На заре эры компьютеров в 1940–1950-е годы отношение к ним также было крайне неоднозначным: одни видели в них спасение и решение всех земных проблем, другие определяли их как «чудовище», «зловещее изобретение», «монстра» и т.п. Компьютеры внушили страх деградации, потому что в них виделась власть «запятнать или принизить человека»: люди испытывали «гнев и беспомощное разочарование», потому что компьютер якобы «унижает человеческое достоинство» и способен привести к тому, что «людей заменят механические существа» [2. С. 472–504]. Дискурс страха перед сложными технологиями неизбежен, и сегодня он обретает новую воплоще-

ние в связи с развитием ИИ, однако, только осознав вездесущее формирующее влияние дискурса на техническое сознание, мы можем надеяться обрести контроль над самой техникой в ее воплощенной форме [2. С. 503].

Таким образом, важно говорить не только о том, чем нам угрожает техника, но и о том, как ее можно капитализировать социально значимыми проектами. Здесь можно также выделить несколько направлений.

1. *Изобразительное искусство* во многом является методом «глубокой обработки» реальности, особенно в модернистских жанрах, что типологически сближает дипфейки и арт. Неспроста одной из первых востребованных компьютерных программ в мире стал графический редактор Adobe Photoshop, который сохраняет свою популярность и сегодня, в эпоху ИИ [17]. Дипфейк предполагает новый уровень работы с человеческим воображением. Стремление фантазировать и воплощать свои фантазии в реальность — важнейший эволюционный источник развития человеческого потенциала. Люди фантазируют о многом, но мало кто в истории был способен реализовать фантазии материально, вербально или визуально. С технологией дипфейка эта творческая нереализованность преодолевается посредством визуального сторителлинга, который лежит в основе современного маркетинга, искусства и инфлюэнсига. В кинопроизводстве дипфейк уже активно используется, что позволяет, в частности, омоляживать актеров и возрождать их к жизни. Например, в сериале «Диверсанты» появляется капитан Калтыгин, герой актера Владислава Галкина, умершего в 2010 году: российские разработчики в области ИИ создали цифровую модель актера, которая могла двигаться, разговаривать и выражать эмоции. Но технология дипфейка за счет ее прогрессирующей доступности предоставляет возможности синтезировать образы практически всем, кто хочет реализовать свой творческий потенциал. Мечта обычного человека о творческой империи обретает реальность, а творчество, как никогда ранее, становится народной практикой. Важной традиционной составляющей популярной культуры всегда были тактики высмеивания и развенчания власти: анекдоты, карикатуры, граффити, мемы, пранки — все это формы народного сатирического обличения и осуждения нравов власть имущих. Дипфейки предлагают новую, более изощренную форму сатирической проблематизации и развития народной демократии, так как способны высвечивать вопросы, которые кажутся далекими или абстрактными. Например, пакистанская инициатива по борьбе с изменением климата «Проект Апология» изобразила мировых лидеров, извиняющихся в 2032 году за бездействие в решении экологических кризисов [24. С. 17]. Риторическая убедительность таких проектов обусловлена ошеломляющим висцеральным реализмом, помноженным на силу народной смеховой культуры, которая обнажает биготизм доминирующих дискурсов и их апологетов.

2. В *образовательной сфере* дипфейки обрели социальное призвание в период пандемии, например, британский рекламный холдинг WPP об-

учал сотрудников при помощи нейросети: пятьдесят тысяч человек, сидя дома, осваивали основы маркетинга по видеокурсам с виртуальным учителем, который вел лекции на трех языках одновременно (английском, испанском и китайском), но обращался к каждому сотруднику лично и по имени. Такой формат оказался очень выгодным: на создание двадцати роликов с виртуальным аватаром по технологии дипфейка WPP потратил сто тысяч долларов, а видео с настоящими актерами обошлось бы в десять раз дороже [31]. Сегодня интеграция ИИ-форматов в процесс обучения объясняется уже не вынуждающими условиями, а доказанной эффективностью, особенно в преодолении языкового барьера. Глобальная кампания по повышению осведомленности о малярии с участием Дэвида Бекхэма еще в 2019 году успешно разрушила языковые барьеры с помощью рекламы, в которой использовались визуальные и голосовые технологии, создающие впечатление, что Бекхэм говорит на нескольких языках. Дипфейки разрушают языковой барьер во время видеоконференций, переводя речь и одновременно изменяя движения лица и рта, чтобы улучшить зрительный контакт и создать впечатление, что все говорят на одном языке.

3. Конфиденциальность данных пациентов — один из важнейших приоритетов в системе здравоохранения любой страны: законы о защите персональных данных часто затрудняют получение и распространение медицинских изображений без идентификации личности пациентов [14]. Однако синтетические изображения обещают решить эту проблему: исследователи могут обеспечить обмен данными, сохраняя конфиденциальность пациентов и создавая реалистичные синтетические данные, имитирующие особенности групп больных. Помимо клинических изображений технология дипфейков может применяться для повышения эмпатии врачей. Исследователи из Тайбэйского медицинского университета [37] использовали технологию дипфейка для создания видеороликов, имитирующих черты лица реальных пациентов, чтобы улучшить способность врачей интерпретировать эмоциональные состояния пациентов, сохраняя их конфиденциальность. Система анализировала изображения пациентов и создавала их дипфейки, предоставляя врачам информацию о необходимости корректировать свое поведение в соответствии с ситуацией. Такой подход был направлен на создание более эмпатической среды, в которой пациенты чувствовали бы себя понятыми и поддержаными. По мере роста интереса к технологии дипфейков и повышения экономической эффективности вычислительных ресурсов можно ожидать более широкого внедрения соответствующих приложений в сфере здравоохранения. Дипфейки могут произвести революцию в медицинской визуализации, повысить точность диагнозов, защитить конфиденциальность пациентов и улучшить их взаимодействие с врачами.

4. *Бизнес*: компании по всему миру заинтересованы в технологиях дипфейка, применимых к бренд-менеджменту, поскольку они могут существен-

но преобразовать электронную коммерцию и рекламу. Например, с помощью дипфейка компании создают цифровых амбассадоров и продвигают бренд по всему миру. Так называемые динамические кампании (масштабная цифровая реклама с микро-таргетингом) становятся ключевым инструментом в арсенале маркетологов. Дипфейки помогают брендам привлечь самых востребованных знаменитостей и инфлюэнсеров из социальных сетей, а «звездам» — расширить сферу деятельности, соглашаясь вести модную рекламную кампанию и моделировать одежду, даже не приходя на фотосессию. Цифровые модели с разным тоном кожи, ростом и весом могут демонстрировать модные наряды для любой аудитории. Дипфейк позволяет создавать гиперперсонализированный контент, который превращает потребителей в моделей: технология дает возможность виртуально примерять одежду, чтобы увидеть, как она будет смотреться, перед покупкой, и может генерировать таргетированную рекламу, которая будет меняться в зависимости от времени, погоды и зрителя. Розничные продавцы также могут гиперперсонализировать обслуживание, используя цифрового помощника с «глубокой обработкой», который поможет с онлайн-запросами, будучи демографической и этнической копией клиента и говоря с ним на его родном языке. Вместо того, чтобы разговаривать с безликим ботом, покупатели уже сегодня могут общаться с «настоящим» лицом, что повышает доверие бренду. Кроме того, ИИ может создавать уникальные искусственные голоса, которые позволяют различать компании и продукты, упрощая брандинг.

Таким образом, с позиции социального конструирования техники (SCOT) дипфейк формируется и развивается в результате взаимодействия двух крупных, но конкурирующих социальных групп. Первая — алармистская группа, в которую входят в основном журналисты, правозащитники, жертвы кибермошенничества, военные, политики и законодатели. Они воспринимают дипфейк как угрозу национальной безопасности, инструмент дезинформации, подрыва институционального доверия и разрушения зоны приватности. Представители этой группы концентрируются на рисках кибербезопасности, политических манипуляциях и моральных издержках, связанных с утратой границы между подлинным и симулированным. Доминирующая позиция этой группы порождает общественный страх перед дипфейками, усиливая регулятивное давление на технологические инновации и формируя нормативный дискурс о контроле, запретах и уголовной ответственности. Вторая — вэлферистская — группа включает в себя программистов, независимых художников, культурных активистов, предпринимателей, врачей и работников образования. Для них дипфейк — прежде всего эффективный инструмент решения общественно значимых задач, укрепления гражданского общества и расширения человеческого опыта. Разработчики видят в дипфейках экспериментальную лабораторию для развития новых форм цифрового творчества, художники — медиум для политической сатиры и социальной

критики, врачи и педагоги — способ сделать коммуникацию более персонализированной и эмпатичной, бизнес — инструмент креативного маркетинга и улучшения клиентского опыта. Эти практики не только способны компенсировать рискогенность данной технологии, но и демонстрируют ее реальный потенциал как источника общественного блага (вэлферизм) — отсюда социально-конструктивистское влияние этой группы: она задает перспективу нормализации и культурной легитимации дипфейков в профессиональной и повседневной жизни.

Противостояние двух групп отражает борьбу за символическое господство в определении того, что считать «правильным» и «опасным» в применении технологий ИИ. Этот конфликт постепенно переходит в фазу поиска баланса в сфере правового регулирования и киберэтики. Имеющиеся правовые модели эффективно реализуются через регулирование распространения фальсифицированных аудиовизуальных материалов (например, Китай и Южная Корея требуют маркировки синтетического контента [4]). Корпоративные механизмы включают в себя автоматизированные системы распознавания дипфейков, фильтрацию и ограничение их загрузки на медиаплатформах, например, через технологию «управления сходством» (likeness management technology). Этические модели формируются посредством согласования профессиональных кодексов, прав человека и локальных политик, призыва к прозрачности протоколов «синтетических медиа» и уважению цифровой идентичности [24].

\*\*\*

Как мы видим, дипфейки по-разному, но эффективно реализуют важнейший тип социальной власти в современном обществе — власть цифровой презентации, которая, согласно М. Фуко, активно говернментализирует сферу власти на смешанные формы агентности. Проблема власти — ключевая в отношениях между ИИ и человеком, поскольку в них задействована сеть агентов — от государства, его институтов до программистов, гражданских активистов и рядовых пользователей, а также самого ИИ, который миметически и рекурсивно воздействует на человека. Кто кого контролирует, кто кому подчиняется и кто здесь главный — неясно, но с социологической точки зрения технологии, обретающие власть, тем более эпистемическую власть над человеком, требуют анализа не технических, а социальных аспектов их конструирования как процесса взаимного приспособления внутри сети. Дальнейшее развитие и адекватное применение дипфейков возможно только при условии признания их сетевой (социотехнической) природы. Необходимо сместить научный и общественный дискурс с алармистских заявлений о рискогенной феномена к изучению интеграции дипфейков в социальные практики, формированию медиаграмотности и киберэтики, по-

вышению уровня кибербезопасности и выявлению оптимальных моделей управления дипфейками в общественных интересах. Как и в случае с другими инновационными технологиями, ключ к контролю над дипфейками лежит не в запрете, а в глубоком понимании механизмов их социального функционирования и целенаправленном формировании контекстов их ответственного этического использования.

### **Библиографический список/References**

1. Аганова А.В., Катерный И.В. Моральный статус искусственного интеллекта: в поисках социологической перспективы // Социологический журнал. 2025. Т. 31. № 1 / Aganova A.V., Katerny I.V. Moralny status iskusstvennogo intellekta: v poiskakh sotsiologicheskoy perspektivy [Moral status of artificial intelligence: In search of a sociological perspective]. *Sotsiologichesky Zhurnal*. 2025; 31 (1). (In Russ.).
2. Александр Дж. Смыслы социальной жизни: Культурсоциология. М., 2013 / Alexander J. Smysly sotsialnoy zhizni: Kultursotsiologiya [The Meanings of Social Life: A Cultural Sociology]. Moscow; 2013. (In Russ.).
3. Брокман Дж. Искусственный интеллект — надежды и опасения. М., 2020 / Brockman J. Iskusstvennyy intellekt — nadezhdy i opaseniya [Artificial Intelligence — Hopes and Fears]. Moscow; 2020. (In Russ.).
4. Дремлюга Р.И., Моисеев В.В., Парин Д.В., Романова Л.И. Национальное правовое регулирование использования и распространения реалистичных аудиовизуальных поддельных материалов (deepfake): опыт Китая // Азиатско-Тихоокеанский регион: экономика, политика, право. 2022. Т. 24. № 4 / Dremlyuga R.I., Moiseytsev V.V., Parin D.V., Romanova L.I. Natsionalnoe pravovoe regulirovanie ispolzovaniya i rasprostraneniya realistichnyh audiovizualnyh poddelynyh materialov (deepfake): opyt Kitaya [National legal regulation of the use and distribution of realistic audiovisual counterfeit materials (deepfake): The case of China]. *Aziatsko-Tikhookeansky Region: Ekonomika, Politika, Pravo*. 2022; 24 (4). (In Russ.).
5. Ефремова М.А., Русскевич Е.А. Дипфейк (deepfake) и уголовный закон // Вестник Казанского юридического института МВД России. 2024. Т. 15. № 2 / Efremova M.A., Russkevich E.A. Deepfake i ugovolnyi zakon [Deepfake and criminal law]. *Vestnik Kazanskogo Yuridicheskogo Instituta MVD Rossii*. 2024; 15 (2). (In Russ.).
6. Жирар Р. Вещи, скрытые от создания мира. М., 2016 / Girard R. *Veshchi, sokrytye ot sozdaniya mira* [Things Hidden Since the Foundation of the World]. Moscow; 2016. (In Russ.).
7. Иванов В.Г., Игнатовский Я.Р. Deepfakes: перспективы применения в политике и угрозы для личности и национальной безопасности // Вестник РУДН. Серия: Государственное и муниципальное управление. 2020. Т. 7. № 4 / Ivanov V.G., Ignatovsky Ya.R. Deepfakes: perspektivy primeneniya v politike i ugrozy dlya lichnosti i natsionalnoy bezopasnosti [Deepfakes: Prospects of application in politics and threats to the individual and national security]. *RUDN Journal of Public and Municipal Administration*. 2020; 7 (4). (In Russ.).
8. Игнатьев А.Г., Курбатова Т.А. Дипфейки в цифровом пространстве: основные международные подходы к исследованию и регулированию // Центр глобальной ИТ-кооперации. 2023. № 3 / Ignatyev A.G., Kurbatova T.A. Dipfeyki v tsifrovom prostranstve: osnovnye mezhdunarodnye podkhody k issledovaniyu i regulirovaniyu [Deepfakes in the digital space: Main international approaches to research and regulation]. *Tsentr Globalnoy IT-Kooperatsii*. 2023; 3. (In Russ.).
9. Карпова Д.Н., Прокурина А.С. Социотехнический поворот в исследовании цифровизации общества // Власть. 2020. № 1 / Karpova D.N., Proskurina A.S. Sotsiotekhnichesk

- y poverot v issledovanii tsifrovizatsii obshchestva [The social-technical turn in the study of digitalization of society]. *Vlast*. 2020; 1. (In Russ.).
10. Калыгин В.О. Дипфейк как правовая проблема: новые угрозы или новые возможности? // Закон. 2022. № 7 / Kalyatin V.O. Dipfeyk kak pravovaya problema: novye ugrozy ili novye vozmozhnosti? [Deepfake as a legal problem: New threats or new opportunities?]. *Zakon*. 2022; 7. (In Russ.).
  11. Катерный И.В. Каузальные объяснения эффекта «зловещей долины» в роботехнике: теории и исследовательские данные // Качество и жизнь. 2017. № 4 / Katernyi I.V. Kauzalnye obyasneniya effekta “zloveshchey doliny” v robototekhnike: teorii i issledovatelskie dannye [Causal explanations of the “uncanny valley” effect in robotics: Theories and research data]. *Kachestvo i Zhizn*. 2017; 4. (In Russ.).
  12. Смирновский К. Deepfakes — что это и как ими пользуются мошенники / Smirnovsky K. Deepfakes — chto eto i kak imi polzuyutsya moshenniki [Deepfakes: What they are and how fraudsters use them]. 2024. URL: <https://iidx.ru/blog/deepfakes-chto-eto-i-kak-imipolzuyutsya-moshenniki>. (In Russ.).
  13. Тард Г. Законы подражания. М., 2011 / Tarde G. *Zakony podrazhaniya* [The Laws of Imitation]. Moscow; 2011. (In Russ.).
  14. Федеральный закон № 152-ФЗ от 27.07.2006 «О персональных данных» / Federalny zakon No. 152-FZ ot 27.07.2006 “O personalnyh dannyh” [Federal Law No. 152-FZ of July 27, 2006 “On Personal Data”]. URL: <https://minzdrav.gov.ru/documents/5402-federalnyy-zakon-152-fz-ot-27-iyulya-2006-g>. (In Russ.).
  15. Ajder H., Patrini G., Cavalli F., Cullen L. *The State of Deepfakes: Landscape, Threats, and Impact*. Deep Trace; 2019.
  16. Baaren R.B. van, Holland R.W., Kawakami K.L., Knippenberg A.F.M. van. Mimicry and prosocial behavior. *Psychological Science*. 2004; 15 (1).
  17. Braganza N. The most popular software and technical tools users picked on MuchSkills. 2025. URL: <https://www.muchskills.com/blog/top-software-technical-tools-muchskills>.
  18. Brandom R. Deepfake propaganda is not a real problem. 2019. URL: <https://www.theverge.com/2019/3/5/18251736/deepfake-propaganda-misinformation-troll-video-hoax>.
  19. Chesney R., Citron D. Deep fakes: A looming challenge for privacy, democracy, and national security. *California Law Review*. 2019; 107.
  20. Feldman B., Hart B., Read M. Are the deepfake fears overblown? *New York Magazine*. 2019; June 13.
  21. Gibbs A. Sympathy, synchrony, and mimetic communication. *The Affect Theory Reader*. Ed. by M. Gregg, G. Seigworth. Durham; 2010.
  22. Goodfellow I., et al. Generative adversarial networks. *Advances in Neural Information Processing Systems*. 2014; 3 (11).
  23. *Identity Fraud Report*. Entrust; 2025.
  24. *Just Joking! Deepfakes, Satire, and The Politics of Synthetic Media*. Witness; 2021.
  25. Kerner C., Risso M. Beyond porn and discreditation: epistemic promises and perils of deepfake technology in digital life worlds. *Moral Philosophy and Politics*. 2021; 8 (1).
  26. Klein H.K., Kleinman D.L. The social construction of technology: Structural considerations. *Science, Technology, & Human Values*. 2002; 27 (1).
  27. Korbak T., Balesni M., Barnes E., et al. Chain of thought monitorability: A new and fragile opportunity for AI safety. 2025. URL: <https://arxiv.org/abs/2507.11473>.
  28. Maras M.H., Alexandrou A. Determining authenticity of video evidence in the age of artificial intelligence and in the wake of Deepfake videos. *International Journal of Evidence & Proof*. 2019; 23 (3).
  29. Pinch T.J., Bijker W.E. The social construction of facts and artefacts: Or how the sociology of science and the sociology of technology might benefit each other. *Social Studies of Science*. 1984; 14 (3).

30. r/MyBoyfriendIsAI. URL: <https://www.reddit.com/r/MyBoyfriendIsAI/comment/1ldw3nn/> / my\_ai\_boyfriend.
31. Simonite T. Deepfakes are becoming the hot new corporate training tool. 2020. URL: <https://www.wired.com/story/covid-drives-real-businesses-deepfake-technology>.
32. Stel M., Vonk R. Mimicry in social interaction: Benefits for mimickers, mimickers, and their interaction. *British Journal of Psychology*. 2010; 101.
33. Taylor B.C. Defending the state from digital deceit: The reflexive securitization of deepfake. *Critical Studies in Media Communication*. 2020; 38 (1).
34. *The State of Deepfakes*. Sensity; 2024.
35. Wang C. et al. Neural codec language models are zero-shot text to speech synthesizers. *IEEE Transactions on Audio, Speech and Language Processing*. 2025; 33.
36. Whyte C. Deepfake news: AI-enabled disinformation as a multi-level public policy challenge. *Journal of Cyber Policy*. 2020; 5 (2).
37. Yang H.C., Rahmanti A.R., Huang C.W., Li Y.C.J. How can research on artificial empathy be enhanced by applying deepfakes? *Journal of Medical Internet Research*. 2022; 24 (3).

DOI: 10.22363/2313-2272-2025-25-4-603-618

EDN: KYPNJM

## **Deepfake as a social-technical artifact: A sociological analysis\***

**I.V. Katernyi, D.N. Karpova, E.V. Skulchuk**

MGIMO University,  
Vernadskogo Prospekt, 76, Moscow, 119454, Russia

(e-mail: katerny@mgimo.ru; d.karpova@inno.mgimo.ru; catherine.ccr@mail.ru)

**Abstract.** Deepfakes as digital artifacts created with deep machine learning have already become a common phenomenon with multiple forms, but social consequences of this technology are still controversial. The authors conducted a sociological analysis based on the social construction of technology theory (SCOT) and cultural sociology to identify three intertwined levels of the deepfake as an artifact: (a) technical — based on generative neural networks capable of creating highly realistic content and providing artificial intelligence (AI) with features of a perfect mimetic object, thus, overcome barriers of social recognition; (b) media — mechanisms of pragmatic impact, including creolization, intertextuality and metanarrativity; (c) social — functions and dysfunctions of deepfakes and also social domains that change under their influence. The social significance of deepfakes is ambivalent: on the one hand, this technology is highly risky, produces significant threats to privacy, cybersecurity, and social stability; on the other hand, it provides previously unseen opportunities for development in the fields of art, education, healthcare and business. This approach allows to overcome the limitations of the dominant alarmist discourse that reduces deepfakes to tools of disinformation and fraud, thus revealing the social construction of technical artifacts. Despite the obvious discursive stigmatization in the very portmanteau “deepfake”, it is important to see the satirical potential of such media as a tool for maintaining democratic culture. In the mixed communication perspective, deepfakes ensure visible prosopopean transmobility of AI, creating

\*© I.V. Katernyi, D.N. Karpova, E.V. Skulchuk, 2025

The article was submitted on 22.07.2025. The article was accepted on 14.10.2025.

new forms of social agency through the mechanisms of social mimesis. The most technically advanced deepfakes generate virtual personas capable of effective mimetic communication beyond the “uncanny valley”. Their quasi-human subjectivity is recognized as normal and credible, since the conditions for constructing a significant Other are mimetic: mimicry (of human visceral features); imitation (of social actions); emulation (of meaningfulness and empathy). As with other innovative technologies, the key to controlling deepfakes is not its prohibition but a deep understanding of its social construction together with reasonable modes of its management in the interests of the public good.

**Key words:** artificial intelligence (AI); deepfake; artifact; media; social construction of technology; cultural sociology; mimesis; satire

**For citation:** Katernyi I.V., Karpova D.N., Skulchuk E.V. Deepfake as a social-technical artifact: A sociological analysis. *RUDN Journal of Sociology*. 2025; 25 (4): 603–618. (In Russ.). <https://doi.org/10.22363/2313-2272-2025-25-4-603-618>