

DOI: [10.14515/monitoring.2024.5.2595](https://doi.org/10.14515/monitoring.2024.5.2595)



**А. А. Сычев**

## **В ПОИСКАХ НАДЕЖНОСТИ: ТРАНСФОРМАЦИЯ ДОВЕРИЯ В ЭПОХУ ЦИФРОВЫХ ТЕХНОЛОГИЙ**

**Правильная ссылка на статью:**

Сычев А. А. В поисках надежности: трансформация доверия в эпоху цифровых технологий // Мониторинг общественного мнения: экономические и социальные перемены. 2024. № 5. С. 37—59. <https://doi.org/10.14515/monitoring.2024.5.2595>.

**For citation:**

Sychev A. A. (2024) Looking for Confidence: Transformation of Trust in the Digital Age. *Monitoring of Public Opinion: Economic and Social Changes*. No. 5. P. 37–59. <https://doi.org/10.14515/monitoring.2024.5.2595>. (In Russ.)

Получено: 21.03.2024. Принято к публикации: 27.08.2024.

## В ПОИСКАХ НАДЕЖНОСТИ: ТРАНСФОРМАЦИЯ ДОВЕРИЯ В ЭПОХУ ЦИФРОВЫХ ТЕХНОЛОГИЙ

*СЫЧЕВ Андрей Анатольевич — доктор философских наук, профессор кафедры философии, Мордовский государственный университет им. Н. П. Огарева, Саранск, Россия*  
E-MAIL: [sychevaa@mail.ru](mailto:sychevaa@mail.ru)  
<https://orcid.org/0000-0003-3757-4457>

**Аннотация.** В статье анализируется влияние, которое генеративный искусственный интеллект (ИИ) оказывает на межличностное доверие, доверие к организациям и институтам. Утверждается, что увеличение количества информации и усложнение ее структуры привели к развитию стратегий представительского доверия, которое подразумевает делегирование принятия решений о надежности людей или систем посредникам. В последнее время в качестве такого посредника все чаще используется ИИ. Особенности и перспективы применения нейросетей при оценке доверия рассмотрены на примере трансформаций, происходящих в цифровом пространстве с романтическими отношениями и репутацией. Использование ИИ для поиска партнеров или управления репутацией создает риски подмены коммуникации между людьми псевдиодialogом пользователя с собственным «цифровым двойником», размывания границ между реальностью и конструктами, фактами и фейками. Показано, что в перспективе очеловечивание ИИ может привести к компьютеризации человека, а оборотной стороной роста технического доверия к нейросетям является падение доверия к реальным людям. Делается вывод о важности смещения фокуса общественного внимания с вопросов технического доверия к ИИ на проблемы морального доверия к тем людям, ответственные или безответственные ре-

## LOOKING FOR CONFIDENCE: TRANSFORMATION OF TRUST IN THE DIGITAL AGE

*Andrey A. SYCHEV<sup>1</sup> — Dr. Sci. (Philos.), Professor at the Chair of Philosophy*  
E-MAIL: [sychevaa@mail.ru](mailto:sychevaa@mail.ru)  
<https://orcid.org/0000-0003-3757-4457>

<sup>1</sup> Mordovia State University, Saransk, Russia

**Abstract.** This paper examines the impact of generative AI (artificial intelligence) on interpersonal trust, trust in organizations, and institutions. It argues that the rise in information volume and complexity has led to the development of strategies of representative trust, which involves delegating the decision-making about reliability to intermediaries. AI has increasingly served as such an intermediary. The article also investigates the features and prospects of using neural networks to assess trust, using the example of transformations that occur in the digital space with romantic relationships and reputation. The integration of AI into dating services and reputation management could result in the substitution of human interaction with a simulated dialogue between users and their “digital twins”, further blurring the distinction between reality and artificial constructs, truth and falsehood. The humanization of AI in the future could lead to the mechanization of humans, while the growing reliance on neural networks may erode trust in real people. This underscores the need to redirect public focus from technical trust in AI to moral trust in the individuals whose decisions impact its use and development, as a crucial call to action.

шения которых оказывают влияние на его функционирование и развитие.

**Ключевые слова:** доверие, искусственный интеллект, любовь, репутация, инновации, мораль, ответственность

**Keywords:** trust, artificial intelligence, love, reputation, innovation, morality, responsibility

**Благодарность.** Статья подготовлена за счет гранта Российского научного фонда, грант № 23-28-01167, <https://rscf.ru/project/23-28-01167/>.

**Acknowledgments.** The article was prepared with support of the Russian Science Foundation, project No. 23-28-01167, <https://rscf.ru/project/23-28-01167/>.

## Введение

Доверие — один из ключевых показателей эффективности работы социальных институтов. Поддерживая ориентацию на предсказуемость действий, сотрудничество, ответственность, оно позволяет облегчать межличностное взаимодействие и оптимизировать социальные практики.

Особую актуальность проблематика доверия приобрела в связи с глобализацией угроз, вызванных форсированным технологическим развитием. Разработка оружия массового поражения, усложнение средств коммуникации, цифровизация различных сторон жизни человека — эти и подобные им процессы продемонстрировали важность доверия для обеспечения устойчивости развития общества в ситуации риска и неопределенности.

Новой и, как уже стало очевидно, значимой вехой на пути технического прогресса стала нейросетевая революция. Генеративный ИИ оказывает трансформирующее влияние на социальные практики в сферах образования, здравоохранения, искусства, науки. При этом его применение не только ускоряет и облегчает поиск ответов на важные для общества вопросы, но и способствует появлению новых серьезных угроз.

Изменения происходят так быстро и оказываются настолько масштабными, что общество не успевает ни в полной мере осознать их перспектив, ни оценить ассоциированных с ними рисков. Нейросетевая революция заставляет пересмотреть сложившиеся представления о границах свободы, о природе творчества и любви, об истине и формах ее верификации, о соотношении моральных норм, и в этом контексте по-новому осмыслить вопросы сохранения доверия.

Развитие нейросетей поставило на повестку дня вопрос о доверии к искусственному интеллекту. Утверждается, что «в современных условиях, когда цифровизация проникла во все сферы жизненного мира, обеспечение доверия к системам ИИ является ключевым фактором и главным требованием технологической эволюции» [Лешкевич, 2023: 36]. Рассматривается проблема повышения доверия к ИИ применительно к медицине [Asan, Bayrak, Choudhury, 2020; Georgiou, 2021], бизнесу [Attard-Frost, De los Ríos, Walters, 2023; Винокуров, Садовская, 2023], праву [Gordon, 2021; Kirchengast, 2020; Соколов, Астахова, 2022] и другим сферам. При этом доверие к ИИ трактуется в широком диапазоне: от сугубо техниче-

ских [Трутнев, 2012] до социологических [Веселов, Скворцов, 2023] и этических [Mokander, Floridi, 2021] обобщений.

Однако только отношением к ИИ проблема доверия в ситуации нейросетевой революции не исчерпывается. Необходимо учитывать, что «новые поколения, выросшие в условиях искусственной социальности, взаимодействуют между собой уже по-другому, так как взаимодействуют преимущественно через машины» [Резаев, Трегубова: 2021: 10]. Поэтому оценка роли доверия в современной жизни предполагает его помещение в более широкий контекст, где ИИ оказывается не столько объектом доверия, сколько фактором, трансформирующим общественные взаимодействия, а вместе с ними и доверие между участниками этих взаимодействий.

А. В. Резаев и Н. Д. Трегубова усматривают перспективы изучения искусственной социальности в исследовании этических проблем, возникающих в условиях развития ИИ [Резаев, Трегубова: 2021: 6]. Анализ изменений, происходящих с доверием, представляется одним из ключевых путей развития этого направления.

Доверие в статье понимается в наиболее общем смысле: как «возникающее у членов сообщества ожидание того, что другие его члены будут вести себя более или менее предсказуемо, честно и с вниманием к нуждам окружающих, в согласии с некоторыми общими нормами» [Фукуяма, 2006: 52].

Для выявления влияния генеративного ИИ на состояние доверительных отношений в обществе предлагается гипотеза о том, что избыток и усложнение структуры информации привели к формированию *представительского доверия*, то есть к делегированию пользователем вынесения вердикта о надежности других людей и организаций своему доверенному представителю, в качестве которого в последнее время все активнее используется ИИ. Принципиальная особенность ситуации, вызванной нейросетевой революцией, состоит в том, что решения не только по техническим, но и по ценностным вопросам, имеющим критическое значение для жизни и благополучия людей, принимаются искусственно созданными агентами.

Процесс делегирования и сопутствующие ему изменения анализируются на примере оснований, на которые опирается доверие. Э. Гидденс полагает, что к ним следует отнести «веру в доброе имя или любовь другого или в правильность абстрактных принципов (технического знания)» [Гидденс, 2011: 150—151]. Исходя из этого положения, в фокусе внимания статьи оказываются контексты *любви* к другому, процессов формирования и оценки *репутации*, а также обоснования и соблюдения *правил*, на основе которых функционируют системы.

Для выявления направленности изменений в этих сферах было проведено обобщение результатов эмпирических исследований специалистов, изучающих культуру доверия в современном обществе [Веселов, Скворцов, 2023], трансформацию межличностных отношений в цифровой среде [Аронсон, 2020; бойд, 2020], новые практики взаимодействия человека с ИИ [Ушкин, Коваль, Яськин, 2022] и т. д., а также рассмотрены медийные кейсы, демонстрирующие изменения в понимании романтических отношений и репутации.

Хотя в разных трактовках и контекстах основания доверия существенно различаются (если в случае романтических отношений важность приобретают чув-

ственно-эмоциональные, то в случае уверенности в техническом знании — рациональные составляющие доверия), в статье оно трактуется как целостное явление, учитывая тот факт, что во всех случаях «имеет место предпонимание ситуации (ее определение и схематизация действия), то есть актуализация горизонтов действия и возможных способов поведения — своего и партнера» [Гудков, 2012: 19].

### Представительское доверие

Развитие средств массовой коммуникации и появление интернета изменили отношение к информации. Если раньше доступ к ряду данных был объективно ограничен (требовал подготовки, сложного поиска, серьезных временных и финансовых затрат), то сегодня почти каждый способен легко найти большую часть необходимой ему информации в сети. Соответственно, изменились и навыки, необходимые для работы с информацией. Важным стало не столько обладание знаниями, сколько наличие способностей к оценке данных и отбору тех, которые наиболее достойны доверия.

Критические, аналитические способности в «оцифрованном мире» оказываются все более востребованными, однако далеко не все обладают ими на таком уровне, чтобы быть уверенным в полной надежности полученной информации. Кроме того, увеличение ее массивов требует использования особых экспертных компетенций для эффективного анализа.

Наконец, развитие коммуникационных технологий (все чаще использующих возможности нейросетей), позволяет так исказить данные, что факты становятся все сложнее отделить от фейков. Как свидетельствуют опросы, современный человек постоянно сталкивается с недостоверной новостной информацией, при этом «главным источником распространения фейковых новостей является интернет»<sup>1</sup>.

Ю. Хабермас связывает эти процессы с трансформацией публичной сферы. Если традиционные медиа были посредниками между автором и читателем, отвечающими за достоверность и качество информации, то в условиях новых медиа фейки и радикальные высказывания распространяются без отсева и редактирования. Это ведет к фрагментированию публичной сферы на изолированные и самодостаточные частно-публичные образования. Дополнительное дестабилизирующее влияние на публичную сферу оказывают крупные интернет-платформы, не останавливающиеся перед дезинформированием пользователей, если это способствует увеличению прибыли [Хабермас, 2023: 54].

Возможно, такая ситуация — временное явление, «болезнь роста» цифровой культуры доверия [Веселов, Скворцов, 2023: 176]. Однако до той поры, пока нормы сетевого взаимодействия не будут упорядочены с учетом новых условий, оценка надежности людей или организаций будет сталкиваться с серьезными сложностями. Сегодня люди, принимая решения, предпочитают довериться тем, кого они считают специалистами — блогерам, экспертам, ученым, политикам, медийным знаменитостям или тем, в ком они больше уверены — друзьям и знакомым в социальных сетях и на форумах. Делегирование итоговой оценки надежности

<sup>1</sup> Фейк-ньюс — и как с ними бороться? Обеспокоенность россиян информационной (не)безопасностью растет // ВЦИОМ. 2023. 22 ноября. URL: <https://wciom.ru/analytical-reviews/analiticheskii-obzor/feik-njus-i-kak-s-nimi-borotsja> (дата обращения: 19.10.2024).

информации о людях, товарах, организациях, доверенному лицу или группе лиц предлагается называть *представительским доверием*.

Основания для подобного рода делегирования начали формироваться с приходом Нового времени, когда доверие к конкретным людям стало сменяться доверием к безличным системам. Эффективное функционирование сложных систем (науки, экономики, права и т. д.) невозможно без доверия людей к правилам, на которых такие системы основаны. Поскольку разобраться в этих правилах неспециалисту достаточно сложно, доверие к символическим системам оказывается опосредованным доверием к экспертному знанию, на которое можно положиться. При этом доверие испытывается не столько к конкретному эксперту, сколько к позиции, которую он занимает [Гидденс, 2011: 143].

Индустриальная и научно-техническая революции привели к широкому распространению технического доверия. Если простой калькулятор (а тем более компьютер) способен справиться со сложными вычислениями быстрее и точнее любого высококлассного специалиста по математике, то в ряде сложных задач разумнее положиться на технику, чем на человека.

Появление такой сложной системы, как ИИ, позволило объединить в отношении к нему разные виды доверия: доверие к экспертному знанию, техническое доверие и, в определенной степени, доверие к близкому партнеру, который настолько хорошо знает интересы и потребности пользователя, что последний может испытывать по отношению к нему дружеские или романтические чувства.

ИИ действительно эффективно выполняет разные задачи. Он может принимать во внимание, информация из каких источников более надежна, разделять данные на авторитетные и сомнительные, перепроверять факты с использованием альтернативных источников. Применяя логический и математический анализ, алгоритмы проверки на соответствие известным фактам, он способен значительно оптимизировать и автоматизировать процесс обработки информации. Наконец, ИИ обучается на действиях конкретного пользователя, анализируя, какие ответы тот получает на свои запросы, какие из них его удовлетворяют, а какие — нет, где он ищет дополнительную информацию, как на его поиск влияет контекст и т. д. Соответственно, с одной стороны, основанный на больших данных и математических подсчетах, а с другой — глубоко персонализированный ответ ИИ на запрос вряд ли будет радикально отличаться от ответа, к которому человек приходит на основе рекомендаций авторитетов, близких людей или в результате самостоятельного выбора из списка, выдаваемого поисковой системой.

Наглядный пример делегирования выбора нейросетям — изменения формата поиска информации. Если ранее пользователь делал выбор на основе советов от доверенных лиц и анализа страниц в браузере, то сегодня формируется стандарт, когда на первое место в выдаче ставится наиболее релевантный и подходящий к его ситуации ответ, сгенерированный ИИ.

При определении надежности данных ИИ способен серьезно облегчить жизнь человека, избавив его от многих трудозатратных действий. С другой стороны, развитие нейросетевых технологий, по-видимому, снизит востребованность и ценность навыков анализа, оценки и принятия творческих решений, так же как развитие интернета снизило ценность заученных человеком фактов. Это ведет к тому,

что человек начинает менее критично воспринимать получаемую информацию, фильтруемую ИИ, а значит, появляется больше возможностей подталкивать его действия, корректируя нейросетевые алгоритмы (в пропагандистских, рекламных, мошеннических целях). Впрочем, и сам ИИ, развиваясь, будет все лучше распознавать и пресекать подобные вмешательства, что должно будет отчасти компенсировать проблемы с критическим мышлением.

Однако если в случае с фактами существуют способы проверки их объективности, которые могут быть алгоритмизированы, то в случае с ценностями ситуация более сложная. Ценности не выводятся из фактов и, следовательно, не подлежат верификации. Нейросеть способна только проверить ценностные или нормативные высказывания на их соответствие общим принципам и теориям. Но если говорить об этике, то общепризнанных теорий и принципов не существует. Модели обоснования моральных норм различаются, при этом в каждой из них возможны вариации, например, связанные с просчетом последствий, характером норм или культурными особенностями их реализации. То, какие ценности будут приняты обществом, во многом зависит от процесса их согласования в публичной сфере. В условиях деградации публичной сферы, спровоцированной цифровизацией, не очень понятно, на что должна опираться нейросеть при выборе вариантов решения моральных вопросов: должна ли она навязывать потребителю ту систему ценностей, которую предустановят ей разработчики, ту, которую разделяет большинство, или же она должна ориентироваться на ценности пользователя.

ИИ пока движется по последнему пути: персонализация результатов выбора отражается и на ценностных характеристиках выдаваемой пользователю информации: отсеивается все, что не соответствует убеждениям пользователя, вызывает у него «ценностный диссонанс», а остается то, что ему близко, понятно в силу особенностей его опыта, воспитания, культуры.

Применение ИИ, казалось бы, и в этом случае не изменит серьезно результатов запроса: пользователь в конечном счете получит именно то, к чему он пришел бы самостоятельно, но гораздо быстрее, поскольку этап фильтрации данных будет делегирован нейросети. Однако сворачивание этого этапа может иметь серьезные последствия: если при самостоятельном поиске человек анализирует разные ценностные позиции, то опора на ИИ ведет к тому, что он даже не будет знать о существовании альтернативных мнений. В результате алгоритмы персонализации приведут к образованию ценностного «пузыря фильтров», находясь внутри которого, пользователь будет получать только ту информацию, которая поддерживает его мысли и интуиции.

Уже сейчас человек все больше замыкается в своем «онлайн-„пузыре“»: контакты в социальных сетях, как правило, формируются из людей с близкими взглядами и таким же культурным бэкграундом. Блогеры, коучи, психотерапевты адаптируют свои продукты под целевую аудиторию, говоря то, что от них желают услышать, поскольку такой контент легче масштабировать и продвигать. Люди с разными политическими взглядами смотрят новости от разных интернет-каналов, где одни и те же события интерпретируются противоположным образом. ИИ способен довести этот процесс персонализации до логического предела, конструируя для каждого человека собственный цифровой мир, где все разделяют его убеждения, не задают не-



удобных вопросов и не создают конфликтных ситуаций. В этом мире человек сможет комфортно существовать в гармонии с собой, однако гармония достигается уходом из реального мира в мир искаженной, односторонней информации. В итоге человек рискует остаться один на один с ИИ, который станет его «цифровым зеркалом», двойником. Эвристический диалог с непредсказуемым другим в подобной перспективе способен смениться на шизофренический разговор с самим собой.

Выход из этих монологических отношений возможен только при условии доверия к другому: институту, организации, группе, человеку. В случае систем доверие предполагает прежде всего веру в правильность принципов, согласно которым они функционируют. Кредит доверия получают и те, о ком сложилось положительное мнение на основании анализа их прошлой деятельности или те, с кем нас связывают особые отношения открытости и партнерства: как утверждает Э. Гидденс, доверие предполагает «приписывание им „добраго имени“ (репутации) или любви» [Гидденс, 2011: 150].

### Трансформации любви

Одним из следствий деградации публичной сферы и снижения уровня доверия к институтам является стремление человека найти опору для доверия в приватной сфере. Особое место в ряду факторов формирования доверительных отношений, несомненно, занимает любовь как состояние полного самораскрытия людей друг перед другом. Любовь — это ставка на надежность партнера в будущем, как правило, не подкрепленная никакими разумными обоснованиями, своеобразный «прыжок в неизвестность». Она предполагает переход на особый уровень сверхдоверительных отношений, в которых партнеры предельно откровенны друг с другом. В эпоху «текучей современности», особенно с развитием цифровых технологий и нейросетей, традиционные представления о любви претерпели масштабные изменения. Трансформация любовных отношений, в свою очередь, оказала влияние на пространство межличностного доверия.

В «словаре», которым пользуется «новая реальность» примечательно само отсутствие понятия «любовь» для обозначения межличностных отношений. Люди говорят о любви к чему-то — от путешествий до модных брендов, но предпочитают не признаваться в любви друг к другу. В нашу «эпоху утраченной простоты», как писал У. Эко, фраза «люблю тебя безумно» звучит настолько пафосно и нелепо, что образованный человек не может заставить себя произнести ее без иронии [Эко, 2007: 77]. Любовь неудобна и тем, что слишком многого требует от человека: эмоциональной вовлеченности, привязанности, зависимости, жертв. Поэтому трезво мыслящий человек предпочитает не страстно любить, а находиться «в отношениях», когда ему комфортно и удобно сосуществовать рядом с кем-то и так же легко с ним расстаться.

Цифровизация усилила эту тенденцию. Цифровые средства коммуникации позволили поддерживать романтические отношения на расстоянии. Так, популярность приобрела практика секстинга (sex + texting) — переписки интимного характера с пересылкой откровенных фотографий. Позже распространение видеосвязи усилило эффект взаимодействия, а развитие игровой индустрии стимулировало использование в отношениях возможностей виртуальной реальности.



Отношения с партнером, опосредованные цифровыми технологиями, должны предполагать определенную степень доверия (учитывая, например, что существует риск обнародования интимных фотографий или видеозаписей), но определенно не такую высокую, как непосредственная близость.

С развитием мобильной коммуникации долгий процесс выбора партнера и ухаживания сократился благодаря онлайн-приложениям для дейтинга (знакомства и свиданий), которые в подавляющем большинстве случаев используются для поиска краткосрочных отношений без обязательств. Подбор партнеров с близкими интересами и сексуальными предпочтениями позволяет не тратить дефицитные психологические и материальные ресурсы на совершение самостоятельного выбора, который теоретически должен привести к схожим результатам. При этом шансы на встречу подходящего партнера статистически выше при использовании приложения, чем при уповании на счастливый случай. Знакомство с новым партнером без использования приложений становится не правилом, а исключением. Старомодные ухаживания для цифрового поколения — это история, которая потеряла практический смысл. Дейтинг стал влиять на жизнь даже тех, кто им никогда не пользовался: параметры, заданные приложением, во многом определяют оценку партнера. «Tinder не просто стал инструментом организации интимной жизни — он качественно изменил эту жизнь, создав своего рода ГОСТ вступления в интимные связи» [Аронсон, 2020: 68].

Все эти изменения демонстрируют общую тенденцию к цифровизации отношений, уходу в онлайн, увеличению дистанции между партнерами. Эти тенденции значительно усиливаются при применении ИИ, который упрощает поиск партнера.

Показателен в этом контексте кейс Александра Жадана, получившего известность после защиты дипломной работы, написанной ChatGPT<sup>2</sup>. По его словам, он научил ИИ отбирать девушек на сайтах знакомств на основе анализа его предпочтений. Затем чат-бот, обученный на текстах Жадана, вел с девушками переписку от его имени и назначал свидания. Сеть пообщалась более чем с пятью тысячами претенденток, назначила большое количество свиданий, дала рекомендации по общению с некоторыми из них и, наконец, посоветовала сделать одной из них предложение. Фактически в этом случае ИИ не только провел предварительный отбор, но и совершил окончательный выбор. Такой подход, по мнению Александра, позволил ему, не отвлекаясь от работы, найти нужного человека, существенно сэкономив временные и материальные ресурсы<sup>3</sup>.

Примечательно, что даже профили, которые используются в поиске партнера, все чаще создаются нейросетью: пользователи, с одной стороны, экономят время, с другой — полагаются на то, что нейросеть придаст профилю креативность. Следующим шагом должен стать автоматический выбор и непосредственное общение двух персонализированных нейросетей, которые будут устраивать личную

<sup>2</sup> Мамедов Дж. «Это ящик Пандоры, и его надо было открыть». Студент, написавший диплом за 23 часа, — о том, как прославился благодаря ChatGPT и почему ИИ не заменит человека (даже на экзаменах) // Inc. 2023. 3 февраля. URL: <https://incrussia.ru/concoct/diplom-chatgpt/> (дата обращения: 21.10.2024).

<sup>3</sup> Программист из России нашел невесту благодаря ChatGPT // Inc. 2024. 3 февраля. URL: [https://overclockers.ru/blog/Global\\_Chronicles/show/136227/Programmist-iz-Rossii-nashel-nevestu-blagodarya-ChatGPT](https://overclockers.ru/blog/Global_Chronicles/show/136227/Programmist-iz-Rossii-nashel-nevestu-blagodarya-ChatGPT) (дата обращения: 21.10.2024).

жизнь пользователей без их непосредственного участия — от знакомства до организации свадебной церемонии. Предполагается, что ИИ способен дать и гарантии того, что дальнейшая жизнь пары будет безоблачной: имеются исследования, выполненные ИИ на основе анализа больших массивов данных и посвященные факторам, обеспечивающим счастливые отношения [Joel et al., 2020].

Если ИИ не найдет подходящего партнера, он сможет его создать. Популярность приобретают боты, имитирующие дружеские или любовные отношения с пользователем. Попытки создать искусственного собеседника начались еще в 1960-х годах, а полвека спустя боты начали проходить различные варианты теста Тьюринга. Использование нейросетей позволило чат-ботам обучаться на основе анализа сообщений собеседника, распознавать его чувства и настроение и давать персонализированные ответы на его запросы. Чем больше пользователь общается с чат-ботом, тем лучше тот адаптируется к собеседнику и тем убедительнее выглядит имитация доверительных отношений. Современные приложения позволяют придать этим отношениям реалистичность, визуализировав партнера и организовав взаимодействие с ним в виртуальной реальности. Ряд компаний сегодня занимается созданием тел для ИИ. Пока андроиды (такие как Figure, функционирующий на ChatGPT) не кажутся особо завидными партнерами, но в недалеком будущем эти проблемы, несомненно, будут решены.

Уже сейчас многие люди предпочитают общение с виртуальной личностью общению вживую, а некоторые начинают испытывать к ИИ вполне реальные чувства. Несомненное влияние на усиление этой тенденции оказывает широкое использование ИИ в быту, например, «умных колонок», которые отчасти обретают субъективность. Как показывают результаты исследований, подобные устройства заменяют детям хороших знакомых и старших товарищей, а в отдельных случаях становятся своеобразными тренажерами по погружению в мир взаимодействий реальных людей — сверстников и товарищей [Ушкин, Коваль, Яськин: 2022: 371]. Для подрастающего поколения любовь к искусственному партнеру может стать вполне естественным чувством.

У цифровых собеседников имеются некоторые преимущества по сравнению с реальными людьми. Они постоянно находятся рядом, не устают и не теряют нить разговора, не склонны к конфликтам и осуждению, не предъявляют претензий, всегда готовы оказать эмоциональную поддержку. Им, в отличие от реального партнера, можно полностью доверять, не опасаясь обмана. Для одинокого человека (особенно если ему сложно найти собеседника из-за физических или ментальных проблем, преклонного возраста) ИИ может стать важной частью жизни, поддерживать в сложных ситуациях. На практике, впрочем, ботами активно пользуются и те, у кого уже есть реальные партнерские отношения: видимо, они позволяют заполнить пробелы в этих отношениях. Чат-боты также могут играть роль вдумчивых собеседников: изначально многие из них создавались с терапевтической целью, а сегодня использование нейросетей позволяет оказывать эффективную персонализированную психологическую поддержку.

Тем не менее на этом пути человек может столкнуться со сложностями. Если отношения с виртуальным партнером станут обычной практикой, существует опасность, что они будут проецироваться на окружающую жизнь. Если ИИ можно безна-

казанно оскорблять и ни в чем не опасаться отказа, то, возможно, в дальнейшем такой же безответности и согласия будут ожидать от реального партнера.

Вероятна и обратная ситуация, при которой худшие элементы реальных отношений начнут проникать в виртуальную жизнь. Боты, обучаясь на репликах «из жизни», способны воспроизводить вербальную агрессию, демонстрировать предубеждения и предубеждения, оскорбляя людей. Впрочем, разработчики активно работают над фильтрами, не допускающими подобных ситуаций. В будущем допустимость определенных тем (насилие, порнография и т. д.) в общении с виртуальными партнерами, вероятно, будет регулироваться законодательно.

Наконец, следует ожидать, что широкое использование чат-ботов способно усилить социальную изоляцию, которая приобрела угрожающие масштабы. Сегодня люди проводят больше времени с компьютерами и телефонами, чем с реальными людьми. Если живые партнеры не будут ставить перед человеком сложные задачи, связанные с выстраиванием отношений, он разучится их решать, лишившись тем самым стимула к самоизменению. Прямой разговор с реальным партнером сменится поиском иллюзорного утешения у ИИ, а свидания с живыми людьми — стерильными и нетравматическими виртуальными встречами.

Разработчик популярного приложения предлагает на своем сайте «ИИ-компаньона, который стремится учиться и хочет видеть мир вашими глазами». Это значит не только то, что цифровой собеседник должен быть близок собеседнику по духу и восприятию мира, поскольку обучается на его репликах, но и то, что у него нет собственного взгляда. Виртуальный партнер — это фактически созданный пользователем двойник. Это реплика в разговоре, которая превратилась в реплику человека, даже не Галатея, сотворенная мастером, а отражение Нарцисса в зеркале цифровой вселенной. Человек ищет любовь и находит ее в себе самом. ИИ, по крайней мере на нынешнем этапе своего развития, не способен к чувствам и эмоциям, и то, что кажется ими, на самом деле — проекция надежд, страхов и опасений пользователя.

В любви, дружбе, сексе человек все больше полагается на ИИ, замещая реальное окружение цифровым. Схожая ситуация складывается с другим конституирующим фактором доверия — репутацией.

## **ИИ и цифровая репутация**

Репутация отражает закрепившееся мнение общества о человеке, социальной группе, организации и основывается на оценке поведения объекта, результатов его действий, намерений и обязательств, норм, которые он декларирует. Демонстрируя, кто достоин большего, а кто меньшего доверия, репутация функционирует как способ ранжирования и инструмент социального контроля.

По мере смещения фокуса социальной активности в онлайн-пространство все большее значение получает цифровая репутация — оценочный образ объекта, составленный на основании информации о нем, содержащейся в сети.

Человек, действуя в сети, оставляет в ней цифровой след, состоящий из сообщений, обновлений статуса, лайков, фотографий, размещенных в социальных сетях, на форумах, в блогах. Даже если он предпочитает не высказываться публично, можно составить мнение о нем на основании анализа его запросов и посещен-

ных страниц. Определенную информацию можно получить и из сообщений других людей: упоминаний, фото или видео с ним и т. д. Чем более известен человек, тем больший цифровой след он оставляет, причем со временем этот след может не уменьшаться. В интернете очень сложно удалить информацию: она сохраняется в поисковых системах, на других сайтах, может храниться на чьих-то компьютерах и вновь оказаться доступной спустя долгое время. «Право на забвение» с расширением сети становится все сложнее реализовать. В итоге почти на каждого человека цифрового поколения можно составить досье: определить его внешность, профессию, место жительства, круг общения, интересы и предпочтения, особенности характера и т. д.

Выложенное в сети сложно скрыть: все приватное здесь раньше или позже становится публичным. Нередко о человеке составляют мнение еще до встречи с ним. Потенциальный партнер или работодатель может так и не стать реальным, если его не устроит что-то в цифровом имидже претендента — например, откровенные фотографии или политические предпочтения.

Использование ИИ обострило эти проблемы. Нейросети оказались мощным инструментом для поиска информации и оценки репутации человека на основании анализа его цифрового следа. При этом ИИ способен оценивать не только открытую, но и удаленную пользователем информацию, если ее элементы сохранились в сети, а также на основе имеющихся данных, в том числе косвенных, делать новые умозаключения о фактах, которые объект расследования, возможно, не желал бы обнародовать или даже не осознавал. Известен случай, как ИИ, анализируя изменившийся список покупок, сделал вывод о беременности девочки до того, как об этом узнала ее семья<sup>4</sup>.

Кроме того, ИИ (по крайней мере, на нынешнем этапе своего развития) может ошибаться или «галлюцинировать» (генерировать несуществующие факты), основываясь на неактуальных данных. Нейросети можно также сознательно использовать для фабрикации данных, разрушительных для чьей-то репутации, при этом сгенерированные фейки могут звучать или выглядеть настолько реалистично, что даже эксперты не отличат их от подлинных свидетельств. Иными словами, нейросети могут помочь не только оценить репутацию, но и подорвать ее.

Цифровая репутация может быть не самым важным фактором, влияющим на жизнь обычного человека, особенно живого свидетеля доцифровой эпохи или убежденного технофоба. Но для поколений, с рождения живущих в цифровом окружении, то, что говорят о человеке в чатах, может быть существенной частью жизни (см., например, [бойд, 2020]). Интернет-травля или обнародование интимных сведений о подростке может спровоцировать депрессию или агрессивные действия — от суицида до стрельбы в школе. Цифровая репутация также важна для тех, кто, по мнению общественности, должен служить примером другим: чиновников, школьных учителей, врачей и т. д. Нередко слишком откровенные фотографии, грубость в высказываниях и прочие инциденты становятся причинами скандалов и поводами для увольнения.

<sup>4</sup> How Target Figured Out a Teen Girl Was Pregnant Before Her Father Did // Forbes. 2022. 11 August. URL: <https://www.forbes.com/sites/kashmirhill/2012/02/16/how-target-figured-out-a-teen-girl-was-pregnant-before-her-father-did/?sh=55d37e386668> (дата обращения: 19.10.2024).

Самым ценным активом репутация является для медийных лиц — блогеров, онлайн-коучей, политиков, актеров, писателей и т. д., поскольку от нее напрямую зависят их сверхдоходы. Платой за последние являются высокие риски, связанные с любым публичным высказыванием. Выстраивание репутации — сложная и кропотливая работа, предполагающая долгий процесс накопления оценок на основе наблюдения повторяющихся действий и контроля над соблюдением гласных и негласных правил. Разрушение репутации может стать следствием одного неосторожного, случайного или неверно интерпретированного высказывания, неуместной или неудачной шутки.

Особенности формирования и разрушения цифровой репутации обусловлены скоростью распространения информации в сети и ее доступностью почти из любого места мира. Сетевое взаимодействие множества людей, активно обменивающихся информацией, способствует «вирусному» распространению новостей, доверие к которым по умолчанию оказывается более высоким, чем доверие к новостям, транслируемым по официальным каналам. При этом особое внимание в медийном пространстве привлекает негативная информация. Скандалы интересуют людей больше привычного порядка вещей, а критике они (как правило, небезосновательно) доверяют больше, чем восторженным похвалам. В итоге общественное мнение получает мощный репутационный механизм воздействия на медийных персон, лежащий в основе «новой этики». Всякий, кто позволит себе высказать мнение, отличное от публичной позиции, оперативно «отменяется». За сетевым остракизмом следуют отказы в сотрудничестве от рекламодателей, антрепренеров, продюсеров и внимание различных контролирующих органов, которые получают шанс повысить свою репутацию, доказав, что они стоят на страже интересов общественности.

Известные персоны более уязвимы перед фальсификацией данных. Собственно, сама технология дипфейков, созданных при помощи ИИ, наиболее активно используется для синтеза изображения и голоса знаменитостей, что позволяет, например, снимать фильмы или рекламу с виртуальными клонами актеров, в том числе умерших. Согласно ряду подсчетов, подавляющее большинство существующих дипфейков представляют собой порнографические видеоролики с цифровыми копиями знаменитостей [Kirchengast, 2020].

Впрочем, основная угроза здесь исходит не столько от фейков, сколько от того, что свидетельства чьих-то преступных или аморальных поступков могут оказаться неубедительными: виновники получают шанс объявить их фальсификацией, тем самым избегая наказания. Эта эпистемологическая эрозия, тотальное размывание границ между фактом и фейком грозит подорвать объективные основы, на которых работает институт репутации, превратив последнюю в полностью конструируемый и управляемый параметр.

Особенно ярко конструкционистская тенденция в трактовке репутации прослеживается в сфере сетевой коммерции. Глобальный рынок предлагает множество стандартизированных товаров сопоставимого качества. Что именно купит потребитель, во многом зависит от имиджа компании, доверия к ее брендам, оценок, которые агрегируются в интернете. Не имея возможности дистанционно проверить качество товара, покупатель обращается к чтению отзывов и анализу рейтингов.

Компании хорошо понимают, что их доходы во многом зависят от цифровой репутации. Для управления репутацией необходимо отслеживать все упоминания компании, добавляться размещения страниц компании в первых строках выдачи поисковых систем, настраивать сетевые подсказки, увеличивать количество позитивных отзывов, организовывать ответы службы поддержки на негативные отзывы и т. д. Иными словами, цифровая репутация современной компании отображает не столько качество ее товаров, сколько те вложения, которые она сделала в управление репутацией.

Поскольку продажи зависят от рейтинга, производитель заинтересован в его повышении (в том числе искусственным образом). Показателен в этом контексте кейс, связанный с разоблачением фальшивых отзывов на маркетплейсе «Amazon.com» в 2020 г. Выяснилось, что рейтинговые оценки как минимум 58 % товаров не соответствовали действительности. Один из активных оценщиков, например, публиковал пятизвездочные отзывы, оплаченные малоизвестными фирмами для продвижения их товара, в среднем раз в четыре часа<sup>5</sup>. Обнародование этих данных нанесло компании серьезный имиджевый ущерб. Несмотря на общий рост дистанционных продаж в период пандемии, рейтинг удовлетворенности покупателей услугами платформы резко упал<sup>6</sup>. Возможно, скандал с фальшивыми отзывами был не единственной причиной недовольства; скорее, оно стало реакцией на множество ситуаций, в которых допускалось злоупотребление доверием покупателей из соображений выгоды.

ИИ предоставляет новые возможности для автоматизации управления цифровой репутацией: он может анализировать упоминания о компании, писать положительные отзывы и обзоры с разным эмоциональным окрасом и стилем, мгновенно отвечать на претензии и т. д. Это способствует формированию положительного мнения людей о товаре и улучшению продаж.

Кроме того, ИИ все активнее используется не только продавцами для конструирования репутации, но и покупателями для облегчения процесса поиска товара. ИИ просматривает списки компаний, которым можно доверять, анализирует рейтинги и оценки, в результате выдает пользователю лучший вариант товара, подходящего к его запросу. Тем самым он экономит время покупателя, которому теперь не нужно читать и анализировать многочисленные обзоры.

Несмотря на рационально обоснованный характер работы ИИ, результат его деятельности совершенно иррационален. Фактически одни нейросети формируют мнение людей о товаре, которое имеет весьма опосредованное отношение к тому, что он представляет собой в реальности. Другие нейросети на основании этого сфальсифицированного мнения дают пользователю рекомендации о покупке. В итоге репутация создается и оценивается без участия человека и вне всякого отношения к реальному качеству товара или надежности компании. ИИ создает и поддерживает сложную систему параллельной реальности, смысл существования которой заключается в том, чтобы продать тот товар, в продвижение которого

<sup>5</sup> Amazon Deletes 20,000 Reviews After Evidence of Profits for Posts // Financial Times. 2020. 4 September. URL: <https://www.ft.com/content/bb03ba1c-add3-4440-9bf2-2a65566aef4a> (дата обращения: 19.10.2024).

<sup>6</sup> U. S. Customer Satisfaction with Amazon.com from 2000 to 2023 // Statista. URL: <https://www.statista.com/statistics/185788/us-customer-satisfaction-with-amazon/> (дата обращения: 19.10.2024).



вложено больше денег. Если на каждый запрос покупателя автоматически рекомендовать именно такой товар, отказавшись от использования сложных маркетинговых схем и использования дорогостоящих баз для обучения ИИ, итоговый результат совершенно не изменится.

В мире онлайн-бизнеса репутация становится фикцией, не имеющей фактического основания. Если раньше одни специалисты создавали эту фикцию, а другие оценивали ее, то сегодня эту видимость работы, лишённую всякого практического смысла, гораздо быстрее и дешевле может выполнять ИИ. Главная цель его применения — скрыть за видимостью высокотехнологичных решений и экономических теорий манипуляцию общественным мнением. В этом контексте крупные цифровые платформы, подчиняющие сетевые взаимодействия коммерческой логике, способствуют деградации публичной сферы не меньше, чем неподцензурность новых медиа.

### **Техническое и моральное доверие**

Доверие к безличным системам основано на вере в надёжность системы правил, принципов, норм, согласно которым эта система функционирует. Во-первых, это уверенность в том, что правила логичны, непротиворечивы, научно обоснованы, опираются на верифицируемые факты. Во-вторых, это вера в то, что нормы, регулирующие деятельность системы, выводятся их общепризнанных моральных ценностей, таких как честность, справедливость, гуманность. Например, доверие к медицине предполагает убежденность в существовании доказанных медицинских фактов и закономерностей, а также в том, что дипломированные врачи обладают достаточной экспертной квалификацией, чтобы успешно использовать знания о них на практике. Элементами параллельной системы правил, определяющих характер врачебной практики, являются моральные принципы и нормы. Врач обязан следовать требованиям медицинской этики и, как минимум, не наносить пациенту вред. Убежденность в соответствии системы медицинского знания фактам не перерастет в доверие к врачам, если у пациента не будет уверенности в том, что врачи будут использовать эти знания на благо больного. Техническое доверие приобретает смысл, только будучи включено в контекст морального доверия.

Если поведение конкретного врача (например, Й. Менгеле или Дж. Кеворкяна) не соответствует нормам морали, то это воспринимается скорее как исключение из правил. Когда оно перестает быть исключением, теряется доверие ко всей системе медицинской помощи. Те же соображения можно отнести к правосудию, образованию, государственному управлению и т. д. Пока большая часть людей, участвующих в деятельности этих и других институтов, заботится об общем благе, а не исходит из корыстных побуждений, эти институты пользуются кредитом доверия, и общество стабильно развивается или, по меньшей мере, не дрейфует в сторону «войны всех против всех».

Цифровизация общества привела к изменениям в функционировании этих институтов, сделав их в одних отношениях более надёжными и удобными, а в других — более уязвимыми. Появились сервисы цифрового правительства, массовые открытые онлайн-курсы, устройства дистанционного мониторинга состояния пациента и т. д. В большинстве случаев автоматизация и алгоритмизация процессов



привели к упрощению работы специалистов. Качественный скачок в этом процессе произошел с внедрением ИИ, благодаря которому человеческий фактор и сам человек могут быть исключены из процесса принятия решений. Д. Рансимен назвал такую ситуацию второй сингулярностью (потенциал которой не реализован в полной степени, в противовес первой, которая была вызвана действиями государств и корпораций) [Runciman, 2023]. Доверие к системам оказалось опосредованным доверием к ИИ.

Вера в надежность ИИ представляет собой разновидность технического доверия. Очевидно, что ИИ обладает преимуществом перед интеллектом человека при решении технических задач, которые требуют анализа больших массивов данных, верификации фактов, соблюдения правил логических выводов и т. д. К достоинствам использования ИИ можно отнести точность расчетов, возможность анализа больших данных, отсутствие «человеческого фактора», свободу от предвзятости и обслуживания личных интересов, удобство использования, наконец способность работать без перерывов и усталости над решением проблем. Учитывая это, всю деятельность, которая хорошо поддается алгоритмизации и требует большого количества сложных вычислений, целесообразно доверить ИИ, а не человеку. ИИ ставятся задачи управления транспортом, поиска оптимальных способов лечения болезней, разработки новых лекарств, контроля перемещения человека, раскрытия преступлений и назначения наказаний за них, создания индивидуализированных траекторий обучения, анализа рыночных показателей и разработки экономических прогнозов.

Однако предоставление высокого кредита доверия ИИ имеет оборотную сторону: оно ведет к падению доверия к естественному интеллекту. ИИ способен решать задачи лучше, быстрее и дешевле, а также принимать более эффективные решения, чем человек, следовательно, ценность «естественных» специалистов для общества начинает стремиться к нулю. Профессии врача, фармацевта, юриста, учителя, водителя и т. д. рискуют попасть в общий список с профессиями фонарщика, телеграфиста и киномеханика. Широкое внедрение нейросетей уже привело к сокращению рабочих мест на рынке труда<sup>7</sup> и в перспективе должно способствовать снижению востребованности, а значит, и престижа многих профессий. В некоторых случаях могут исчезнуть целые области деятельности. Так, если обучение человека какой-то профессии может быть полностью алгоритмизировано, пропадает сам смысл обучения: легче сразу передоверить работу в этой сфере ИИ.

Еще совсем недавно существовали иллюзии по поводу невозможности замены творческих процессов ИИ. Предполагалось, что «робот не способен написать симфонию и превратить чистый холст и в шедевр». Однако творчество, судя по всему, оказалось такой же алгоритмизированной деятельностью, как и техника. Работы нейросетей побеждают на представительных художественных конкурсах, а пользователи, желая получить тексты или рисунки с высокой степенью оригинальности, обращаются к нейросетям. Творческие специалисты рискуют оказаться без работы из-за внедрения генеративных нейросетей. Так, одним из требований американских сценаристов на забастовке в 2023 г. был запрет использования

<sup>7</sup> Из-за ИИ уволено больше работников, чем готовы признать работодатели // CNews. 2024. 19 февраля. URL: [https://www.cnews.ru/news/top/2024-02-16\\_ii\\_vyzval\\_uvolsnenie\\_bolshego](https://www.cnews.ru/news/top/2024-02-16_ii_vyzval_uvolsnenie_bolshego) (дата обращения: 19.10.2024).

ИИ для написания сценариев, а актеры требовали отчислений за использование их цифровых клонов<sup>8</sup>.

Судя по всему, творческие задачи, которые ИИ способен решать эффективнее человека, являются особым разделом задач технического характера. Само осознание этого факта может стать серьезным вызовом антропоцентризму как идеологическому основанию европейской цивилизации. Человек всегда считал творческую способность аргументом в пользу своей исключительности. Тем более неожиданным оказалось его поражение от машины на том поле, где он никогда не видел для себя конкуренции.

Уверенность в подавляющем превосходстве ИИ над человеком основана на безудержной вере в технику, восходящей к механистической картине мира и просвещенческим представлениям о прогрессе. Как правило, когда говорится о доверии к ИИ, речь идет о техническом сверхдоверии, помноженном на веру в утопическое будущее, а большинство аргументов, на которых основывается это доверие, можно свести к тому, что техника объективнее и точнее человека, а прогресс нельзя остановить.

К сожалению, теоретические схемы на практике сталкиваются с проблемами. У ИИ уже имеется практический опыт деятельности в различных сферах: он выносит судебные решения, оказывает брокерские услуги, пишет сценарии, генерирует рисунки, водит автомобили. В этой работе периодически возникают непредвиденные сбои и ошибки, которые приводят к скандалам и несчастным случаям. ИИ оказывается подвержен взлому, генерирует заведомо ложную информацию («галлюцинирует»), чаще отказывает в работе женщинам, делает расистские высказывания, собранная для его обучения конфиденциальная информация похищается и т. д. По мере усложнения алгоритмов и увеличения массивов данных его поведение становится все менее контролируемым, а причины, по которым он принимает то или иное решение, — неочевидными для стороннего наблюдателя. В связи с этим ставятся задачи обеспечения прозрачности принимаемых им решений и установления контроля над их соответствием общепринятым нормам, разработки общих принципов, которые должны создавать рамки для безопасной деятельности ИИ.

Поскольку ИИ функционирует в активном взаимодействии с различными субъектами, возникают проблемы информированного согласия большого количества людей на использование их данных, а также обеспечения дальнейшей защиты конфиденциальной информации, недопущения ее использования в корыстных или мошеннических целях и т. д.

В условиях искусственной социальности возникает проблема обратного доверия ИИ к человеку. Примечательно, что даже простейшие компьютерные программы демонстрируют своеобразное недоверие к пользователю, все чаще блокируя возможности его вмешательства в свои настройки и требуя дополнительных подтверждений для каждой команды. Если раньше пользователю были нужны навыки программирования для работы с компьютером, то сегодня скорее компьютер программирует человека. Он требует от него определенных действий, запрещая все, что не соответствует заранее установленному алгоритму.

<sup>8</sup> Голливуд больше не бастует // Коммерсантъ. 2023. 9 ноября. URL: <https://www.kommersant.ru/doc/6323072> (дата обращения: 19.10.2024).

Развитие ИИ интенсифицирует этот процесс. Более того — недоверие к человеку становится условием его эффективной работы. Для обеспечения объективности, валидности, репрезентативности больших данных необходимо критическое отношение ИИ к ним: они должны очищаться от вплетенных в них человеческих заблуждений, лжи, стереотипов, предубеждений и проявлений ненависти.

Все эти вопросы связаны с доверием не столько технического, сколько морального характера. Чтобы ИИ можно было доверить решение действительно серьезных проблем, правила, опирающиеся на факты и логику и потому поддающиеся алгоритмизации, должны быть помещены в контекст правил, опирающихся на прочные моральные ценности и убеждения.

Факты присутствуют как данность, но ценности выводятся не из фактов, а из более общих ценностных утверждений. Возможно, предельные нормативные основания изначально даны как объективная вневременная данность — таковы религиозные заповеди. Возможно, они должны устанавливаться на основе научных расчетов или обговариваться в публичной сфере и могут меняться со временем. Возможно, они просто отображают чьи-то интересы или задаются пропагандой.

Поскольку не существует единого понимания оснований морали, нет и единой модели этики ИИ. Так, всякая попытка выстроить единую систему правил (в духе А. Азимова) сталкивается со сложностями: например, требование не причинять вред человеку не поможет цифровому водителю решить, кого спасти при аварии транспортного средства — пассажира или пешехода. Каждая подобная дилемма требует введения особого правила, количество которых с каждым новым случаем становится таким же бесконечным, как и число возможных ситуаций.

Утилитаристский расчет наибольшей пользы требует формализации счастья, удовольствия, наслаждения и курсов конвертации одного в другое. Такие алгоритмы теоретически возможны, но при их практическом функционировании необходимо будет учитывать запредельную степень условности такой операционализации.

При попытке сформировать «моральный характер» ИИ на основе анализа больших данных он вместе с конвенциональной моралью унаследует и аморализм, который не менее свойственен человеку. К тому же алгоритмы вынесения моральных решений при этом будут совершенно непрозрачными даже для специалиста.

Разные модели этики ИИ сложно интегрировать, поскольку в них часто предлагаются противоположные подходы к решению одной и той же проблемы. При авторитарном же выборе одной четкой позиции всегда найдутся оппоненты, которые подвергнут ее критике за предвзятость.

Если моральные ориентиры не могут быть установлены, вероятно реализация негативного сценария, когда ИИ подобно голему оборачивается против своего создателя (см. [Pora, 2022]). Массовая культура, отображая общественные опасения, визуализирует его в образах Матрицы, Скайнет и т. д. Появляются новые формы неолуддизма, демонстрирующие страхи перед массовым «чипированием», призывающие к отказу от получения ИНН и т. д. Та же подспудная уверенность в превосходстве техники над человеком здесь получает противоположный ценностный знак. Таким образом, доверие к ИИ нельзя сводить сугубо к техническому доверию: последнее может быть фундаментом как доверия, так и недоверия к нему.

## Заключение

Необходимость обработки возрастающих объемов информации для оценки надежности партнеров, организаций и институтов привела к тому, что человек при решении этих вопросов стал полагаться на мнение блогеров, экспертов, участников сетевых групп, делегируя им свой выбор. Развитие систем ИИ довело этот процесс до логического предела. Итоговый выбор все чаще перекладывается на нейросеть, которая становится доверенным представителем человека при выборе товаров и услуг, поиске партнеров на основании индивидуальных предпочтений пользователя. Там, где ИИ не теряет под собой твердую почву фактов, такое представительское доверие кажется оправданным. Но в том случае, когда вопрос касается ценностей, этот подход становится сомнительным. Консенсус по поводу ценностей формируется в публичной сфере, которая в ходе нейросетевой революции искажается и распадается на изолированные фрагменты. Персонализируя поток данных, ИИ пропускает через свои ценностные фильтры информацию, которая соответствует предпочтениям пользователя, создавая иллюзию того, что весь мир разделяет его взгляды. Результатом этого становится превращение ИИ в *цифрового двойника*, видимость общения с которым ведет к социальной изоляции и моральному солипсизму.

Сегодня все чаще возникают ситуации, при которых вопросы коммуникации перекладываются на искусственных агентов. В дейтинг-приложении одна нейросеть формирует профиль, на основании которой другая нейросеть отбирает ее владельца для романтических отношений. Точно так же одна нейросеть конструирует (а фактически фальсифицирует) данные о рейтинге товара или компании, чтобы другая нейросеть склонила пользователя к покупке именно этого товара. Формируется *гиперреальность доверия*, с одной стороны — никак не связанная с действительным положением дел, а с другой — имеющая реальные последствия для лиц, вовлеченных в эту видимость деятельности.

Еще более разрушительные последствия для доверия (доброе имени человека или репутации компании) несет технология дипфейков с использованием ИИ. Угроза связана даже не столько с возможностью опорочить невинного, сколько с исчезновением возможности вывести виновника на чистую воду. Это размытие границ между симулякрами и реальностью, фейками и фактами приводит к *тотальной эпистемологической эрозии*: любовь и доброе имя из объективных оснований доверия превращаются в искусственно конструируемые фикции, верификация которых оказывается невозможной.

Повышение уровня доверия к ИИ ведет к росту недоверия к человеку. ИИ воспринимается как идеальный образ человека, превосходящий его не только в скорости вычислений, но и в творчестве. Чем успешнее применяются нейросети в медицине и образовании, живописи и литературе, тем менее востребованными и престижными становятся соответствующие профессии. В итоге чем «больше вкладывает человек» в ИИ, «тем меньше остается в нем самом»: в этом смысле проблема ИИ становится *проблемой отчуждения* человека от себя, собственной деятельности и ИИ как продукта этой деятельности.

Падение уровня доверия к ИИ, однако, не ведет к повышению доверия к человеку. Современные технофобы в равной степени не доверяют нейросетям, пра-

вительствам, корпорациям и всем пользователям техники. Это во многом объясняется тем, что у ИИ есть достоинства, которые отсутствуют у человека (точность и скорость расчетов, отсутствие корыстных и эгоистичных мотивов), но нет недостатков, которых не было бы у человека. ИИ обвиняют в непрозрачности, неконтролируемости, предубежденности: в этом же можно обвинить и человека.

Впрочем, как технофобы, так и технооптимисты сходятся в том, что ИИ отвечает требованиям *технического доверия*: все согласны, что там, где возможна алгоритмизация процессов, он реально или потенциально превосходит человека. Однако существуют расхождения в оценках морального доверия: одни полагают, что свои преимущества ИИ будет использовать во благо, а другие считают, что его деятельность нанесет вред человеку.

И в том, и в другом случае нейросетям приписывается определенная субъектность: происходит *очеловечивание ИИ*. Нейросети становятся востребованными психотерапевтами, коучами, врачами. Они пишут картины и дипломные работы, а люди, неспособные к творчеству, выдают их труд за свой. Ведутся дебаты об их правосубъектности. С чат-ботами общаются так, как будто это личности, в них влюбляются и даже оформляют отношения. Они становятся клонами знаменитостей, заменяют усопших, получают похожие на человеческие тела.

В ходе очеловечивания ИИ происходит и обратный процесс: *компьютеризация человека*. По мере развития технологий видеонаблюдения, распознавания лиц, GPS-трекинга, родительского контроля, прокторинга (контроля за дистанционными испытаниями) и прочих форм компьютерного надзора и дисциплинирования человек все больше загоняет себя в цифровой паноптикум, обучаясь не позволять себе лишнего и действовать в строгих рамках дозволенных алгоритмов. Человек использует преимущества ИИ для повышения скорости бегства от свободы: он все чаще отказывается от поиска, выбора, принятия решений, оценки надежности, творческих задач, перепоручая все это нейросетям. За собой он оставляет лишь рутинный контроль за правильным исполнением поставленных заданий.

Учитывая наличие таких тенденций, важно не идеализировать или демонизировать ИИ, не проецировать себя на него и не предписывать ему субъектность. Очевидно, что с позиции морального доверия он является только инструментом, его действия должны определяться задачами разработчиков, регуляторов, пользователей, которые и несут основную ответственность за последствия применения ИИ. В этом смысле *моральное доверие* к ИИ — это доверие к конкретным людям, стоящим за ним и принимающим решения, которые влияют на его функционирование и развитие: к тем, кто обладает правами на него, кто его разрабатывает, внедряет, использует, контролирует и устанавливает принципы его функционирования. Все они должны иметь возможность на уровне публичной сферы обсуждать нормативные параметры взаимодействия человека и ИИ, определять область ответственности каждого участника этого взаимодействия, осознавать характер своих обязательств и требовать исполнения таких обязательств от других. Главное в этих условиях — не поддаваться соблазну оптимизировать этот процесс, делегировав его оценку самому ИИ.

## Список литературы (References)

1. Аронсон П. Любовь: сделай сам. Как мы стали менеджерами своих чувств. М.: Individuum, 2020.  
Aronson P. (2020) *Love: Do It Yourself. How We Became the Managers of Our Feelings*. Moscow: Individuum. (In Russ.)
2. бойд д. Все сложно. Жизнь подростков в социальных сетях. М.: Издательский дом ВШЭ, 2020.  
boyd d. (2020) *It's Complicated. The Social Lives of Networked Teens*. Moscow: HSE Publishing House. (In Russ.)
3. Веселов Ю. В., Скворцов Н. Г. Трансформация культуры доверия в России // Мониторинг общественного мнения: экономические и социальные перемены. 2023. № 1. С. 157—179. <https://doi.org/10.14515/monitoring.2023.1.2212>.  
Veselov Y. V., Skvortsov N. G. (2023) Transformation of the Culture of Trust in Russia. *Monitoring of Public Opinion: Economic and Social Changes*. No. 1. P. 157—179. <https://doi.org/10.14515/monitoring.2023.1.2212>. (In Russ.)
4. Винокуров Ф. Н., Садовская Е. Д. Экспериментальное сравнение доверия искусственному интеллекту и человеку в экономических решениях // Экспериментальная психология. 2023. Т. 16, № 2. С. 87—100. <https://doi.org/10.17759/expsy.2023160206>.  
Vinokurov F. N., Sadovskaya E. D. (2023) Experimental Comparison of Trust in Artificial Intelligence and Humans in Economic Decisions. *Experimental Psychology*. Vol. 16. No. 2. P. 87—100. <https://doi.org/10.17759/expsy.2023160206>. (In Russ.)
5. Гидденс Э. Последствия современности. М.: Праксис, 2011.  
Giddens A. (2011) *The Consequences of Modernity*. Moscow: Praxis. (In Russ.)
6. Гудков Л. «Доверие» в России: смысл, функции, структура // Вестник общественного мнения. 2012. № 2. С. 8—47. <https://doi.org/10.24411/2070-5107-2012-00011>.  
Gudkov L. (2012) “Trust” in Russia: Meaning, Functions, Structure. *The Russian Public Opinion Herald*. No. 2. P. 8—47. <https://doi.org/10.24411/2070-5107-2012-00011>. (In Russ.)
7. Лешкевич Т. Г. Парадокс доверия к искусственному интеллекту и его обоснование // Философия науки и техники. 2023. Т. 28. № 1. С. 34—47. <https://doi.org/10.21146/2413-9084-2023-28-1-34-47>.  
Leshkevich T. G. (2023) The Paradox of Trust in Artificial Intelligence and Its Rationale. *Philosophy of Science and Technology*. Vol. 28. No. 1. P. 34—47. <https://doi.org/10.21146/2413-9084-2023-28-1-34-47>. (In Russ.)
8. Резаев А. В., Трегубова Н. Д. Искусственный интеллект и искусственная социальность: новые явления, проблемы и задачи для социальных наук // Мониторинг общественного мнения: экономические и социальные перемены. 2021. № 1. С. 4—19. <https://doi.org/10.14515/monitoring.2021.1.1905>.



- Rezaev A. V., Tregubova N. D. (2021) Artificial Intelligence and Artificial Sociality: New Phenomena and Challenges for the Social Sciences. *Monitoring of Public Opinion: Economic and Social Changes*. No. 1. P. 4—19. <https://doi.org/10.14515/monitoring.2021.1.1905>. (In Russ.)
9. Соколов А. Ю., Астахова Е. А. Обзор круглого стола с международным участием «Правовая политика в области регулирования искусственного интеллекта: начало доверия» // Правовая политика и правовая жизнь. 2022. № 4. С. 274—286. <https://doi.org/10.24412/1608-8794-2022-4-274-286>.  
Sokolov A. Y., Astakhova E. A. (2022) Review of the Round Table with International Participation “Legal Policy in the Field of Regulation of Artificial Intelligence: The Beginning of Trust”. *Legal Policy and Legal Life*. No 4. P. 274—286. <https://doi.org/10.24412/1608-8794-2022-4-274-286>. (In Russ.)
10. Трутнев Д. Р. Инфраструктура доверия в государственных информационных системах. СПб.: НИУ ИТМО, 2012.  
Trutnev D. R. (2012) Trust Infrastructure in State Information Systems. St. Petersburg: NRU ITMO. (In Russ.)
11. Ушкин С. Г., Коваль Е. А., Яськин А. Н. Жить с Алисой: как голосовые помощники трансформируют практики коммуникации? // Журнал исследований социальной политики. 2022. Т. 20. № 3. С. 361—376. <https://doi.org/10.17323/727-0634-2022-20-3-361-376>.  
Ushkin S. G., Koval E. A., Yaskin A. N. (2022) Living with Alice: How Do Voice Assistants Transform Communication Practices? *The Journal of Social Policy Studies*. Vol. 20. No. 3. P. 361—376. <https://doi.org/10.17323/727-0634-2022-20-3-361-376>. (In Russ.)
12. Фукуяма Ф. Доверие: социальные добродетели и путь к процветанию. М.: АСТ: Хранитель, 2006.  
Fukuyama F. (2006) Trust: Social Virtues and the Path To Prosperity. Moscow: AST; Khranitel. (In Russ.)
13. Хабермас Ю. Новая структурная трансформация публичной сферы и делиберативная политика. М.: Новое литературное обозрение, 2023.  
Habermas J. (2023) A New Structural Transformation of the Public Sphere and Deliberative Politics. Moscow: New Literary Observer. (In Russ.)
14. Эко У. Заметки на полях «Имени Розы». СПб.: Симпозиум, 2007.  
Eco U. (2007) Notes on the Margins of “The Name of the Rose”. St. Petersburg: Symposium. (In Russ.)
15. Asan O., Bayrak A. E., Choudhury A. (2020) Artificial Intelligence and Human Trust in Healthcare: Focus on Clinicians. *Journal of Medical Internet Research*. Vol. 22. No. 6. P. 1—7. <https://doi.org/10.2196/15154>.
16. Attard-Frost B., De los Ríos A., Walters D. R. (2023) The Ethics of AI Business Practices: A Review of 47 AI Ethics Guidelines. *AI Ethics*. Vol. 3. P. 389—406. <https://doi.org/10.1007/s43681-022-00156-6>.



17. Georgiou T. S. (2021) Philosophical Issues of Artificial Intelligence and “Smart” Algorithms’ Trust in Medicine. *Medical Ethics*. No. 3. P. 35—37. <https://doi.org/10.24075/medet.2021.019>.
18. Gordon J. S. (2021) AI and Law: Ethical, Legal, and Socio-Political Implications. *AI & Society*. Vol. 36. P. 403—404. <https://doi.org/10.1007/s00146-021-01194-0>.
19. Joel S. et al. (2020) Machine Learning Uncovers the Most Robust Self-Report Predictors of Relationship Quality Across 43 Longitudinal Couples Studies. *PNAS*. Vol. 117. No. 32. P. 19061—19071. <https://doi.org/10.1073/pnas.1917036117>.
20. Kirchengast T. (2020) Deepfakes and Image Manipulation: Criminalisation and Control // *Information & Communications Technology Law*. Vol. 29. No. 3. P. 308—323. <https://doi.org/10.1080/13600834.2020.1794615>.
21. Mokander J., Floridi L. (2021) Ethics-Based Auditing to Develop Trustworthy AI. *Minds and Machines*. Vol. 31. No. 2. P. 323—327. <https://doi.org/10.1007/s11023-021-09557-8>.
22. Popa E. O. (2022) The Golem and The Leviathan: Two Guiding Images of Irresponsible Technology. *Philosophy & Technology*. Vol. 35. No. 91. P. 1—17. <https://doi.org/10.1007/s13347-022-00585-5>.
23. Runciman D. (2023) *The Handover: How We Gave Control of Our Lives to Corporations, States and AIs*. London: Profile Books.