

УДК 177

DOI: 10.24412/2713-1033-2023-3-64-78

А. А. Сычев

Национальный исследовательский Мордовский
государственный университет им. Н.П. Огарева,
Саранск, Россия, e-mail: sychevaa@mail.ru

ЦЕННОСТЬ ДОВЕРИЯ В ЭПОХУ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА И НОВОЙ ЭТИКИ: ПРОБЛЕМЫ И ВЫЗОВЫ¹

В статье рассмотрена трансформация отношений доверия в условиях развития информационного общества. «Цифровое доверие» рассматривается в двух аспектах: как уверенность человека в надежности цифровых систем и как доверие к людям, опосредованное современными информационными технологиями. Сверхдоверие к цифровым технологиям рассмотрено в контексте технооптимизма. Это доверие подрывается ограниченностью, непрозрачностью, предвзятостью алгоритмов искусственного интеллекта, возможностью технических сбоев и злоупотреблений при его функционировании, что порождает иную крайность – цифровое сверхнедоверие. Демонстрируется, что опосредованность общения в цифровой среде, создающая возможности для введения собеседника в заблуждение, подрывает доверие к людям. По мере совершенствования алгоритмов ИИ (искусственного интеллекта) появляется все больше возможностей для создания «дипфейков», ИИ все легче выдать за человека, а результаты его работы – за авторский продукт.

Еще более актуальной проблемой является усиливающаяся поляризация ценностей в интернет-пространстве. Формирование «пузыря фильтров» создает условия, при которых пользователь утверждает в приоритетности личных интересов и безальтернативности собственной ценностной позиции, а любые противоположные позиции воспринимает как изначально неверные. В статье показано, как особенности сетевой коммуникации способствуют формированию, с одной стороны, норм «новой этики», а с другой – усилению позиций сторонников коллективной морали. Эта поляризация ценностей обостряет конфликтность в современном обществе, поддерживая доверие к одним за счет исключения других. Автор делает вывод, что отношения доверия можно выстроить лишь в том случае, если общество и индивид будут готовы согласовывать свои интересы и двигаться навстречу друг к другу.

Ключевые слова: доверие, информатизация, «новая этика», цифровые системы, искусственный интеллект

¹ Материалы исследования представлены в докладе на круглом столе «Человек vs технологии в эпоху больших данных и искусственного интеллекта» в рамках XIII Международной Грушинской социологической конференции (Москва, 25-27 мая 2023 г.)

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

A. A. Sychev

National Research Mordovia State University,
Saransk, Russia, e-mail: sychevaa@mail.ru

THE VALUE OF TRUST IN THE ERA OF ARTIFICIAL INTELLIGENCE AND NEW ETHICS: PROBLEMS AND CHALLENGES¹

The article examines the transformation of trust in the context of the development of the information society. “Digital trust” is considered in two aspects: as a person’s confidence in the reliability of digital systems and as trust in people, mediated by modern information technologies. Overtrust in digital technologies is considered as a modern form of techno-optimism. Trust is undermined by the narrowness, opacity, bias of artificial intelligence algorithms, the possibility of technical failures and malicious interference during its functioning, which gives rise to another extreme – digital super-distrust. It is demonstrated that the indirectness of communication in the digital environment, which creates opportunities for misleading the interlocutor, undermines trust in people. As AI (artificial intelligence) algorithms improve, more and more opportunities arise for creating “deepfakes”; it becomes increasingly easier to take AI as a person, and the results of its work as an original product.

Another problem is the polarization of values in the Internet space. The formation of a “filter bubble” creates conditions under which the user affirms the priority of personal interests, the lack of alternative to his own value position, and perceives any opposing positions as initially incorrect. The paper shows how the features of network communication contribute to the formation of the norms of “new ethics” and, at the same time, to the support of the value system based on the priority of the common good. This confrontation exacerbates conflict in modern society, maintaining trust among one group of people at the expense of excluding other groups. The author concludes that relationships of trust can only be built if society and the individual are ready to coordinate their interests and move towards each other.

Keywords: trust, informatization, “new ethics”, digital systems, artificial intelligence

Acknowledgment: the study was supported by the Russian Science Foundation (RSF), project no. 23-28-01167, <https://rscf.ru/project/23-28-01167/>

¹ The research materials were presented in a report at the XIII International Sociological Grushin Conference “Reconstruction of the World: Research of (in) a New Reality” (Moscow, May 27, 2023)

Введение

В последние десятилетия информационные технологии прочно закрепились во всех сферах общественной жизни. Межличностное взаимодействие в значительной мере опосредовано социальными сетями: с их помощью знакомятся, ищут работу, просят и получают советы. Современную экономику невозможно представить вне системы электронных платежей и технологий блокчейна. Государственные услуги оказываются населению в режиме онлайн, документооборот переводится в электронный вид, для идентификации личности используются цифровые подписи. Наука, образование, правосудие, медицина все активнее используют технологии больших данных. Темпы информатизации многократно ускорились в условиях локдауна во время пандемии COVID-19, сделав особо актуальным вопрос о надежности информации, полученной через социальные сети, к электронному правительству и цифровой экономике, к возможностям ИИ. Анализ ценностных изменений в системе взаимодействия человека и техники и в межличностных отношениях подтверждает, что «в цифровую эпоху вопрос доверия приобретает особенно важное значение» [Nandhakumar, 2004: 77].

Доверие – сложное и многомерное понятие, рассматриваемое исследователями с разных позиций и на различных уровнях теоретического обобщения [Селигмен, 2002; Фукуяма, 2008; Штомпка, 2012]. В настоящем исследовании оно рассматривается в наиболее общем смысле: как «уверенность в надежности человека или системы в отношении некоторого данного множества ожидаемых результатов или событий, где эта уверенность выражает веру в доброе имя или любовь другого или в правильность абстрактных принципов (технического знания)» [Гидденс, 2011: 150-151].

Применительно к современному обществу можно говорить о процессе серьезной трансформации доверия под воздействием новых технологических инноваций [Веселов, 2020; Перов, 2021; Ушкин, 2023] и ценностных конфликтов [Эпштейн, 2021; Коваль, 2021, Жадунова, 2023 и др.].

Целью статьи является выявление особенностей функционирования доверия в условиях цифрового общества. В ней, с одной стороны, ставится вопрос о надежности новых *цифровых систем* (социальных сетей, элементов цифровой экономики, электронного правительства, ИИ), а с другой – прослеживаются тенденции катастрофического падения доверия к *людям*, общение с которыми в пространстве сетевой коммуникации приобретает конфликтный характер.

Доверие к цифровым системам

Вера человека в надежность цифровых систем является современным проявлением технооптимизма – характерной для культуры Нового времени веры в то, что технологии превосходят возможности человека и потому на них можно переложить всю сложную и монотонную работу, повышая тем самым качество человеческой жизни. Так, компьютер надежнее человека, поскольку способен быстрее и точнее производить вычисления, не имеет предвзятого отношения к

кому-либо, не подвержен влиянию настроения. ИИ может найти более эффективные решения в тех областях, где человеческий мозг не справляется с планированием из-за необходимости принимать во внимание слишком большое количество взаимозависимых факторов. Поэтому при сложных многомерных проблемах на компьютерные технологии обработки и анализа больших данных будет разумнее полагаться, чем на человека.

Технооптимизм применительно к цифровым технологиям по мнению В.Ю. Перова базируется на нескольких допущениях: «а) этические проблемы связаны с людьми; б) цифровые технологии являются беспристрастными, то есть этически нейтральными; в) если «убрать» людей из их функционирования, то цифровым технологиям можно доверить практически все. Технологии становятся trustworthy в смысле «благонадежными»; г) пропадает доверие как проблема, поскольку люди не могут помешать ей функционировать посредством любых форм контроля, а сами технологии являются самодостаточными и безошибочными (или минимально ошибочными по сравнению с людьми). Эту ситуацию можно обозначить как формирование установки на «сверхдоверие» [Перов, 2021: 215].

С точки зрения технооптимизма системы ИИ должны взять на себя большую часть обязанностей врачей, юристов, чиновников, преподавателей и других профессионалов, обслуживающих ключевые сферы общественной жизни. Они способны справиться лучше, чем человек, с вождением машин и поездов, пилотированием самолетов. В перспективе возможно превращение «слабого» ИИ в «сильный». Массовая культура, учитывая эти тенденции, описывает возможные сценарии будущего, где ИИ превращается в равноправного партнера человека, и ставится вопрос о его моральном и правовом признании [Сычев, 2020: 141-172].

Действительно, информационные технологии существенно облегчили жизнь человечества, однако в ходе их применения проявились серьезные проблемы, которые ставят оптимистические прогнозы под сомнения и подрывают установки на «сверхдоверие».

Одна из центральных проблем связана с необходимостью учитывать сложные, многомерные и динамично изменяющиеся условия реальности при принятии морально значимых решений, то есть решений, которые предполагают выбор между моральными ценностями и могут иметь серьезные последствия для тех, кого они затрагивают. Например, при обучении ИИ вождению возникают типичные вопросы о том, как следует поступать самоуправляемой машине в ситуации выбора и какие ценности должны быть для нее определяющими. Вряд ли общие принципы способны стать четким ориентиром для конкретных действий в экстремальных ситуациях. Так, широко известны законы робототехники, предложенные Айзеком Азимовым, первый из которых гласит: «Робот не может причинить вред человеку или своим бездействием допустить, чтобы человеку был причинен вред». Однако реальность гораздо сложнее нормативных схем: на практике периодически складываются условия, в которых этот закон нельзя применить, поскольку спасение одного человека (например,

пассажира самоуправляемого автомобиля) может привести к гибели другого (пешехода) или наоборот. В этом случае возникает необходимость разработки критериев для сопоставления моральной ценности возможных жертв.

Разработчики пытаются очертить все значимые факторы, учитывая которые ИИ должен сделать моральный выбор. В эксперименте «Moral Machine» обобщаются ответы респондентов с разным культурным опытом на вопросы о том, что в подобных случаях должна сделать машина: продолжать движение или свернуть; спасать пассажиров или пешеходов, мужчин или женщин, молодых или старых; как относиться к нарушителям правил дорожного движения и животным; как учитывать численность групп и т.д. Всего предлагается девять факторов, способных, по мнению разработчиков, повлиять на выбор [Awad, 2018: 59-64]. Однако очевидно, что, если даже будет установлено устраивающее всех количественное соотношение всех этих факторов и их комбинаций (что само по себе более, чем проблематично), всегда будут возникать новые, не учтенные ранее факторы и нештатные ситуации, которые невозможно заранее предугадать и прописать в программе. В этом аспекте доверять ограниченному жесткими схемами ИИ жизнь людей было бы, по меньшей мере, неразумно.

Для того чтобы справиться с этой проблемой, самообучаемые системы ИИ должны корректировать свои алгоритмы, увеличивая количество рассчитываемых параметров и используя большие объемы данных (часто недоступные для обработки обычным человеческим мозгом). Учитывая эти преимущества, возможно, они смогут найти более эффективные способы разрешения нештатных ситуаций, чем большинство обычных водителей автомобилей. Однако при этом возникает иная проблема: количество принимаемых во внимание параметров становится настолько большим, что даже специалисты оказываются неспособны разобраться в логике принятия решений. В таком случае ни понять их, ни проконтролировать работу алгоритмов ИИ невозможно: они представляют собой «черный ящик». Фактически жизнь человека ставится в зависимость от решений ИИ, базовые основания для формирования которых (пути от конкретных норм до общих принципов) не могут быть прослежены. Насколько разумным будет слепое доверие таким решениям? Более безопасным представляется запрет на использование «непрозрачных» алгоритмов, по крайней мере в наиболее критичных областях человеческой жизни (медицина, военное дело, юриспруденция и т.д.).

Таким образом, если алгоритмы доступны для понимания и контроля, то они не способны учесть всего многообразия факторов, влияющих на решение; если же ИИ обучается анализировать большое количество факторов, его алгоритмы становятся «непрозрачными».

Еще одна проблема связана с тем, что действия ИИ нельзя назвать в полной мере беспристрастными – они зависят или от установок разработчиков, на программном уровне закладывающих ценностные приоритеты для принятия решений, или от мнений тех людей, действия или суждения которых были проанализированы в ходе обучения ИИ. Все эти мнения не свободны от

предвзятости, которые впоследствии будут устойчиво воспроизводиться и в алгоритмах принятия решения.

Так, для создания алгоритмов вождения беспилотных автомобилей необходимо определиться с базовыми ценностями. Алгоритмы будут действовать согласно вложенным в них разработчиками ценностным установкам: религиозным или философским, утилитаристским или деонтологическим и т.д. Поскольку согласия по поводу приоритетности тех или иных установок не существует, любая система «предустановленных» ценностей будет восприниматься оппонентами как предвзятая.

Если ИИ сам вырабатывает алгоритмы выбора, обучаясь на реальных примерах, он автоматически приобретает и предрассудки с этим выбором связанные. Так, алгоритм, который применялся для приема на работу в компании Amazon, проанализировав данные о приеме в компанию за предыдущее десятилетие, стал чаще отказывать женщинам. Нечто похожее случилось при использовании ИИ для назначения типа и срока наказания: система Compas, основываясь на статистике рецидивов, предлагала более жесткие наказания чернокожим преступникам. Чат-бот Tay компании Microsoft, обучаясь в процессе общения с пользователями Twitter, стал позволять себе расистские и сексистские высказывания и т.д.

Для того чтобы решения ИИ были в полном смысле непредвзятыми, он должен учиться распознавать и отсеивать предрассудки (даже в тех случаях, когда о них не подозревают их носители). Здесь можно говорить о специфическом понимании цифрового доверия (пока еще в метафорическом значении) как уверенности в надежности системы моральных ценностей, субъектом которого на этот раз становится сам ИИ, а объектом – люди, совершающие свой выбор (а возможно и другие системы ИИ). Теперь ИИ чтобы получать информацию, которой можно доверять, приходится проверять ее надежность: сравнить данные разных независимых источников, выявить конфликты интересов т.д. Тем не менее, маловероятно, что ИИ сможет полностью избавиться от предрассудков, уже в силу того, что сами предрассудки трактуются неоднозначно (например, одни считают таковыми отказ относиться ко всем группам одинаково, а другие – отсутствие предпочтений для дискриминируемых групп).

Даже если эта проблема имеет удовлетворительное программное решение, и ИИ начнет приходить к относительно непредвзятым выводам, это еще не значит, что ему можно будет безоговорочно доверять. Самые эффективные алгоритмы способны дать сбой из-за непредвиденных технических неполадок. «Сверхдоверие» в подобных условиях может привести к катастрофическим последствиям. В.Н. Рисухин, анализируя крушение самолета А-320 во Франции в 1992 году, делает вывод, что ее причиной могло стать чрезмерное доверие пилотов к бортовой автоматике: «Хотя ошибка, вызвавшаяся в чересчур крутом снижении, могла быть легко выявлена по необычному пространственному положению самолета (сильно опущенной носовой части фюзеляжа), пилоты полностью положились на систему автоматического управления полетом»

[Рисухин, 2006: 25]. При этом чем сложнее становятся технические системы, тем больше уязвимостей в них появляется. В случае с «непрозрачными» алгоритмами ИИ многие ошибки, вызванные техническими сбоями, нельзя будет даже идентифицировать.

Наконец, нельзя исключить и злонамеренное вмешательство в работу цифровых систем. Ускорение процесса цифровизации деятельности человека одновременно приводит к росту киберпреступности, увеличивая тем самым количество рисков в разных сферах жизни, включая угрозы для жизни людей. Так, в результате хакерской атаки на ИТ-систему университетской клиники в Дюссельдорфе на некоторое время вышли из строя аппараты жизнеобеспечения. Как предполагают журналисты, это стало причиной смерти одной из пациенток¹. По мере цифровизации различных сфер человеческой деятельности возникает все больше таких случаев, а их последствия оказываются все катастрофичнее.

Возможность ошибок и злоупотреблений, ограниченность, непрозрачность, предвзятость являются факторами, существенно подрывающими доверие к цифровым системам. Опасения по их поводу являются питательной почвой для новой волны антицифрового луддизма. В крайних своих формах это недоверие выражается в отрицании методов электронной идентификации, отказах от ИНН, страхе перед массовым чипированием и т.д., в контексте той или иной теории заговора. В массовой культуре эти фобии воплощаются в образах Skynet, Матрицы и прочих систем ИИ, обернувшихся против своих творцов и поработивших их.

Как чрезмерное доверие к цифровым системам, так и чрезмерное недоверие к ним представляются опасными крайностями. В современном информационном мире необходим поиск *оптимума* доверия, равноудаленного как от чрезмерного технооптимизма, так и от иррациональной технофобии.

Доверие к человеку в условиях цифровизации

По мере развития информационного общества общение людей становится все более опосредованным. Новые технические средства создают условия для формирования цифровых проекций человека, замещающих собой оригинал. Человек может сконструировать в сети свой виртуальный образ, не совпадающий с реальным, скрыть или исказить информацию о себе, выдать себя за кого-то другого.

Привлечение к этому процессу ИИ позволило создать технологию дипфейков, которая позволяет имитировать голос другого человека, подделывать фотографии или видеозаписи с его участием, а также синтезировать образы личностей, не существующих в реальности [Ушкин, 2023: 71-80]. В этих условиях всегда присутствует вероятность, что собеседник в сети может быть

¹ Eddy M., Perlroth N. Cyber Attack Suspected in German Woman's Death // The New York Times. Sept. 18, 2020. URL: <https://www.nytimes.com/2020/09/18/world/europe/cyber-attack-germany-ransomware-death.html>.

совсем не тем, за кого он себя выдает, а значит уровень доверия к нему должен быть существенно ниже, чем при общении лицом к лицу.

Возможность скрыться за маской или псевдонимом нередко приводит к т.н. «эффекту растормаживания», при котором моральные барьеры, ограничивающие поведение человека в реальной жизни, в интернет-пространстве ослабевают или исчезают. Иллюзия невидимости и ощущение безнаказанности нередко приводит к тому, что человек начинает позволять себе чересчур резкие и оскорбительные высказывания по отношению к некоторым лицам и группам. Немодерируемое обсуждение любых чувствительных тем рано или поздно приводит к обмену прямыми оскорблениями, на фоне которого вопрос о доверии ставиться вообще не может. При общении лицом к лицу люди лучше контролируют свои высказывания уже в силу того, что рискуют стать объектом ответного физического воздействия. В итоге поведение человека может радикальным образом отличаться от поведения его проекции, сконструированной в интернете. В этом случае возникает вопрос о том, что является истинным лицом, а что маской. Возможно, именно в условиях анонимности человек демонстрирует свои истинные убеждения (судя по высказываниям на интернет-форумах и в социальных сетях, все чаще предполагающие ненависть к другому и ксенофобию)? Если это так, то серфинг в виртуальном пространстве способен убедить многих пользователей, что не стоит чрезмерно доверять людям и в реальной жизни.

По мере совершенствования нейросетей взаимодействие человека с ИИ из имитации общения фактически превращается в полноценную коммуникацию [Ушкин, 2022: 361-376]. Поэтому по мере бурного развития технологий все чаще возникают ситуации, в которых собеседник выдает себя даже не за другого человека, а за человека, не являясь им на самом деле.

Для того, чтобы проверить, является ли собеседник человеком или машиной, используются различные тесты. Однако параллельно с усложнением таких тестов развиваются и способности ИИ их обходить. Так, Google для того, чтобы подтвердить, что пользователь не является роботом, использует снимки из Street View (где необходимо отметить все снимки со светофорами, пешеходными переходами, дорожными знаками и т.д.). Результаты прохождения этих тестов в дальнейшем используются для обучения ИИ управлению беспилотными автомобилями. Иными словами, с одной стороны, Google разрабатывает тесты для выявления ИИ, а с другой – обучает ИИ, как с этими тестами справиться.

Алан Тьюринг полагал, что рано или поздно ИИ научится обходить любые тесты, то есть в общении его невозможно будет отличить от человека. В этой связи возникает вопрос о нормах отношения к ИИ, обладающему такими возможностями: если его поведение будет в полной мере человеческим, следует ли и относиться к нему, как к человеку?

Хотя анонимность все еще является важной характеристикой сетевого общения, по мере усиления административного и общественного контроля над интернетом ее сфера сужается. Процессы информатизации не только облегчают

возможности коммуникации, но и предоставляют государству широкие возможности для осуществления контроля за деятельностью человека на всех ее уровнях. Новые технологии позволяют аккумулировать конфиденциальные данные, регулировать доступ к различным сервисам и локациям, проверять частную переписку, разговоры и публичные высказывания на предмет нарушений закона, отслеживать передвижения, финансовые транзакции и сетевую деятельность. Все чаще высказывания с применением «языка вражды» становятся либо основанием для судебного разбирательства (если власть видит в них угрозу для себя), либо поводом для сетевого остракизма (если они противоречат убеждениям сторонников «новой этики»). В этой ситуации люди предпочитают скрывать свое мнение либо притворно демонстрировать согласие с властью или активистами. В итоге далеко не всегда следует доверять тому, что в интернет-пространстве позиционируется как мнение большинства.

Таким образом, как полная анонимность, так и жесткий внешний контроль подрывают доверие к людям, общение с которыми опосредовано новыми цифровыми технологиями.

Еще одна проблема, влияющая на уровень доверия в обществе, связана с тем, что с некоторыми задачами ИИ справляется не только так же хорошо, как человек, но значительно быстрее и лучше человека. Поэтому все чаще возникают случаи, когда люди выдают результаты деятельности ИИ за свои. Так, широкую известность получил кейс студента РГГУ Александра Ждана, который признался в том, что защитил диплом, написанный нейросетью ChatGPT. Сегодня система «Антиплагиат» распознает тексты, сгенерированные ИИ, однако, как и в случае со Street View, разработка алгоритмов, обходящих эти ограничения – вопрос техники и времени. Доступность текстов в интернете уже превратила написание большей части рефератов, курсовых и дипломных работ в копирование и перефразирование, а успехи ИИ на этом поприще и вовсе ведут к кризису перепроизводства учебных и научных текстов, в массе которых сложно будет вычленивать что-то действительно новое и значимое.

Другой известный кейс представлен победой на фотоконкурсе Sony World Photography Awards 2023 «фотографии», сгенерированной ИИ. Подмена оказалась бы незамеченной, если бы конкурсант, представивший работу на суд экспертов, не отказался от награды, заявив, что он просто желал проверить, готово ли жюри конкурса к такого рода мошенничеству. Кейс с фотографией демонстрирует, что ИИ способен заменить человека не только в рациональной деятельности, но и в креативной. Если раньше казалось невозможным, что робот «может создать симфонию или превратить чистый холст в шедевр», то теперь очевидно, что эти вопросы вполне решаемы. Новые возможности нейросетей могут привести не только к увеличению числа подлогов, но и к таким масштабным изменениям в понимании искусства и креативности, которые по степени влияния будут сопоставимы с изобретением фотографии или с возникновением массового искусства.

Современная ситуация требует изменений в подготовке специалиста, оценке его профессиональных способностей. Если в отдельных профессиях ИИ

способен заменить человека так, что их нельзя будет отличить друг от друга, возможно эти профессии (в том числе творческие) будут делегированы ИИ, а человеку стоит сосредоточиться на таких направлениях профессиональной деятельности, которые ИИ недоступны. В этом случае ИИ станет официальным помощником в написании текстов, подготовке презентаций, проведении исследований, создании иллюстраций, написании музыки и т.д. Так, уже сегодня в США большинство сделок на бирже совершаются ИИ.

Возможно, человеку следует оставить за собой роль лишь контролера качества: в таком случае не будет необходимости выдавать результаты деятельности ИИ за свои. Однако, насколько правильной будет передача в ведение ИИ вопросов творчества и резервирование за человеком рутинного контроля? Не получится ли, что по мере «очеловечивания» ИИ, повышения его творческих потенций, человек, выполняющий однообразную работу, сам будет «компьютеризироваться»?

Усиление цифрового надзора над деятельностью человека – еще один путь к тотальной регламентации и сокращению пространства свободного самовыражения. М. Эпштейн полагает, что в этих условиях «человеку придется самому стать механизмом, управлять собою с помощью компьютерной программы, чтобы соответствовать всем условиям и ожиданиям электронно вооруженного общества» [Эпштейн, 2021: 161].

Наконец, еще одна важная проблема, тесно связанная с информатизацией, касается доверия к людям и социальным группам в условиях ценностного многообразия.

Постоянное пользование интернетом способствует выработке особого типа мышления. Для того, чтобы получать релевантную информацию, человеку следует отсеять все, что не соответствует фактам. Фактически современный пользователь научился не обращать внимания на нерелевантную информацию, воспринимая ее как информационный шум. Такой отсев часто происходит автоматически, бессознательно.

Информация помимо фактического содержания содержит определенные ценности и нормы. Если факты можно проверить, то ценности и нормы верификации не подлежат. Поэтому человек выбирает из имеющихся ценностных систем ту, которая ближе и понятнее ему (в силу опыта, воспитания и т.д.). Соответственно, все, что противоречит ей (то есть, что не нравится человеку, вызывает «ценностный диссонанс»), также начинает автоматически отсеиваться как часть информационного шума.

Развитие искусственного интеллекта позволяет оптимизировать этот процесс фильтрации на программном уровне. Алгоритмы персонализации результатов поиска образуют дополнительный «пузырь фильтров», из-за чего каждый пользователь получает преимущественно ту информацию, которая подтверждает важность ценностей и норм, которые ему импонируют, и опровергает те взгляды, которые ему не нравятся.

Схожую роль играют разнообразные «инфлюенсеры», не всегда обладающие достаточными квалификацией и знаниями в сфере своей

деятельности, но умеющие завоевывать доверие аудитории, играя на ее интересах и слабостях. Как правило, они продвигают те идеи, которые самим пользователям кажутся наиболее привлекательными. Аудитория готова больше платить за те лекции, мастер-классы, тренинги, где доказывается, что единственно правильный путь к личностной самореализации и успеху – это делать то, что нравится человеку, невзирая на реакцию и давление других.

Все эти механизмы фильтрации и убеждения в совокупности создают у пользователя впечатление, что единственно правильными являются те нормы и ценности, которые соответствуют его желаниям и потребностям. Любые контраргументы и попытки заставить его делать то, чего он не желает (даже если они направлены на достижение общего блага), начинают восприниматься им как ущемление свободы и насилие над личностью. Если на человека оказывают давление, требуя ограничить свои желания, он ощущает себя жертвой несправедливой системы господства и подчинения.

В условиях сетевой коммуникации информация становится более открытой и публичной, скорость ее распространения многократно возрастает. При этом сетевой характер обсуждения позволяет поднять ту или иную проблему без участия посредников (государства, крупных СМИ). «Из социальных сетей теперь индивид черпает новости, они не производятся специальными компаниями и продвигаются не пропагандистами телевизионных каналов, а обычными людьми – соответственно, они заслуживают большего доверия в силу личного характера и отсутствия возможной ангажированности» [Веселов, 2020: 135].

В социальных сетях представители массы получают трибуну, где могут заявить о своих желаниях и обвинить государство и общество в том, что они ставят барьеры на пути реализации этих желаний, подавляют индивидуальность, нарушают права и ущемляют свободы личности. Статус жертвы позволяет найти единомышленников, готовых не только обсуждать проблемы, но и искать пути их решения. Идентификация с жертвой позволяет выйти за пределы эгоизма и выстроить новую систему моральных ценностей, в центре которой находится благо не общества, а страдающего другого, чьи желания игнорируются и подавляются. Совокупность норм, направленных против дискриминации другого, ограничения свободы выбора, всей сложившейся в обществе системы доминирования и подчинения и составляет в наиболее крайних своих проявлениях феномен «новой этики».

Новые цифровые технологии дали сторонникам индивидуалистической морали (которые ранее уже в силу своего индивидуализма не могли консолидировать свои действия) возможность объединиться в защиту тех, кто страдает от насилия представителей привилегированных групп. В совокупности таких жертв – реальных или потенциальных – оказывается больше, чем гонителей (хрестоматийных «белых гетеросексуальных мужчин»). Это позволяет борцам против дискриминации при помощи бойкотов, остракизма и других форм культуры «отмены» влиять на тех, кто высказывает мысли, оскорбляющие достоинство «маленького человека», ограничивающие его

личный выбор [Жадунова, 2023: 81-93]. Виртуальная и реальная жизнь сегодня так тесно переплетены, что результатом «отмены» в интернет-пространстве, как правило, являются негативные последствия и в реальной жизни (разрушение семьи, потеря работы, уголовное преследование и т.д.) [Коваль, 2020: 335-344].

Защита желаний другого от насилия со стороны общества, основанного на господстве и подчинении, в «новой этике» предполагает достижение не столько «общего блага» (хотя, в определенной степени, и может рассматриваться как вклад в него), сколько конкретного блага того или иного человека, которое может отличаться от блага другого. Поскольку возможны конфликты между разными пониманиями блага, людям важно не нарушать личные границы других и постоянно договариваться о действиях. «Культуре насилия» здесь противопоставляется «культура согласия», где один индивид должен дать другому явно выраженное согласие на любое вмешательство в его жизнь. В этом смысле «новая этика» является этикой индивидуалистической и противопоставляется системе морали, декларирующей приоритет коллективных ценностей над индивидуальными желаниями. Эта система глубоко укоренена в традиционной культуре и регулярно воспроизводится в кризисных ситуациях, когда необходима мобилизация людей перед лицом экзистенциальных опасностей. Терроризм, вооруженные конфликты, пандемия и другие чрезвычайные ситуации последних десятилетий потребовали существенных ограничений личных прав, значительно укрепив тем самым позиции коллективной морали.

«Пузыри фильтров» эффективно действуют в отношении сторонников не только «новой», но и «старой» этики, убеждая последних в безальтернативности традиционных ценностей и опасности любых отступлений от них. Кроме того, эта позиция находит свою поддержку у государства и ряда влиятельных институтов, которые все активнее используют возможности не только традиционных СМИ, но и социальных сетей для критики индивидуалистической морали и поддержки коллективных ценностей.

В итоге одной из тенденций в ходе цифровизации социальной жизни становится возрастающая ценностная поляризация общества, выраженная, в том числе, в жестком противопоставлении «старой» морали общественного блага «новой» этике освобождения от дискриминации. В этих условиях в обеих группах формируется сверхдоверие по отношению к тем, кто разделяет схожие ценности, и свехнедоверие по отношению к «чужим».

Заключение

Развитие информационного общества, особенно возможностей ИИ, подрывает доверие людей не только к новым технологиям, но и к другим людям и социальным группам, способствуя радикализации позиций и поляризации моральных ценностей. Общественное взаимодействие приобретает конфликтный и агрессивный характер, а общий уровень доверия людей друг к другу падает. В этих условиях важен поиск таких вариантов использования новых информационных технологий, которые способствовали бы

формированию и укреплению доверия. Для этого необходима разработка новых норм регулирования сетевых взаимодействий, предполагающая введение моральных ограничений и общественного контроля за использованием новых цифровых систем, прежде всего основанных на ИИ. Характер этих ограничений во многом зависит от того, какую именно из конкурирующих ценностных позиций должен занять субъект нормотворчества.

Как представляется, ни фиксированность на ничем не ограниченных желаниях индивида, ни призывы к подавлению желаний ради общего блага не могут стать прочной ценностной основой для новой морали информационного общества. Они направлены на отстаивание интересов целого в ущерб личному или личного в ущерб целому, и в этом смысле являются не столько моральными, сколько идеологическими системами взглядов, для которых оппозиция «друг-враг» определяет содержательную наполненность понятий «добро» и «зло», а не наоборот. Специфика морали состоит в том, что она подразумевает движение личности и общества навстречу друг другу и готовность согласовывать интересы всех заинтересованных лиц. В отсутствие такого движения отношения доверия между личностью и обществом выстроить невозможно.

Список литературы (References)

1. Веселов Ю.В. Доверие в цифровом обществе // Вестник Санкт-Петербургского университета. Социология. 2020. Т. 13. Вып. 2. С. 129-143. <https://doi.org/10.21638/spbu12.2020.202>.

Veselov Yu.V. (2020) *Doverie v tsifrovom obshchestve* (Trust in a digital society). *Vestnik of Saint Petersburg University. Sociology*. Vol. 13. Issue 2. P. 129-143. <https://doi.org/10.21638/spbu12.2020.202>. (In Russ.)

2. Гидденс Э. Последствия современности. М.: Практикс, 2011.

Giddens E. (2011) *The Consequences of Modernity*. Moscow: Praksis. (In Russ.)

3. Жадунова Н.В., Гришнева А.А. «Новая этика»: нормальное и аномальное // Социальные нормы и практики. 2023. № 2(8). С. 81-93. <https://doi.org/10.24412/2713-1033-2023-2-81-93>.

Zhadunova N.V., Grishneva A.A. (2023) “Novaya etika”: normal'noe i anormal'noe (“New ethics”: normal and abnormal). *Social norms and practices*. No. 2(8). P. 81-93. <https://doi.org/10.24412/2713-1033-2023-2-81-93>. (In Russ.)

4. Коваль Е.А., Сычев А.А. «Новая этика»: нормотворческие перспективы // Мир человека: нормативное измерение – 7.0. Проблема обоснования норм в различных перспективах: от реализма до конструктивизма и трансцендентализма: Сборник трудов международной научной конференции, Саратов, 07-09 июня 2021 года / отв. ред. И.Д. Невважай. Саратов: Саратовская государственная юридическая академия, 2021. С. 335-344.

Koval E.A., Sychev A.A. (2021) “Novaya etika”: normotvorcheskie perspektivy (“New ethics”: norm-creating visions). *Human world: Normative dimension – 7.0. Problem of justification of norms in various prospects: from realism to constructivism and a transscendentalism. Proceedings of the international academic conference,*

Saratov, June 07-09, 2021; Ed. by I.D. Nevvazhay. Saratov: Saratov State Academy of Law. P. 335-344. (In Russ.)

5. Перов В.Ю. Иллюзия «сверхдоверия» в контексте профессиональных этик в цифровом обществе // XIII международная конференция «Теоретическая и прикладная этика: Традиции и перспективы – 2021. Этика как наука и профессия». Санкт-Петербургский государственный университет, 18-20 ноября 2021 г. Материалы конференции / Отв. ред. В.Ю. Перов. СПб.: ООО «Сборка», 2021. С. 214-215.

Perov V.Yu. (2021) Illyuziya “sverkhdoveriya” v kontekste professional'nykh etik v tsifrovom obshchestve (The illusion of “supertrust” in the context of professional ethics in a digital society). *XIII International Conference Theoretical and Applied Ethics: Traditions and Prospects – 2021. Ethics as research and profession. Saint Petersburg State University, November 18-20, 2021. Conference papers; Ed. by V.Yu. Perov. St. Petersburg: Sborka Publ. P. 214-215. (In Russ.)*

6. Рисухин В.Н. Проблемы, вызванные сложностью систем автоматического управления полетом, и возможные пути их преодоления // Вестник Международной академии проблем человека в авиации и космонавтике. 2006. № 2 (21). С. 25-28.

Risuhin V.N. (2006) Problemy, vyzvannye slozhnost'yu sistem avtomaticheskogo upravleniya poletom, i vozmozhnye puti ikh preodoleniya (Problems caused by the complexity of automatic flight control systems and possible ways to overcome them). *Bulletin of International Academy for Human Problems in Aviation and Cosmonautics. No. 2 (21). P. 25-28. (In Russ.)*

7. Селигмен А. Проблема доверия. М.: Идея-Пресс, 2002.

Seligmen A. (2002) The problem of Trust. Moscow: Ideya-Press. (In Russ.)

8. Сычев А.А., Коваль Е.А., Ушкин С.Г. Социология за пределами реальности: постгуманизм и биополитика в сериале «Мир Дикого Запада» // Мониторинг общественного мнения: экономические и социальные перемены. 2020. № 6. С. 141-172. <https://doi.org/10.14515/monitoring.2020.6.1747>.

Sychev A.A., Koval E.A., Ushkin S.G. (2020) Sotsiologiya za predelami real'nosti: postgumanizm i biopolitika v seriale “Mir Dikogo Zapada” (Sociology beyond reality: posthumanism and biopolitics in the “Westworld” television series). *Monitoring of Public Opinion: Economic and Social Changes. No. 6 (160). P. 141-172. <https://doi.org/10.14515/monitoring.2020.6.1747>. (In Russ.)*

9. Ушкин С.Г., Коваль Е.А., Яськин А.Н. Жить с Алисой: как голосовые помощники трансформируют практики коммуникации? // Журнал исследований социальной политики. 2022. № 3. С. 361-376. DOI: 10.17323/727-0634-2022-20-3-361-376.

Ushkin S.G., Koval E.A., Yaskin A.N. (2022) Zhit' s Alisoy: kak golosovye pomoshchniki transformiruyut praktiki kommunikatsii? (Living with Alice: how do voice assistants transform communication practices?). *The Journal of Social Policy Studies. No 3. P. 361-376. DOI: 10.17323/727-0634-2022-20-3-361-376. (In Russ.)*

10. Ушкин С.Г. Пуховик папы Римского: как нейросети сужают доверие и укрепляют дискурс постправды? // Социальные нормы и практики. 2023. № 2(8). С. 71-80. <https://doi.org/10.24412/2713-1033-2023-2-71-80>.

Ushkin S.G. (2023) Pukhovik papu Rimskogo: kak neyroseti suzhayut doverie i ukreplyayut diskurs postpravdy? (The puffer jacket of the Pope Francis: how do neural networks narrow confidence and reinforce post-truth discourse?). *Social norms and practices*. No. 2 (8). P. 71-80. <https://doi.org/10.24412/2713-1033-2023-2-71-80.2>. (In Russ.)

11. Фукуяма Ф. Доверие: социальные добродетели и путь к процветанию. М.: АСТ, 2008.

Fukuyama F. (2008) Trust: The Social Virtues and the Creation of Prosperity. Moscow: AST. (In Russ.)

12. Штомпка П. Доверие – основа общества. М.: Логос, 2012.

Sztompka P. (2012) Trust is the Basis of Society. Moscow: Logos. (In Russ.)

13. Эпштейн М. Новая этика или старая идеология? // Знамя. 2021. № 8. С. 159-162.

Epshtejn M. (2021) Novaya etika ili staraya ideologiya? (New ethics or old ideology?). *Znamya*. No 8. P. 159-162. (In Russ.)

14. Awad E., Dsouza S., Kim R., Schulz J., Henrich J., Shariff A., Bonnefon J.-F., Rahwan I. (2018) The Moral Machine experiment. *Nature*. No. 563 (7729). P. 59-64.

15. Nandhakumar J., Panteli N., Powell Ph., Vidgen R. Trust in the Digital Era (2004). *Social and Economic Transformation in the Digital Era*. Hershey PA: Idea Group Inc. P. 77-94. <https://doi.org/10.4018/978-1-59140-158-2>.

Сведения об авторе

Сычев Андрей Анатольевич – доктор философских наук, профессор, профессор кафедры философии Национального исследовательского Мордовского государственного университета. Занимается изучением социокультурных проблем современности. Автор более 200 работ.

E-mail: sychevaa@mail.ru

<https://orcid.org/0000-0003-3757-4457>

About the author

Sychev Andrey Anatolievich – Doctor of Philosophical Sciences, professor, professor of the philosophy department of National Research Mordovia State University. Engaged in studying contemporary sociocultural problems. The author of over 200 works.

E-mail: sychevaa@mail.ru

<https://orcid.org/0000-0003-3757-4457>

Поступила 28.09.2023; одобрена после рецензирования 03.11.2023; принята к публикации 01.12.2023.