

ИСКУССТВЕННЫЕ ИНТЕЛЛЕКТУАЛЬНЫЕ СИСТЕМЫ И ПРОБЛЕМА «ЕСТЕСТВЕННОГО» ДОВЕРИЯ

Е. Дегтева, О. Куксова

Аннотация. Развитие технологий ИИ обострило проблему гуманитарного, создав вызов на всех уровнях общественных отношений. Этические вопросы и, в частности, проблема доверия стали актуальными для сферы высоких технологий, учитывая тот факт, что ИИ выполняет все более значимые управленческие функции, которые ранее безоговорочно относились к человеку. Этот вопрос напрямую связан с системами искусственного интеллекта, которые уже воплотились в конкретных масштабных проектах. В данном исследовании авторы анализируют концепцию доверия через призму технологического развития. Для этого в исследовании представлен обзор исторических и современных интерпретаций концепции доверия и доказывається, что эта концепция актуальна и необходима для контроля рисков, возникающих при интеграции продуктов искусственного интеллекта в социальную жизнь. Авторы показывают, что требуется переосмысление понятий этики и морали в новом контексте. Это необходимое требование для создания доверенного ИИ и для достижения доверия при взаимодействии человека с технологическими продуктами. Авторы приходят к выводу о необходимости построения междисциплинарного диалога для интеграции теории и практики из многочисленных областей. Для этого необходимо создать общую базу знаний и платформу для общения между всеми заинтересованными сторонами, а также важно сформировать благоприятные условия для устойчивого и конструктивного взаимодействия. Поэтому доверие является актуальной концепцией, которую необходимо строить в многомерной системе координат, ориентированной на различные заинтересованные стороны, особенно с учетом роста взаимодействия между человеком и технологиями, другими словами, на всех уровнях и во всех масштабах.

Ключевые слова: ИИ, искусственный интеллект, система искусственного интеллекта, этика, мораль, доверие

Для цитирования: Дегтева Е., Куксова О. (2023). Искусственные интеллектуальные системы и проблема «естественного» доверия. – *Исследования в цифровой экономике*. №1, С. 109–136. DOI: [10.24833/14511791-2023-1-109-136](https://doi.org/10.24833/14511791-2023-1-109-136)

Информация об авторах:

Дегтева Е. – магистр экономики, Заместитель директора Департамента развития социальной сферы и сектора НКО, Министерство экономического развития Российской Федерации. degteva@list.ru
ORCID: 0000-0002-4137-6317

Куксова О. – магистр экономики, Заместитель начальника управления, Аналитический центр топливно-энергетического комплекса ФГБУ “РЭА” Минэнерго России.
129110, г. Москва, ул. Гиляровского, д.39, стр.1.
kuksovaki@gmail.com
ORCID: 0000-0001-8169-1517

Введение

В последнее время возникла явная необходимость «вписать» гуманитарные ценности в продукты цифровых технологий, в частности, в рамках разнообразных проектов на основе ИИ, которые используются в государственном управлении, бизнесе, образовании и других сферах. Многочисленные теоретики, эксперты, практики, деятели культуры и искусства ведут горячую и оживленную дискуссию о том, как вовремя выявлять возможные риски и угрозы, кто и как должен разрабатывать и применять инструменты контроля рисков, как внедрять их в систему управления и общественного контроля с целью подготовки и адаптации общества к еще большему внедрению технологий ИИ. Сегодня именно этические вопросы широко обсуждаются и активно разрабатываются в контексте развития и применения ИИ в политической, экономической и социальной сферах.

Данные этические вопросы неразрывно связаны с актуальной проблемой доверия, которая сегодня требует внимания на всех уровнях: в международных отношениях, в экономике, в социальной сфере. Феномен доверия стал актуальным в сфере высоких технологий. Более того, дискуссия о взаимодействии человека с технологиями обостряется в условиях, когда ИИ начинает выполнять все больше управленческих функций, которые раньше мог делать только человек [7]. В данном случае мы подразумеваем не только «тактические» вопросы, но говорим и о технологиях, которые используются для управления на высоком уровне. Среди таких технологий можно выделить искусственные общества или цифровые двойники общества. Перед человечеством стоит задача обучить нейронные сети понимать» и «учитывать» этические ценности, либо четко определить, что такой результат невозможен. Этот вызов накладывает огромную ответственность не только на разработчиков, но и на всех участников процесса создания, внедрения и использования продуктов ИИ.

В то же время заявленная проблема уходит корнями в далекое прошлое - и вопрос доверия как такового, и проблема доверия к технологиям имеют долгую историю и сложное многогранное содержание.

Учитывая, что проекты искусственных обществ уже существуют, а их решения применяются на практике для проверки управленческих решений, очевидно, что дальнейшее развитие и использование требует большого внимания, поскольку такая деятельность влияет как на общество, так и на отдельного человека. Мы считаем, что доверие является ключевым фактором и требованием для

проверки приемлемости использования проектов ИИ. Это особенно актуально для значимых проектов, которые направлены на решение политических и социальных вопросов с использованием ИИ и, в частности, систем искусственного интеллекта, и в рамках которых жизненно важно обеспечить доверие между всеми заинтересованными участниками [8].

Обзор литературы

Взаимодействие технологических продуктов с антропосферой привлекало внимание мыслителей со времен античности. Эпохи Возрождения и Нового времени не стали исключением. В начале XX века ведущие мыслители были озабочены гуманитарными аспектами технического прогресса и взаимодействием человека и техники [15]. Мы не будем делать философско-исторический очерк, но сфокусируем наше внимание на том, что разнообразные вопросы, обусловленные технологическим развитием, стали объектом исследования философии достаточно давно. Даже как концепция понятие искусственного интеллекта еще не существовало. Когда концепция ИИ была сформулирована, философия приняла ее для рефлексии со всей серьезностью, потому что безусловно этот концепт приносит очень многое в культуру, поскольку в его сути соединяются ранее противоположные категории - естественного и искусственного, причем в практически сакральном (и не имеющим общепринятого определения) - в интеллекте. Вспомним, что американский ученый Д. Маккарти ввел в науку данный термин еще в 1955 году, и немногим позже его коллега М. Мински дал первое определение искусственному интеллекту в качестве науки, нацеленной на решение интеллектуальных задач (человека) с использованием машины [10].

Проблема ИИ стала благодатной почвой, на которой исследователи начали проверять возможности конвергенции естественного и искусственного, биологического и технологического - то, что сегодня представляется наиболее перспективными направлениями для развития науки и практики.

Уже в момент зарождения самого понятия ИИ были сформулированы по сути противоречивые подходы к его определению. В 1950-х и 1960-х годах в области ИИ выделились две точки зрения. Один подход, предложенный Д. Маккарти, использовал чистую логику и общую математическую основу. Маккарти считал, что исследования в области ИИ будут плодотворными, когда будет определено само понятие интеллекта (которое до сих пор имеет ряд трактовок). По его мнению, определение должно образовывать систему координат, определяемую эпистемологией и эвристикой. То есть эпистемология — это то, как приобретаются новые знания с точки зрения их полезности для решения проблем, а эвристика - как эти проблемы решаются: какими процедурами и с помощью каких средств [15].

М. Мински предложил другой подход, который впоследствии стал известен как коннекционизм (биологически инспирированный подход). Ученый призывал изучить, как работают человеческий мозг и психика, и на этой основе реали-

зовать знания в компьютерной модели. Этот подход дает основу для трактовки ИИ как чего-то субъективного, обладающего индивидуальностью, сознанием и даже эмоциями. [4].

В целом, сегодня можно говорить о существовании 3 основных подходов к развитию ИИ:

Искусственный интеллект (ИИ), или «Символический искусственный интеллект», или «Классический искусственный интеллект» - это подход «сверху вниз», направленный на воспроизведение когнитивных способностей без погружения на уровень отдельных нейронов.

Обобщенный искусственный интеллект, или сильный искусственный интеллект (AGI), в настоящее время является предполагаемой, гипотетической формой ИИ, согласно которому машина получит равный человеческому интеллект, и, соответственно, такой ИИ будет обладать самосознанием и свободой выбора

Биологически вдохновленный искусственный интеллект, или биологически вдохновленные когнитивные архитектуры, — это проекты, которые, основываясь на достижениях нейронауки, направлены на создание искусственных систем, воспроизводящих поведение и мышление биологических существ [11]. В частности, развитие теории искусственных нейронных сетей коррелирует и взаимодействует с исследованиями функционирования человеческого мозга [13, 24].

Так, первый подход развивает подход Маккарти, а второе и третье направления основаны на работах Минского. Мы видим, что на сегодняшний день нет единого подхода и четкого понимания направлений развития ИИ и, соответственно, нет единого взгляда на понимание этого вопроса [27]. Однако попытка осмысления этой общей актуальной темы в рамках философии техники началась в начале XX века. Один из крупнейших мыслителей прошлого века М. Хайдеггер указал на угрозы, исходящие от такого этапа в развитии цивилизации, на котором техника стала играть доминирующую роль. Мыслитель отмечал, что техника не является нейтральной по отношению к человеку, а человек теряет свою субъективность и перестает определять развитие цивилизации. Техника сама начинает преобразовывать природный мир, а непонимание смысла техники и ее возможностей чревато ошибкой, за которую Человек заплатит своим порабощением [15, 16].

В наше время горизонт интерпретаций и прогнозов очень широк и разнонаправлен. Следует отметить, что фундаментальная мысль стала объектом не только профессиональных философов, но и талантливых изобретателей, предпринимателей и бизнесменов. В частности, рационалист Э. Юдковский верит в возможность создания сознательного ИИ, но согласно его позиции сначала необходимо раскрыть тайну интеллекта [23].

Технооптимист Р. Курцвейл считает, что ИИ обязательно будет создан. При условии, что сам человек сможет сохранить ценности и культуру, ИИ поможет сделать мир лучше и обеспечит процветание цивилизации [14].

Трансгуманист Н. Бостром предсказывает, что создание ИИ приведет к следующему этапу в эволюции разумной жизни - появлению Сверхразума, который получит возможность контролировать человеческое существование [5]. А.В. Абрамова исследует этические аспекты развития ИИ, изучая и систематизируя мировые тенденции в развитии этических инструментов в области ИИ [1].

А.Ю. Алексеев руководит проектом «Искусственная личность», в рамках которого исследует перспективы компьютерного моделирования культурных, политических, социальных, экономических, нравственных и других аспектов общественной жизни и человека. Философ убежден в том, что исследования искусственной жизни являются весьма актуальной сферой для науки, и искусственная жизнь имеет значение как ключевой концепт для осмысления и дальнейшей реализации нано- и биотехнологий с далеко идущей целью – воссоздания биологической жизни с применением цифровых технологий. Как считает мыслитель, искусственная жизнь лежит в основе искусственных обществ. Философ много лет изучает искусственные общества, поскольку считает этот феномен является базовым для развития социальных и информационных технологий, которые задействованы в управлении общественной жизнью. Данное направление исследований привело к созданию концепта «искусственной личности», вокруг которой сейчас ведется оживленная дискуссия, и которой ряд ученых приписывает не возможность обладать самосознанием, но также иметь свободу выбора и способность к моральным суждениям и рефлексии своих мыслей и действий [3].

Изучением особенностей человеческого интеллекта и ИИ занимается И.М. Дзялошинский, который стремится доказать принципиальную разницу между человеческим и искусственным интеллектом [10]. Необходимо подчеркнуть, что в последнее время данной проблематикой стали заниматься не только ученые, но и публицисты, журналисты и писатели. Многочисленные деятели культуры размышляют о перспективах развития ИИ, о рисках и возможностях, которые открывают продукты технологий, основанных на ИИ. В частности, можно упомянуть книги «Искусственный ты: Машинный интеллект и будущее разума», написанную С. Шнайдер, и «Наше последнее изобретение: Искусственный интеллект и конец человеческой эры», автором которой является Дж. Баррат [20, 33].

В последнее время появился ряд публикаций, посвященных изучению потенциальных опасностей злоупотребления нейронными сетями владельцами технологий и противодействия им, полагаясь на доверенный искусственный интеллект. [2].

Именно разработка доверенного искусственного интеллекта представляет собой наиболее перспективную концепцию создания и понимания идеи этического ИИ на сегодняшний день. Однако само понятие «доверенный» уже вышло за рамки теории и начало использоваться в качестве юридического термина в нормативно-правовом регулировании. В частности, понятие доверенного ИИ определено в Этическом руководстве для доверенного ИИ Группы экспертов

высокого уровня по искусственному интеллекту Европейской комиссии (Ethics guidelines for trustworthy AI, 2019) [30]. Согласно этому документу, заслуживающий доверия ИИ отвечает следующим требованиям, которые на первый взгляд относятся скорее к деятельности человека, чем к технологии:

- Этичный (существующий в соответствии с этическими ценностями и этическим кодексом);
- Законный (не противоречащий закону);
- Достоверный (надежный и устойчивый).

В этом контексте необходимо задуматься над пониманием такого понятия, которое на первый взгляд относится в основном к социальной сфере, как доверие. Это понятие в определенной степени объединяет представленную тему, поэтому далее будут предложены некоторые актуальные подходы к определению данного понятия. Именно доверие выступает как объединяющее понятие и как практический инструмент, который позволит ему стать объединяющим звеном между человеческими ценностями и технологическим развитием [29].

Это понятие вошло в теорию и социальную практику с ранних этапов цивилизации, практически одновременно его ввели Конфуций в Древнем Китае и Перикл в Древней Греции, описывая отношения, основанные на доверии, как необходимое условие процветания государства и общества [28].

Впоследствии, в разные исторические периоды, в зависимости от идеологии, стратегии развития и личностных качеств лидеров и государственных идеологов, подходы к определению и оценке значения этого понятия неоднократно менялись самым радикальным образом.

Обращаясь к реалиям XX века, отметим, что в период после Второй мировой войны проблемы доверия и противоположного состояния - недоверия в контексте гонки вооружений и политики ядерного сдерживания были одними из ключевых. В рамках этой политики в США возникла школа политического реализма, разработанная Г. Киссинджером, Т. Шеллингом и рядом других ученых и стратегов. Данный подход предполагает, что уровень доверия политических акторов (в данном случае игроков на международном поле) низок, а стабильность (предсказуемость) действий партнеров определяется балансом сил и системой сдержек и противовесов. Только такой подход может обеспечить предсказуемость в политической игре [34].

Неолиберальный подход подходит к вопросу сложной взаимозависимости как к возможности с помощью многочисленных каналов и инструментов построить доверие между международными партнерами и тем самым обеспечить высокий уровень доверия во внешнеполитическом дискурсе [32].

С. Хантингтон отмечает, что особая ценность доверия возникает в периоды трансформации и кризисов, и подчеркивает важность институционального доверия, поскольку именно стабильные институты позволяют сохранить общество и цивилизацию в трудные времена [35].

Постмодернистский взгляд по сути отвергает доверие, а Ж.-Ф. Лиотар выделяет в качестве тенденции глобальное недоверие к «метанарративам», которые должны оправдывать реальность. З. Бауман указывает на эфемерность, нестабильность и ненадежность постмодернизма и утверждает, что центр тяжести смещается от государственных и международных институтов к глобальному рынку с его потребительскими ценностями и заменой рационального знания поверхностным воздействием СМИ [26].

Последняя мысль приводит нас к проблеме технологического развития, поскольку оно позволило глобализировать экономику и культуру, развить СМИ и массовую культуру, а затем социальный и мобильный веб, и все это благодаря технологиям и продуктам искусственного интеллекта. Таким образом, широта обсуждения, масштаб проблемы и горизонт мнений указывают на то, что данный вопрос является сложной, актуальной и перспективной областью исследования.

Обсуждение

Рост значения технологий в общественной жизни, и внедрение их продуктов в важные процессы, управляющие текущей деятельностью и развитием цивилизации, оказывают серьезное влияние на изменение содержания и нп появлением нового набора социальных ценностей, а также непосредственно влияют на формулировки новых стратегий.

В продолжение озвученной темы – перспектив развития искусственных интеллектуальных систем, постараемся определить данное понятие. Мы подойдем к нему с точки зрения функционального подхода: ИИС обладает набором свойств и ресурсов, которые позволяют ее находиться во взаимодействии с внешним миром, причем взаимодействие представляет собой активную реакцию на происходящие события. Систем тем более автономна, чем большей степенью независимости от команд управления она обладает. В этом смысле ИИС является в большей или меньшей степени самодостаточной системой, обладающей способностью находить решения тем или иным точечным задачам [8].

В частности, нейронная сеть, созданная компанией Salesforce, тестирует и разрабатывает идеальную налоговую систему в симулированной среде [36]. Так называемый AI Economist основан на обучении с подкреплением, которое предполагает применение вознаграждений и наказаний к машинным алгоритмам для максимизации желаемых результатов. По такому же принципу, например, были созданы алгоритмы Google DeepMind AlphaGo и AlphaZero. Эксперимент был направлен на то, чтобы найти решение, которое позволило бы более эффективно и справедливо управлять системой налогообложения. Программа нацелена на создание огромного количества экосистем с работниками-теоретиками, которые торгуют валютой и строят дома. Уровень заработной платы и набор навыков варьируются, а ИИ определяет оптимальные налоговые ставки. Ученые обращают

внимание на то, что такой подход учитывает возможность учитывать иррациональное поведение у субъектов. Классические экономические теории нередко игнорируют эти аспекты при выстраивании моделей [11].

Актуальность таких проектов возрастает в связи с тем, что человечество сегодня развивается ускоренными темпами, и беспрецедентная скорость изменений регулярно вызывает глобальные кризисы. Именно поэтому стало популярным понятие «черных лебедей», регулярно возникающих крупномасштабных рисков и угроз, которые невозможно предвидеть, обнаружить или идентифицировать заранее, поскольку они исходят из областей, находящихся за пределами знаний, и во многом (конечно, не полностью) обусловлены стремительным развитием технологий, достижения которых интегрируются в многообразные сегменты общественной жизни, расширяя техносферу и стирая границы между гуманитарной и технологической областями [12].

В этом контексте классические инструменты экономической теории, стратегического менеджмента и даже антикризисного управления теряют убедительность и перестают приносить результаты. Для изучения мира, в котором техносфера заняла важное место, необходимы инструменты, включающие также возможности цифровых технологий, в частности ИИ. В этом смысле искусственные интеллектуальные системы очень перспективны, поскольку в данной концепции используется агент-ориентированное моделирование. Этот метод исследует поведение децентрализованных агентов и то, как это поведение определяет поведение всей системы. Использование агент-ориентированного моделирования в искусственных интеллектуальных системах открывает большие возможности не только для прогнозирования развития общества в обычных условиях, но и для предсказания возможных кризисов и минимизации ущерба от таких событий. Конечно, такой подход является сложной задачей, поскольку он должен учитывать большое количество сложных факторов и основываться на результатах многогранных исследований реальных сложных систем [21].

В этом контексте доверие является важным и необходимым условием, поскольку мы имеем дело с радикальной неопределенностью. В общественной жизни, на индивидуальном уровне, да и во всем мире, неопределенность будущего выходит на первый план, как на практическом, стратегическом, так и на философском, буквально метафизическом уровне. Однако в этих условиях перспективные проекты на основе ИИ могут стать решением, поскольку классические инструменты сегодня не могут показать эффективных результатов. Темпы изменений, появление новых условий, ускорение развития требуют учета всех факторов, а классические теории и методологии не способны учесть важные, но незамеченные и не просчитанные элементы, такие как иррациональность и другие особенности человеческой природы, и все вытекающие из этого ограничения и возможности.

В условиях радикальной неопределенности доверие является требованием, обеспечивающим смысл и основу для дальнейшего развития. Более того, в обществе, где развитие определяется явлениями, которые не были изначально задуманы

маны, а возникли сами по себе, в частности, когда совокупный эффект действия совокупности индивидов отличается от эффекта действия индивидов и неэргодическими процессами (для неэргодических процессов вероятность, наблюдаемая в прошлом, не применима к будущим процессам), неопределенность становится основной характеристикой. В то же время общество увеличивает неопределенность за счет собственной несогласованности, креативности, и именно в этих условиях требуются искусственные интеллектуальные системы для функционирования и, более того, для обеспечения уверенности в своей деятельности и в ее результатах [21]. На сегодняшний день мы уже нередко способны моделировать потенциальные последствия работы подобных продуктов технологий. При этом последствия можно охарактеризовать как с положительной, так и отрицательной сторон. В первую очередь, такие системы направлены на обеспечение безопасности и благополучия, с другой стороны, применение решений таких систем в реальной жизни может затронуть такие естественные ценности, как свобода, частная жизнь и личная информация реальных людей. И самое главное, необходимо четко определить и установить, кто будет играть доминирующую роль в новых условиях и какой баланс будет в управлении социальной жизнью между государством, технологическими корпорациями, обществом и личностью.

Кроме того, наряду положительными перспективами, возникающими в свете развития и распространения интеллектуальных систем, следует обратить внимание и на потенциальные риски, обращенные в общественное измерение. Наиболее очевидными представляются проблемы в области профилирования и агрегирования персональных данных. Социальные сети, развлекательные сервисы, являющиеся интеллектуальными системами, могут накапливать большие объемы персональной информации о жизнедеятельности каждого пользователя, его личных предпочтениях, взглядах, состоянии здоровья, семейных отношениях и другой информации. В случаях, когда система научилась распознавать пользователя по его поведенческим факторам, идентификация может происходить даже тогда, когда пользователь появляется на партнерских сервисах под псевдонимом или анонимно. Такое профилирование данных может нарушать права и свободы человека [22].

Агрегация пользовательских данных для улучшения сотрудничества с партнерскими сервисами сама по себе не опасна, но если накапливается достаточное количество личной информации из различных источников, система может стать привлекательной для злоумышленников с целью создания брешы в данных. Поэтому агрегирующие сервисы обязаны своевременно повышать безопасность всех задействованных систем. Однако в настоящее время эта проблема далека от глобального решения. В настоящее время этот вопрос рассматривается на самом высоком уровне. По мнению Экономического и социального совета ООН, необходимо уточнить, кому принадлежат данные, обеспечить ответственное использование данных, защиту частной жизни, подотчетность и справедливость в

социальных сетях, платформах с широким участием и порталах онлайн-коммерции. Основная цель этих мероприятий - укрепление доверия и стабильности в цифровом пространстве, в частности в области ИИ [17].

Другим примером возможных негативных социальных последствий при использовании интеллектуальных систем является тема, связанная с большими данными. Предполагается, что большие данные должны быть обезличены, но возможность извлечения персональных данных из обезличенных потоков данных (деперсонализация) и получение индивидуальной информации из больших данных — это крайне непростая работа. В этом случае человек не защищен от несанкционированного использования личной информации и, конечно, доверие к системам разведки уже не гарантировано. Список социальных последствий можно было бы продолжить, поскольку представленные вызовы не имеют на данный момент общепринятых решений. В этой связи мы должны исследовать этическую проблематику, возникающую при работе подобных систем. То есть необходимо изучить, насколько поведение таких систем может быть согласовано с этическими идеями и правилами, зафиксированными в соответствующих кодексах [32].

Существует потребность в определении и согласовании целей, на выполнение которых направлено создание и дальнейшее использование интеллектуальных систем. При этом важно определить границы автономии и спектр вопросов, которые возможно решать с помощью таких систем. Отдельной темой, которая сейчас активно исследуется, является категоризация потенциальных рисков и инструменты их контроля. При этом необходимо учитывать, что на сегодняшний день интеллектуальные системы представляют собой существующие технологии, которые не только имеют потенциал развития, но уже сейчас могут оказывать влияние на общественное развитие. Поэтому вопрос создания и применения кодекса этики для искусственного интеллекта является вполне прагматичной темой – ведь в данном случае мы говорим о технологиях, чья работа направлена на управление, прогнозирование и выбор моделей человеческого поведения [23]. С учетом динамики общественного развития следует на постоянной основе осуществлять анализ происходящих явлений и тенденции, происходящих в обществе, и по результатам проведенной работы оперативно менять законодательство и фиксировать общественные нормы. Поскольку этот процесс сложный, для его осмысления и реализации следует интегрировать работу экспертов, представляющих разнообразные области теории и практики. Речь идет и о работе, затрагивающей разномасштабные активности – это и научная, и законодательная, и практическая работа, в которую должны быть вовлечены и все общество, и государство, и экспертное сообщество. Только при такой совместной работе возможно достичь практических и эффективных результатов.

Соглашаясь с необходимостью активной работы в области создания и применения положений Этического кодекса ИИ, мы должны обратить внимание, что этот инструмент сам по себе не сможет решить озвученные ранее пробле-

мы [25]. Не секрет, что в общественном пространстве в последние годы границы между концептами этики, морали и нравственности размылись, и в повседневном языке они используются как синонимы. Кроме этого эти концепты стали использоваться буквально во всех сферах в очень широком смысле. Чтобы исключить неопределенность, постараемся дать определение этим понятиям. Этика в широком смысле определяет, что есть «добро» и «зло» и противопоставляет их в рамках системы мировоззрения. Мораль, в свою очередь, является инструментом управления социальным посредством принципов и норм [9]. Нравственность же – практические действие и поведение человека в реальности, и последствия такого поведения, отражающееся впоследствии в полученном и осмысленном опыте индивида и общества. Если выразить это иначе, то общественная мораль представляет собой объективную оценку реального поведения в обществе. В связи с этим применение понятия «этики» в абстрактном смысле, как мы полагаем, не несет никакой пользы и даже опасно. В этом случае происходит лишь появление в культурном пространстве очередных симулякров, распространение которых чревато разрушением человеческой природы, разрушением социального и социальными потрясениями. Мы считаем, что современные социальные тенденции образуют единый нравственный узел - фундамент, на котором должны строиться ступени нравственного прогресса [9].

Именно по этой причине вокруг данных понятий существует множество спекуляций, цена которых в связи с использованием искусственных интеллектуальных систем может оказаться слишком высокой. Проблема заключается в том, что нет единого понятийного аппарата для базовых определений, нет должного уровня цифровой и гуманитарной грамотности, а среди заинтересованных участников часто присутствует предвзятость или конфликт интересов. Цель заключается в том, чтобы проговорить и согласовать качественно новые определения для ново созданных категорий, с учетом того, что проблематика является по сути многомерной. Успешное выполнение этого требования будет способствовать более объективному анализу с учетом сложности явлений. По нашему мнению, не стоит недооценивать необходимость в кратчайшие сроки сфокусироваться на изучении и поиске решений данным вопросам, поскольку инвестиции в создание прикладных продуктов ИИ в разнообразных секторах экономики растут на регулярной основе. Этот бум привел к тому, что уже сейчас мы можем говорить о существовании фактически автономных систем, поэтому речь идет не о теоретических угрозах, а о фактических рисках, которые уже сегодня чреват масштабными последствиями [37].

Вышеприведенные факты подтверждают важную роль исследований в области этического поведения интеллектуальных систем. Знание в этой сфере и изучение данной проблематики важно как с теоретической, так и с практической позиций. Интеллектуальные системы, как мы уже отметили, будут использоваться все чаще и чаще и будут влиять на общество и человека. В этих условиях крайне важно обеспечить прозрачность и гибкость при создании цифрового двойника

общества. Несомненно, важно обеспечить укрепление доверия к таким проектам. В научном и особенно общественном дискурсе можно встретить как фантастические, так и обоснованные опасения в связи с рисками безответственного использования технологий ИИ, вплоть до угрозы потери человеком автономии и свободы [6]. В то же время общепринятого и четкого плана формирования доверия в этой сфере не существует, однако исторический опыт внедрения инноваций показывает, что доверие обычно рождается, если технологии полезны и безопасны, а их регулирование прозрачно.

Говоря о практической стороне этического регулирования модели искусственного общества, необходимо решить проблему формализации этики в формате, доступном для технологий ИИ, а также реализовать принятие этических решений в самой автономной системе на основе ограничений ИИ в соответствии с моральными нормами или путем обучения ИИ распознавать этические конфликты и принимать свои индивидуальные решения [2]. Это требует повышения ответственности при разработке и использовании этих систем, а также максимально оперативного контроля ввиду серьезных последствий неправильного этического выбора.

Существующий опыт систематизации основных этических вопросов, связанных с созданием и развитием сложных проектов на основе технологий ИИ, позволяет выделить проблемы конфиденциальности и защиты данных, прозрачности и, возможно, как общий знаменатель, доверия к предиктивным данным.

Заключение

Проведенное исследование позволяет сделать вывод, что сегодня необходимо изучать технологическое и социальное развитие в комплексе, и такой глобальный взгляд позволит вовремя выявить и найти решение проблемам, вызванным развитием и распространением ИИ и таких продуктов, как системы искусственного интеллекта.

В исследовании авторов был проведен анализ концепции доверия через призму технологических реалий. Авторы представили обзор исторических и современных интерпретаций концепции доверия и показали, что эта концепция является весьма перспективной и даже необходимой в качестве инструмента контроля рисков, возникающих при создании цифровых технологических продуктов, что особенно очевидно в случае интеллектуальных систем.

В то же время гуманитарные ценности сегодня требуют определенного переосмысления и новых определений, которые будут соответствовать новой возникающей реальности. Авторы показывают, что использование понятий этики и морали без качественного пересмотра их смысла в соответствии с современными требованиями несостоятельно и является формальным интеллектуальным упражнением, которое в действительности не приведет к созданию доверенного ИИ и формированию доверия к взаимодействию человека с технологиями. Авторы утверждают, что данный вопрос является вызовом, поскольку современная

цивилизация на нынешнем этапе своего развития включает в себя широкий круг заинтересованных сторон, представляющих политические, экономические и социальные сферы, от глобального сообщества до отдельного человека, и очевидно, что для построения этого разнонаправленного диалога необходимо сформировать новый язык и новые определения терминов, определяющих требования и позволяющих описывать ценности. По мнению авторов, вклад в развитие данной проблематики заключается в том, что проведено значительное исследование интерпретаций соответствующих понятий, а представленные выводы могут быть использованы для следующих конкретных шагов.

Для всестороннего понимания и управления текущими процессами авторы предлагают следующие виды деятельности:

1. Необходимо построить междисциплинарный диалог для интеграции теории и практики из многочисленных областей, принимая во внимание глобальный уровень вопросов, которые мы исследуем.
2. Для этого необходимо создать общую базу знаний и платформу для общения между всеми заинтересованными сторонами.
3. В этих условиях станет возможным устойчивое и конструктивное взаимодействие, в ходе которого можно будет сформулировать и согласовать новые определения обсуждаемых понятий с целью их дальнейшей интеграции в практику, как социальную, так и научную.

Таким образом, авторы считают, что в современных условиях доверие - это категория, которую нужно и можно строить в нескольких измерениях - между заинтересованными сторонами, между человеком и технологией, то есть на всех уровнях и во всех масштабах.

Ссылки

1. Абрамова А.В. Этика в области искусственного интеллекта - от дискуссии к научному обоснованию и практическому применению: аналитический доклад / А.В. Абрамова, А.Г. Игнатьев, М.С. Панова; МГИМО МИД России, Центр искусственного интеллекта; XIII Съезд Российской ассоциации международных исследований (РАМИ) (Москва, 14-16 октября 2021 г.). - М.: МГИМО-Университет, 2021. - 24 с.
2. Авдошин С.М., Песоцкая Е.Ю. Доверенный искусственный интеллект как способ цифровой защиты // Бизнес-информатика. - 2022. - №2. - С. 62-73.
3. Алексеев А.Ю. Когнитивно-технологические проекты искусственной личности // Человек: образ и сущность. Гуманитарные аспекты. - 2014. - No. 1. - С. 156-174.
4. Болдачев А.В. Эссе: мифы и тупики искусственного интеллекта // Философские проблемы информационных технологий и киберпространства. - No. 2 (18). - С. 27-41.
5. Бостром, Н. Искусственный интеллект. Этапы. Угрозы. Стратегии. - СПб: Манн, Иванов и Фербер, 2016. - 490 с.
6. Булавинова М.П. Риски и угрозы новых технологий на основе искусственного интеллекта. (обзор) // Социально-гуманитарные науки. Отечественная и зарубежная литература. Сер. 8, Наука о науке: Реферативный журнал. - 2018. - No. 2. - С. 23-41.

7. Гуров О. Н. Искусственные интеллектуальные системы и решение социальных задач: проблема доверия // Искусственные общества. – 2021 – Т. 16 – Выпуск 1 URL: <https://artsoc.jes.su/s207751800014095-3-1/>. DOI: 10.18254/S207751800014095-3
8. Гуров О. Н. Этичное взаимодействие с интеллектуальными системами // Искусственные общества. – 2020 – Т. 15 – Выпуск 3 URL: <https://artsoc.jes.su/s207751800010905-4-1/>. DOI: 10.18254/S207751800010905-4
9. Гуров О.Н. Метавселенная — из сумерек во тьму перелетая? // Наука телевидения 2022 18 (1). С. 11–46. DOI: <https://doi.org/10.30628/1994-9529-2022-18.1-11-46>
10. Дзялошинский И.М. Искусственный интеллект: Гуманитарная перспектива // Вестник Новосибирского государственного университета. Серия: История, филология. - 2022. - №6. - Р. 20-29.
11. Дубровский, Давид. (2021). Задача создания искусственного общего интеллекта и проблема сознания. Российский журнал философских наук. 64. 13-44. 10.30727/0235-1188-2021-64-1-13-44.
12. Дубровский, Давид. (2022). Эпистемологический анализ социальной и гуманитарной значимости инноваций искусственного интеллекта в контексте искусственного общего интеллекта. Российский журнал философских наук. 65. 10-26. 10.30727/0235-1188-2022-65-1-10-26.
13. Карпов В. Е., Готовцев П. М., Ройзензон Г. В. К вопросу об этике и системах искусственного интеллекта // Философия и общество. - 2018. - No. 2 (87). - Р. 84-105.
14. Курцвейл, Р. Эволюция разума. Как расширение возможностей нашего разума решит многие мировые проблемы. - М.: ЭКСМО, 2015. - 352 р.
15. Финн В.: «Не все функции естественного интеллекта могут быть формализованы и автоматизированы» // Коммерсантъ URL: <https://www.kommersant.ru/doc/4198609?ysclid=lafntgvfcq626184900> (дата обращения: 11/14/2022).
16. Хайдеггер М. Время и бытие: статьи и выступления. - М.: Республика, 1993. - 447 р.
17. Этика и «цифра» - краткое резюме. Робот-врач, робот-учитель, робот-полицейский: социальные риски и этические вызовы для промышленности: Краткий обзор политики для второго тома отчета «Этика и цифровые технологии: Этические вызовы цифровых технологий». - Москва: РАНХИГС, 2020. - 45 с.
18. AI Designed a Fairer Tax System // Hightech.fm URL: <https://hightech.fm/2020/05/07/ai-economist> (Accessed: 11/14/2022).
19. Anton, Eduard & Kus, Kevin & Teuteberg, Frank. (2021). Is Ethics Really Such a Big Deal? The Influence of Perceived Usefulness of AI-based Surveillance Technology on Ethical Decision-Making in Scenarios of Public Surveillance. 10.24251/HICSS.2021.261.
20. Barrat J. (2013). Our final invention: artificial intelligence and the end of the human era (First). Thomas Dunne Books.
21. Bookstaber, Richard. (2017). The End of Theory: Financial Crises, the Failure of Economics, and the Sweep of Human Interaction. 10.1515/9781400884964.
22. Comi, Antonio & Nuzzolo, Agostino. (2014). Individual pre-trip path choice modeling in transit advisor tools: theoretical and empirical evidences.
23. Goertzel, Ben. (2015). Superintelligence: Fears, Promises and Potentials: Reflections on Bostrom's Superintelligence, Yudkowsky's From AI to Zombies, and Weaver and Veitas's "Open-Ended Intelligence". Journal of Ethics and Emerging Technologies. 25.55-87. 10.55613/jeet.v25i2.48.

24. Gurjar, Arunaben & Patel, Shitalben. (2022). Fundamental Categories of Artificial Neural Networks. 10.4018/978-1-6684-2408-7.ch001.
25. Kopalle, Praveen & Gangwar, Manish & Kaplan, Andreas & Ramachandran, Divya & Reinartz, Werner & Rindfleisch, Aric. (2021). Examining Artificial Intelligence (AI) Technologies in Marketing Via a Global Lens: Current Trends and Future Research Opportunities. *International Journal of Research in Marketing*. 39. 10.1016/j.ijresmar.2021.11.002.
26. McHale, Brian. (2013). Afterword: Reconstructing Postmodernism. *Narrative*. 21. 357-364. 10.1353/nar.2013.0020.
27. Muggleton, Stephen, and Nicholas Chater (eds), *Human-Like Machine Intelligence* (Oxford, 2021; online edn, Oxford Academic, 19 Aug. 2021), <https://doi.org/10.1093/oso/9780198862536.001.0001>, accessed 12 Dec. 2022.
28. Munira, Fayzuloeva. (2022). Ethical Ideas of the Ancient World. *Annals of Bioethics & Clinical Applications*. 5. 10.23880/abca-16000222.
29. Omrani, Nessrine & Riveccio, Giorgia & Fiore, Ugo & Schiavone, Francesco & Garcia-Agreda, Sergio. (2022). To trust or not to trust? An assessment of trust in AI-based systems: Concerns, ethics and contexts. *Technological Forecasting and Social Change*. 181. 121763. 10.1016/j.techfore.2022.121763.
30. Palladino, Nicola. (2021). The role of epistemic communities in the «constitution-alization» of internet governance: The example of the European Commission High-Level Expert Group on Artificial Intelligence. *Telecommunications Policy*. 45. 102149. 10.1016/j.tel-pol.2021.102149.
31. Ralph, Jason. (2013). The liberal state in international society: Interpreting recent British foreign policy. *International Relations*. 28. 3-24. 10.1177/0047117813486822.
32. Rawat, Romil & Yadav, Rishika. (2021). Big Data: Big Data Analysis, Issues and Challenges and Technologies. *IOP Conference Series: Materials Science and Engineering*. 1022. 012014. 10.1088/1757-899X/1022/1/012014.
33. Schneider, S. (2019). *Artificial You: AI and the Future of Your Mind*. Princeton University Press. <https://doi.org/10.2307/j.ctvfjd00r>
34. Schultz, Robert A. "Political Realism and the Society of Societies." In *Information Technology and the Ethics of Globalization: Transnational Issues and Implications*. Hershey, PA: IGI Global, 2010. <https://doi.org/10.4018/978-1-60566-922-9.ch006>
35. Thanetsunthorn, Namporn. (2021). Organization development and cultural values of trust in international contexts. *Review of International Business and Strategy*. ahead-of-print. 10.1108/RIBS-04-2021-0054.
36. The AI Economist: Improving Equality and Productivity with AI-Driven Tax Policies // Salesforce URL: <https://blog.salesforceairesearch.com/the-ai-economist/> (Accessed 11/14/2022).
37. Vakkuri, Ville & Kemell, Kai-Kristian & Jantunen, Marianna & Abrahamsson, Pekka. (2020). "This is Just a Prototype": How Ethics Are Ignored in Software Startup-Like Environments. 10.1007/978-3-030-49392-9_13.