

Научная статья
УДК 1:316+17
DOI: 10.18101/1994-0866-2023-2-76-85

К ВОПРОСУ О НЕОБХОДИМОСТИ СОЦИАЛЬНО-ФИЛОСОФСКОГО АНАЛИЗА ПРОБЛЕМ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

© **Бадмаева Майна Харлановна**

ведущий специалист,
Бурятский государственный университет имени Доржи Банзарова
Россия, 670000, г. Улан-Удэ, ул. Смолина, 24а
badmaevamaina@gmail.com

© **Золхоева Мария Валентиновна**

доктор философских наук, профессор,
Бурятский государственный университет имени Доржи Банзарова
Россия, 670000, г. Улан-Удэ, ул. Смолина, 24а
badmaeva_maria@mail.ru

Аннотация. В настоящее время мы являемся свидетелями увеличения масштабов внедрения новых технологий во все сферы жизнедеятельности человека. Искусственный интеллект (далее — ИИ) чрезвычайно активно проникает в жизненный уклад общества. Это открывает целый ряд новых возможностей и преимуществ для человека, однако порождает потенциальные риски и угрозы, усугубляющие существующие в обществе проблемы: предвзятое и несправедливое отношение к человеку, рост безработицы, трансформация человеческой идентичности, угроза безопасности и др. Кроме того, ИИ создает специфические трудности, обусловленные его природой, среди которых проблемы алгоритмической предвзятости, ошибки при работе с большими объемами данных, принципиальная непрозрачность действий ИИ. Поэтому необходимо осторожное, осмысленное и подконтрольное человеку внедрение ИИ, чтобы обеспечить безопасность общества. В этих условиях философская рефлексия проблем применения ИИ позволяет глубже взглянуть на предполагаемые и существующие угрозы, возникающие вследствие взаимодействия человека и данной технологии и выработать адекватный ответ на вызовы искусственного разума.

Ключевые слова: философия, социальная философия, искусственный интеллект, узкий искусственный интеллект, системы искусственного интеллекта, алгоритмическая предвзятость, дискриминация, социальная несправедливость, автономия человека, непрозрачность систем искусственного интеллекта.

Для цитирования

Бадмаева М. Х., Золхоева М. В. К вопросу о необходимости социально-философского анализа проблем искусственного интеллекта // Вестник Бурятского государственного университета. Философия. 2023. Вып. 2. С. 76–85.

Искусственный интеллект (далее — ИИ) является одной из наиболее популярных тем для научных дискуссий, творческих работ, разговоров и обмена мнениями в повседневной реальности. Несмотря на высокую степень заинтересованности научного и философского сообщества к порождаемым им проблемам, сегодня еще не сложился консенсус даже в отношении определения искусственного интеллекта. Многочисленность и разнообразие существующих дефиниций ИИ

свидетельствуют о сложной, комплексной природе этого феномена, понимание смысла и значения которого необходимо для выявления и оценки проблем, связанных с разработкой и внедрением ИИ в жизнь современного человека. В принятом нами исследовании мы придерживаемся наиболее детального и всеобъемлющего, на наш взгляд, определения ИИ, согласно которому последний представляет собой «системы, разработанные людьми, которые, имея сложную цель, действуют в физическом или цифровом мире, интерпретируя собранные структурированные или неструктурированные данные. Системы используют полученные данные при принятии наиболее подходящих и эффективных решений (в соответствии с заранее определенными параметрами) для достижения поставленной цели. Системы ИИ также могут быть разработаны таким образом, чтобы научиться адаптировать свое поведение, анализируя, как их предыдущие действия влияют на окружающую среду»¹⁴. Данное определение помимо выявления наиболее характерных для ИИ черт указывает на его системный характер, который определяется тем фактом, что составные элементы ИИ представляют не простую совокупность или последовательность, а целостность и взаимосвязанность составляющих компонентов: программно-аппаратного комплекса и набора данных. Кроме того, в данном определении упоминается один из передовых методов ИИ — самообучающийся тип, который учится на подкреплении без специально введенных человеком данных или его руководства. Иными словами, самообучающийся ИИ может в отсутствие каких-либо дополнительно введенных данных, без овладения специальными знаниями решать поставленные задачи в определенной, конкретной сфере реальности, во многих отношениях превосходя создавшего его человека.

Подобные системы представляют существующий в настоящее время искусственный интеллект, который принято обозначать термином «узкий» (или «слабый») ИИ, способный выполнять определенную (узкую) интеллектуальную задачу (или ограниченный набор задач). В отличие от него сильный (или общий) ИИ — это настоящие мыслящие машины с интеллектом, подобным человеческому¹⁵ [19]. Сильный ИИ еще не создан. Многие ученые считают, что вопрос о его существовании или создании сегодня является спекулятивным и существует лишь на уровне теории, но не на практическом применении. В фокусе нашего исследования находится вышеупомянутый узкий ИИ, активное использование которого уже сегодня порождает целый ряд как положительных, так и негативных последствий.

Узкий ИИ ранее применялся в некоторых областях реальности для решения отдельных задач. При этом внутренняя его работа была понятна, поддавалась интерпретации человеком, даже если для этого было необходимо осуществлять чрезвычайно сложные вычисления (например, в игре го или в шахматах). В настоящее время ИИ проник практически во все области человеческой жизнедеятельности, начиная от сельского хозяйства до создания самых современных

¹⁴ AI HLEG. A definition of AI: Main capabilities and scientific disciplines. Bruxelles: European Commission. 2018.

¹⁵ Search Enterprise AI. Artificial intelligence. URL: <https://searchenterpriseai.techtarget.com/definition/AI-Artificial-Intelligence> (дата обращения: 15.03.2023). Текст: электронный.

видов автономного вооружения. Столь широкое применение систем ИИ в жизненно важных для человека отраслях, безусловно, свидетельствует о том, что они становятся все более значимыми для человека, имеют явные преимущества по сравнению с любыми предшествующими технологиями, такими как быстрое распространение информации, автоматизация рутинных видов работ, обработка огромных баз данных и др. Но напрашивается вопрос, насколько мы готовы к такому крупномасштабному проникновению ИИ в нашу жизнь? Внедрение ИИ порождает ряд проблем, возникающих в процессе взаимодействия ИИ и человека, которые сегодня не решены и создают угрозы и риски для человека и общества, заставляющие пересмотреть онтологический и моральный статус как «умных» машин, так и человека.

Среди основных проблем применения ИИ, возникающих в социальной сфере жизни общества, выделяются проблемы занятости, безработицы и деградации, а также связанные с ними проблемы нарушения автономии человека, социально-экономического неравенства и многие другие.

В контексте внедрения технологии ИИ проблема занятости предполагает, что искусственный разум успешно заменяет людей при решении конкретных задач. Преимуществом ИИ в данном случае является его способность выполнять простые задачи (например, оценивать качество фруктов [15]), не совершая «человеческие» ошибки (пресловутый «человеческий фактор») и не предъявляя «человеческие» требования (социальные гарантии, защита прав трудящихся и т. д.). Например, система ИИ под названием COIN, используемая в юридической практике, может помогать в интерпретации кредитных соглашений и экономить до 360 тыс. рабочих часов в год [20]. Такое применение ИИ, на первый взгляд, вряд ли приведет к безработице глобальных масштабов и маргинализации населения. Тем более если человеку будет предоставлена возможность взяться за другую сопоставимую или более интересную работу. Привлечение ИИ в этом случае высвободит время от выполнения неинтересной, скучной, однообразной и утомительной работы и создаст возможности для дальнейшего обучения, совершенствования и развития людей. Тем не менее М. Гиббс считает, что в любых ситуациях, в которых системы ИИ полностью заменяют работника, даже если вся работа будет состоять из простых и рутинных задач, вероятность безработицы возрастает [7]. Так, в нашумевшем исследовании Фрея и Осборна приводится, что 47% всех рабочих мест в США находятся под угрозой технологической замены в течение ближайших двух десятилетий [22]. Однако наиболее примечательным аспектом автоматизации на основе систем ИИ является способность их работать и обучаться автономно, а не действовать по заданной инструкции. На этом основании некоторые ученые утверждают, что автоматизацию с помощью ИИ можно назвать «высококвалифицированной» [23]. Это может привести к проблеме безработицы в областях, в которых активно задействован интеллектуальный и творческий потенциал человека. Например, в области медицины запущены системы, которые в будущем могут заменить специалистов, осуществляющих на основе симптомов диагностику заболевания и консультирующих пациентов о предполагаемом плане лечения. Известен случай, когда система на основе глубокого обучения, используя около 112 тыс. снимков более 30 тыс. пациентов, успешно выявила пневмонию. Такая автоматизация способна менять человеческое поведение

и устоявшиеся алгоритмы выполнения работы, что в конечном итоге с большой долей вероятности приведет к проблеме деквалификации, которая окажет негативное влияние на идентичность человека, его чувство собственного достоинства и самоуважения. Другой пример — использование авиационных систем ИИ и, как следствие, деквалификация рабочих в этой области. Современные бортовые системы могут самостоятельно, без участия пилотов, обеспечить взлет, перелет с места на место и посадку воздушного судна. Зачем в таком случае нужен высококвалифицированный пилот, на подготовку которого необходимо потратить значительное количество времени и ресурсов, и при этом нет никакой гарантии, что он всегда сможет успешно справляться со своими профессиональными задачами? Как видно из приведенных выше примеров, системы ИИ обретают возможность влиять на характер происходящих в сфере занятости изменений, порождая новые, неожиданные, противоречивые и негативные последствия для занятости людей, их самооценки, самореализации, понимания своего места и роли в мире.

Делегирование полномочий от человека ИИ в выполнении ряда работ заставляет задуматься о том, чем грозит для человека такая активная сопричастность и вовлеченность этой технологии в решение столь важных вопросов и задач? Вопрос о выборе конкретной совокупности поручений, выполнение которых можно доверять системам ИИ, опирается не только в практический, экономический аспект, а с необходимостью поднимает более важный экзистенциальный вопрос о том, что значит быть человеком и в чем состоит собственно человеческое. Ведь этот выбор образует контур жизни любого человека и определяет место, занимаемое им в мире. К. Маркс утверждал, что родовая сущность человека заключается в продуктивной творческой деятельности. В «Капитале» он писал, что «...труд как созидатель потребительских стоимостей, как полезный труд есть независимое от всяких общественных форм условие существования людей, вечная естественная необходимость: без него не был бы возможен обмен веществ между человеком и природой, то есть не была бы возможна сама человеческая жизнь». Способность к труду, трудовая деятельность составляют родовую сущность человека, его характерную особенность. Автоматизация на основе ИИ потенциально способна исключить человека из процесса творческой созидательной деятельности, лишив его своего предназначения и превратив творца в организм, пассивно воспринимающий идущую извне информацию. Вовлечение в реальную практику, решение разнообразных задач, постановка целей традиционно обогащают человека знаниями и формируют его личность, собственно человеческий образ жизни.

Конечно, ИИ в ряде случаев помогает работникам успешно выполнять свою работу, расширяя возможности человека и не пытаясь заменить его. Так, системы ИИ могут быстро и эффективно сопоставлять разрозненные источники данных для анализа характеристик и историй жертв домашнего насилия, что позволяет представителям закона применять адекватные и необходимые меры реагирования [8]. Это демонстрирует потенциальное преимущество ИИ, заключающееся в том, что он способен положительно повлиять на возможность людей решать сложные проблемы, повысить результативность выполняемой ими работы и создать условия для генерирования новых идей, поиска лучшего решения из

множества возможных. Тем не менее эти преимущества не снимают проблему безработицы и обусловленных ею негативных последствий.

Проблема безработицы, возникающая при применении систем ИИ, тесно связана с еще одной проблемой социального спектра: с проблемой автономии человека. Автономию в контексте взаимодействия человека с системами ИИ следует понимать как степень свободы, независимости человека и возможность самостоятельно планировать свою работу [9]. Автономия отражает то, насколько свободно люди могут определять свои подходы к решению проблем и задач и какова степень этой свободы. Чем более автономны работники, тем выше их способность проявлять самостоятельность в принятии решений, выбирать ту работу, которая способствует саморазвитию, нравственному совершенствованию и позволяет придерживаться собственных ценностей, жить в соответствии с принципами, которые они действительно разделяют. Применение систем ИИ негативно влияет на автономию человека, поскольку алгоритмы самообучающихся систем не предполагают учет его мнения и позиции. Сам человек не способен контролировать их деятельность. В результате он оказывается ограничен в выборе возможных вариантов решения проблемы теми рамками, которые предлагает искусственный разум. Автономия человека оказывается зависимой от воли и решения искусственного интеллекта. Автономия машины обуславливает, сокращает, ущемляет свободу и самостоятельность человека в принятии им решений.

Системы ИИ, принимающие автономные решения, уже сегодня приводят к аварийным ситуациям, причинению вреда человеку. Например, швейцарская полиция арестовала бота для онлайн-покупок под названием Random Darknet Shopper, который совершал незаконные покупки (в том числе наркотики) в даркнете. Другой известный пример — авария со смертельным исходом при участии автоматизированного транспортного средства на основе систем ИИ¹⁶. В этом свете по-иному встает вопрос о субъекте ответственности, который должен отвечать в рамках закона за нанесенный ущерб. Является ли таковым разработчик, производитель, собственник или сама система ИИ? В настоящее время гражданское и уголовное право все еще не имеет готового ответа на подобные вопросы.

Применение систем ИИ также порождает проблему нарушения конфиденциальности, которая может возникать на любом этапе обработки данных: во время сбора информации, при анализе данных и их использовании, в процессе удаления данных. Приведем в качестве примера запуск технологии под названием Status Today, которая позволяет фирмам отслеживать производительность сотрудников [24]. Технология ежеминутно проверяет поведение персонала, собирая данные о том, кто кому и в какое время отправляет электронные письма, кто получает доступ к файлам и редактирует их, а также кто с кем встречается. Такое наблюдение фактически увеличивает психологическое давление на работников, нарушая их личные границы и лишая возможности полноценно трудиться и отдыхать.

Социально-философское осмысление проблем ИИ позволяет поднимать вопросы внедрения технологии ИИ в социальную практику, особенно вопросы

¹⁶ NTSB. Highway Preliminary Report: HWY19FH008. URL: <https://ntsb.gov/investigations/AccidentReports/Pages/HWY19FH008-preliminary-report.aspx> (дата обращения: 19.03.2023). Текст: электронный.

целесообразности поспешного и необдуманного внедрения ИИ, способного привести к нежелательным для человека последствиям, изменив его привычный жизненный уклад. Так, постоянно растущее использование ИИ в тех сферах, где человеческие качества традиционно считались необходимыми (например, в области медицины), может повлиять на пересмотр человеком восприятия самого себя или даже способствовать утрате им собственной идентичности посредством стирания грани между ним и ИИ. Автоматизация интеллектуальных или творческих видов деятельности, в свою очередь, грозит психологической неуверенностью и беспокойством, вызванным неверием человека в собственную уникальность, незаменимость, необходимость «быть человеком» на том основании, что искусственный интеллект выполняет работу быстрее, лучше и без ошибок.

Безусловно, приведенные соображения не означают, что ИИ нужно исключить из жизни современного общества или уничтожить. Тем более это сегодня уже не представляется возможным в силу тотального присутствия систем ИИ в жизни человека и человечества. Они лишь призывают к осторожному и осмысленному обращению с данной технологией.

Одной из наиболее актуальных проблем применения систем ИИ выступает проблема алгоритмической предвзятости. Речь идет о том, что системы ИИ способны принимать необъективные решения, что может усугублять существующие в обществе формы несправедливого отношения, предвзятости [11], несмотря на то, что внедрение систем ИИ изначально осуществлялось с целью обеспечить максимально возможную беспристрастность при рассмотрении спорных вопросов и проблем. Алгоритмическая предвзятость выражается в том, что системы ИИ проявляют склонность обрабатывать данные, отклоняясь от нормативных, принятых в обществе социальных, моральных стандартов, в результате чего определенные лица или группы оказываются обделенными или ущемленными в своих правах и привилегиях. Среди наиболее часто упоминаемых синонимов предвзятости в литературе указываются несправедливость [17] и дискриминация со стороны ИИ [10].

Источником предвзятости систем ИИ зачастую становится использование больших данных, которые изначально могут содержать предубеждения, порождающие дискриминацию отдельных лиц и целых социальных групп. В литературе данный феномен принято обозначать термином «врожденная предвзятость данных». Он может включать, например, плохо маркированные данные [2].

Предвзятость алгоритмов, дискриминация и несправедливость были выявлены в различных системах ИИ, таких как прогностические приложения в правосудии [12], распознавание лиц [4], модели прогнозирования в здравоохранении [14]. Например, в правосудии модели ИИ давали ошибочное предсказание поведения будущего преступника. В медицине предвзятость проявилась в тех случаях, когда система ИИ определяла, каким лицам и группам необходима медицинская помощь. Оказалось, что она может дискриминировать их по полу, расовой принадлежности, возрасту и т. д. Э. Стрикланд приводит пример, когда система выдала результат о том, что афроамериканцы менее нуждаются в медицинской помощи по сравнению с представителями белой расы [18]. В итоге выяснилось, что изначально в алгоритме не было заложено никакой информации о дискриминирующем отношении к афроамериканцам, однако система пришла к такому выво-

ду на основании расходов пациентов на медицинскую помощь за предыдущие годы. Следовательно, тем, кому практически не оказывалась медицинская помощь из-за низкого уровня доходов, были выделены системой ИИ как лица, которым она менее требовалась. Проблема предвзятости систем ИИ с высокой степенью вероятности привела к тому, что такие компании, как Amazon, потеряли доверие к ним и отказались от автоматизированных систем для принятия управленческих решений [5].

Исследования, проводимые учеными сегодня, показывают, что наиболее часто встречающийся вид алгоритмической предвзятости касается гендерных и расовых предубеждений [21]. У. Питерс исследует предвзятость, при которой системы ИИ способны на основе «цифровых следов» определять политические взгляды человека, даже если он их явно и открыто не выражает. Насколько оправданно использование подобных систем, которые способны причинить вред человеку, а он при этом не имеет возможности в силу принципиальной неинтерпретируемости данных, которыми оперирует система ИИ, даже предположить, почему, на основании каких критериев (раса, цвет волос, потребительские привычки, политические предпочтения и т. д.) система поступила несправедливо по отношению к нему? Очевидно, что в подобных случаях алгоритмы ИИ ограничивают человеческие права и свободы, человеческую автономию, лишая человека возможности самоопределения.

Проблема предвзятости усугубляется непрозрачной природой ИИ, когда принципы внутренней работы технологии остаются скрытыми от человека. Непрозрачность технологии ИИ вызвана применением метода машинного обучения на искусственных нейронных сетях, при котором входные и выходные данные составляют два отдельных слоя, последние при этом оказываются неожиданными даже для самих разработчиков. Так, в своем исследовании Р. Бенджамин исследует проблему предвзятости в контексте занятости, когда работники могут не догадываться, почему было принято то или решение в их отношении. Это препятствует взвешенному, консультативному принятию решений руководством.

Ситуацию усложняет, на наш взгляд, еще и то обстоятельство, что системы ИИ не обладают подлинной способностью интерпретировать, понимать информацию. Они принимают решения, не понимая сути происходящего. Они действуют, не отдавая себе отчета в необходимости и правильности этих действий.

Применение предвзятых и непрозрачных систем ИИ, принимающих решение на основании неизвестных и зачастую случайных событий или особенностей, свойств, выливается в проблему причинения вреда человеку, поднимая вопрос об экзистенциальной угрозе для него, вызванной негативным влиянием систем ИИ на родовую сущность человека как единственного существа, способного в известном нам мире к целенаправленной, осознанной природопреобразующей активности. Проблема причинения вреда человеку системами ИИ, по нашему мнению, представляет собой ключевой принцип благодеяния. Системы ИИ, способные к самообучению, обладающие относительной долей автономии, уже сегодня актуализируют вопрос о дальнейшей судьбе человека в мире, смысле его жизни и предназначении.

Таким образом, быстрый прогресс в области ИИ оказывает глубокое влияние на жизнь общественной системы и всех ее подсистем. Системы ИИ, способные

принимать автономные решения, взаимодействовать со всеми заинтересованными сторонами в социуме, могут стать причиной самых непредвиденных ситуаций вплоть до фундаментального цивилизационного сдвига с далеко идущими последствиями. Становится понятно, что наше общество еще не готово к этой стремительно приближающейся смене парадигм, слишком много вопросов недостаточно исследованы и рассмотрены. Философский анализ социальных проблем применения ИИ уже стоит в сегодняшней повестке дня, уже нуждается в широком и профессиональном освещении. Обозначенные в данной статье негативные последствия применения систем ИИ не являются изолированными инцидентами, а напротив, свидетельствуют о действительно глубоком, сущностном уровне поднимаемых проблем. Исследование с позиций социальной философии результатов внедрения ИИ в широкую практику позволит рассмотреть их в целостности и комплексе, не упуская экономические, политические, собственно социальные, этические и другие смыслы, аспекты, грани проблемы. Кроме того, такой подход позволит заново переосмыслить экзистенциальные вопросы о природе и родовой сущности человека, о его месте в изменившемся мире.

Литература

1. Маркс К., Энгельс Ф. Капитал. Критика политической экономии. 1960. Москва: Политиздат. 1960. Т. 23. С. 51. Текст: непосредственный.
2. Ashrafián H. Engineering a social contract: Rawlsian distributive justice through algorithmic game theory and artificial intelligence // *AI Ethics*. 2022.
3. Benjamin R. Assessing risk, automating racism // *Science*. 2019. 366. Vol. 6464. P. 421–422.
4. Buolamwini J., Gebru T. Gender shades: Intersectional accuracy disparities in commercial gender classification // *Proceedings of conference on fairness, accountability and transparency, FAT2018*. 2018. P. 77–91.
5. Dastin J. Amazon scraps secret AI recruiting tool that showed bias against women.reuters.com. URL: <https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G> (дата обращения: 11. 03. 2023)
6. Eveleth R. My robot bought illegal drugs. BBC. 2015. URL: <https://tinyurl.com/qbgglw7> (дата обращения: 28.03.2023).
7. How is new technology changing job design? / M. B. Gibbs, S. Bazylik et al. // *IZA World of Labor*. 2022. 464 p.
8. Grogger J., Ivandic R., Kirchmaier T. Comparing conventional and machine-learning approaches to risk assessment in domestic abuse cases // *Journal of Empirical Legal Studies*. 2020. Vol. 18(1). P. 90–130.
9. Hackman J. R., Oldham G. R. Motivation through the design of work: Test of a theory // *Organizational Behavior and Human Performance*. 1976. Vol. 16(2). P. 258.
10. Hacker P. Teaching Fairness to Artificial Intelligence: Existing and Novel Strategies Against Algorithmic Discrimination under EU Law r // *Common Market Law Review*. 2018. № 55(4).
11. Jackson J. R. Algorithmic Bias // *Journal of Leadership, Accountability & Ethics*. 2018. Vol. 15, Issue 4. P. 55–65.
12. How we analysed the COMPAS recidivism algorithm / J. Larson, S. Mattu, L. Kirchner, J. Angwin // *ProPublica*. 2016. Vol. 5(2016). P. 9.

13. Oakden-Rayner L. CheXNet: An In-Depth Review. URL: <https://laurenoakdenrayner.com/2018/01/24/chexnet-an-in-depth-review> (дата обращения: 28.03.2023).
14. Dissecting racial bias in an algorithm used to manage the health of populations / Z. Obermeyer, B. Powers, C. Vogeli, S. Mullainathan // Science. 2019. Vol. 366(6464). P. 447–453.
15. Roberts P. Working smarter with data. Australian Manufacturing Forum. URL: <https://www.aumanufacturing.com.au/working-smarter-with-data-ai-gives-agriculture-the-competitive-edge> (дата обращения: 30.03.2023).
16. Mastering the game of go without human knowledge / D. Silver, J. Schrittwieser, K. Simonyan, I. Antonoglou, et al. // Nature. 2017. Vol. 2017550(7676). P.354–359.
17. Smith L. Unfairness by Algorithm: Distilling the Harms of Automated Decision-Making. Future of Privacy Forum. URL : <https://fpf.org/2017/12/11/unfair-ness-by-algorithm-distilling-theharms-of-automa-ted-decision-making> (дата обращения: 10.04.2023).
18. Strickland E. Racial Bias Found in Algorithms That Determine Health Care for Millions of Patients // IEEE Spectrum. 2019.
19. Wang P., Goertzel B., Franklin S. Artificial general intelligence, 2008: Proceedings of the first AGI conference / P. Wang, B. Goertzel, S. Franklin. Washington DC: IOS Press, 2008. 507 p.
20. Wall L. Some financial regulatory implications of artificial intelligence // J. Econ. Bus. 2018. Vol. 100. P. 55–63.
21. West S. M., Whittaker M., Crawford K. Discriminating systems: Gender, race and power in AI. AI Now Institute / S. M. West, M. Whittaker, K. Crawford. URL: <https://ainowinstitute.org/discriminatingystems.html> (дата обращения: 12.04.2023).
22. Frey C. B., Osborne M. A. The future of employment: How susceptible are jobs to computerisation? // Technol. Forecast Soc. Chang. 2017. Vol. 114. P. 254–280.
23. Acemoglu D., Restrepo P. The race between man and machine: Implications of technology for growth, factor shares, and employment // Am Econ. Rev. 2018. Vol. 108(6). P. 1488–1542.
24. Haenlein M., Huang M. H., Kaplan A. Guest Editorial: Business Ethics in the Era of Artificial Intelligence // Journal Business Ethics. 2022. P. 1–3.

Статья поступила в редакцию 20.04.2023; одобрена после рецензирования 24.04.2023; принята к публикации 26.04.2023.

ON THE NECESSITY OF SOCIO-PHILOSOPHICAL ANALYSIS OF ARTIFICIAL INTELLIGENCE ISSUES

Maina Kh. Badmaeva

Leading Specialist,

Dorzhi Banzarov Buryat State University

24a Smolina St., Ulan-Ude 670000, Russia

badmaevamaina@gmail.com

Mariya V. Zolkhoeva

Dr. Sci. (Philos.), Head of Philosophy Department,

Dorzhi Banzarov Buryat State University

24a Smolina St., Ulan-Ude 670000, Russia

badmaeva_maria@mail.ru

Abstract. As we continue to witness the rapid implementation of new technologies across all aspects of human activity, the integration of artificial intelligence (AI) into society is becoming increasingly prevalent. While this presents many exciting opportunities and advantages for humans, it also poses potential risks and threats that exacerbate existing societal problems such as biased and unfair treatment, unemployment, and security threats. Moreover, AI's unique nature creates specific difficulties such as algorithmic bias, errors in processing large amounts of data, and the fundamental opacity of AI actions. Therefore, it is crucial that AI is implemented with caution, thoughtfulness, and human control to ensure the safety of society. In these circumstances, philosophical reflection on the challenges of AI application enables us to examine both the presumed and existing threats arising from human-AI interaction and develop an appropriate response to these challenges.

Keywords: philosophy, social philosophy, artificial intelligence, narrow artificial intelligence, artificial intelligence systems, algorithmic bias, discrimination, social injustice, human autonomy, opacity of artificial intelligence systems.

For citation

Kharlanova M. B., Zolkhoeva M. V. On the Necessity of Socio-Philosophical Analysis of Artificial Intelligence Issues. *Bulletin of Buryat State University. Philosophy.* 2023; 2: 76–85 (in Russ.)

The article was submitted 20.04.2023; approved after review 24.04.2023; accepted for publication 26.04.2023.