

МИНОБРНАУКИ РОССИИ

**Федеральное государственное
автономное образовательное
учреждение высшего образования
«Национальный исследовательский
Томский государственный университет»
(ТГУ, НИ ТГУ)**

Ленина пр., 36, г. Томск, 634050
Тел. (3822) 52-98-52, факс (3822) 52-95-85
E-mail: rector@tsu.ru
http://www.tsu.ru
ОКПО 02069318, ОГРН 1027000853978
ИНН 7018012970, КПП 701701001

№ _____
на № _____ от _____

УТВЕРЖДАЮ

и.о. проректора по научной и
инновационной деятельности
канд. геол.-минерал. наук,
Томского государственного
университета
Т.С.Краснова



« 6 » июня 2023 г.

ОТЗЫВ ВЕДУЩЕЙ ОРГАНИЗАЦИИ
о диссертации Бадмаевой Майны Харлановны
«Социально-философские проблемы и принципы применения систем
искусственного интеллекта»,
представленной на соискание ученой степени кандидата
философских наук по специальности 5.7.7 — Социальная и политическая
философия

Кандидатская диссертация Бадмаевой Майны Харлановны посвящена исследованию актуальнейших проблем, обусловленных внедрением систем искусственного интеллекта (далее — ИИ) в различных областях социальной реальности. Их воздействие на человека и общество носит амбивалентный характер.

В условиях четвертой промышленной революции возрастает влияние NBICS-технологий, оказывающих преобразующее влияние на человека и общество, что сопровождается ростом проблем, обусловленных неопределенностью и рисками инноваций. При этом драйвером развития являются информационные и цифровые технологии. Цифровизация, с одной стороны, обеспечивает глобальную коммуникацию, автоматизацию труда, возможности прогнозирования на основе оперирования большими

данными, с другой стороны, проявляется угроза трансформации социальных связей, невиданного социального расслоения, виртуализации культуры, уничтожения личного пространства, манипулирования сознанием, компьютерной пропаганды, деформации мышления, снижения когнитивных способностей и другие. Выявление и анализ рисков цифровизации означает не отказ от цифровых технологий, что невозможно, но осознание необходимости формирования культурных регуляторов, включающих цифровую этику, обуславливающих механизмы коммуникации в сетевом обществе и принятие решений в интересах человека, а не цифрового капитализма.

Существующие модели ИИ способны выполнять сложные интеллектуальные задачи, требующие творческого и индивидуального подхода. В то же время системы ИИ, характеризующиеся не прогнозируемостью последствий, возникающих при их разработке и функционировании, уже сегодня воспринимаются в качестве новой угрозы свободе и благополучию человека в мире.

Постоянно расширяющийся перечень проблем, вызванных применением указанных технологий, среди которых, в частности, проблема обеспечения права человека на частную жизнь, гражданскую свободу, проблема непрозрачности ИИ, проблема отсутствия ответственности в случае возникновения негативных последствий, проблема сохранения идентичности человека в условиях широкого внедрения NBICS-технологий, где ИИ выступает в качестве «сквозной» технологии и многие другие, а также достигнутая системами ИИ степень автономии вне зависимости от наличия физической формы (аппаратного комплекса) заставляет ученых, разработчиков и пользователей систем ИИ переосмыслить существующие представления о месте ИИ в социальной системе, о целесообразности внедрения данных систем в различные сферы жизнедеятельности человека, об условиях безопасного и эффективного

использования указанных систем в ближайшей и отдаленной перспективе. Эти вопросы анализируются во второй главе диссертации.

В диссертационном исследовании показано, что сегодня формируется совершенно уникальная этическая ситуация, которая продиктована специфическим интеллектуальным характером и возрастающей сложностью ИИ, разрывом между технологическим развитием и отставанием этики науки и техники. Автор убедителен в том, что это обуславливает необходимость философского осмысления процесса внедрения ИИ в жизнь современного социума (Гл.I. параграф 1.3)

Автор на большом материале показывает, что проблемы, возникшие в связи с социальными трансформациями, вызванными развитием ИИ, носят междисциплинарный характер, их решение требует консолидации знаний и опыта инженеров, специалистов в области цифровых технологий, ученых, исследующих социальные и гуманитарные проблемы (философов, психологов, социологов и т.д.), юристов и политиков, представителей коммерческих структур, общественных организаций.

Все вышеперечисленное объясняет и доказывает безусловную *актуальность* представленного М.Х. Бадмаевой диссертационного исследования.

Важно отметить, что работа выполнена на основе изучения широкого круга литературы отечественной и зарубежной. Материал, положенный диссертантом в основу исследования, собран из многочисленных источников, представленных как научной литературой, так и широким кругом документов, посвященных регулированию этико-правовых аспектов применения систем ИИ в общественной практике и отражающих российский и зарубежный опыт их использования в различных областях деятельности современного общества. Безусловно, это большое достоинство данной работы.

На основании вышеизложенного отметим, что *значимыми для современной социальной философии являются следующие результаты,*

самостоятельно полученные автором диссертации:

- обосновано, что проблемы применения ИИ целесообразно исследовать с позиции социальной философии, что позволяет подойти к постановке и решению этих вопросов, на основе гуманистических принципов и ценностей, вырабатываемых в пространстве социально-философского взгляда на мир и место человека в мире;

- проанализирован международный опыт применения ИИ, представленный в диссертации обширным кругом источников (в том числе на английском и китайском языках), значительная часть которых не знакома русскоязычному научному сообществу;

- установлены и проанализированы с позиций социальной философии наиболее злободневные проблемы ИИ, а также используемые в социальной практике стратегии и пути их преодоления;

- выявлены сущности базовых принципов и норм, регулирующих безопасное и эффективное применение ИИ в различных сферах жизни социума. Показано, что указанные принципы образуют систему требований, отражающих различные стороны философских принципов безопасности, прозрачности, ответственности, конфиденциальности, справедливости и автономии.

- раскрыто содержание философских принципов взаимодействия человека с системами ИИ, представляющих собой условия его безопасного и эффективного использования.

Диссертантом осуществлен комплексный анализ значительного объема разнообразных конкретных примеров использования ИИ, обоснована целесообразность философского подхода к исследованию проблем влияния ИИ на человека и социум, носящих междисциплинарный характер.

М.Х. Бадмаева, охарактеризовав основные сферы применения ИИ (глава 1 и глава 2), выявила наиболее острые проблемы социального, этического, правового характера, возникающие в процессе взаимодействия

ИИ с человеком. В третьей главе диссертант обосновывает философские принципы, на которых должно основываться взаимодействие с системами ИИ и внедрение цифровых технологий: это принципы не причинения вреда, прозрачности, ответственности, конфиденциальности, справедливости и автономии (Глава 3).

Хочется подчеркнуть, что М.Х. Бадмаева обосновывает корреляцию содержания данных социально-философских принципов с междисциплинарными исследованиями проблем неопределенности и рисков внедрения ИИ на анализе весьма и весьма широкого круга источников, включающих научные, философские исследования, а также документы, регулирующие этико-правовые аспекты применения систем ИИ в общественной практике, отражающие как российский, так и зарубежный опыт их использования в различных областях. Это делает выводы диссертационного исследования более убедительными.

Все сказанное позволяет заключить, что выше обозначенные *результаты, полученные автором, являются теоретически значимыми и вносят вклад в сферу социальной философии, в такие ее разделы как глобальные проблемы современности и этические проблемы внедрения инновационных технологий.*

Практически значимыми представляются следующие результаты, полученные автором: выявлены и сформулированы смыслы требований, составляющих суть философских принципов применения ИИ, исключающих причинение любого вреда человеку и одновременно обеспечивающих развитие передовых технологий как в существующих социальных условиях, так и в отдаленной временной перспективе.

Все вышеизложенное свидетельствует о наличии *новизны* в проведенном исследовании, в котором с позиций гуманистически ориентированной социально-философской концепции обобщен и осмыслен существующий опыт взаимодействия человека с ИИ, сформулированы и содержательно охарактеризованы шесть основных принципов, на которых

должно строится внедрение и применение любых теоретических разработок в сфере ИИ.

Достоверность полученных в исследовании результатов обусловлена теоретико-методологической базой, позволившей автору аргументировано, с учетом работ предшественников, представить и обосновать основные положения исследования. Теоретические положения работы осмыслены с учетом достижений современной науки в области социальной философии, философии техники, философии компьютерных наук, философии информации и других смежных гуманитарных исследований. Достоверность рецензируемого исследования обусловлена также значительным объемом эмпирического материала, представленного целым комплексом документов, существующих в форме национальных стратегий, нормативно-правовых актов, этических руководств, рекомендаций и стандартов, изданных государственными институтами, бизнес-структурами, международными и неправительственными организациями, регулирующими этико-правовые аспекты использования ИИ. Исследуемый материал был отобран, классифицирован и систематизирован в Приложении к диссертации.

Недостатком работы является неудачно сформулированные новизна и положения, выносимые на защиту. Во многих формулировках новизны нет конкретики. В некоторых пунктах даже просто повторяется формулировка задачи с заменой слова «обосновать» на слово «обосновано» (стр.14 и 18).

Среди положений, выносимых на защиту, на наш взгляд, второе, третье и четвертое положения могут без ущерба для диссертации быть исключены. Положение №2 («Системы ИИ находят применение в ключевых сферах жизни социума ...» с.19) является констатацией известного факта и не требует защиты в рамках диссертационного исследования. Его следовало бы сформулировать иначе, сконцентрировав внимание на том, что «применение ИИ привело к возникновению

потенциальной угрозы пересмотра базовых мировоззренческих представлений о месте человека в мире, определяющем его роль и предназначение» с.19, о чем автор, но без акцента на этой идее и говорит, и что действительно заслуживает философского анализа.

Положение №3, в котором утверждается, что экзистенциальная угроза при встраивании систем ИИ в жизнь социума «потребовала разработки и принятия целого комплекса документов, существующих в форме национальных стратегий, нормативно-правовых актов, этических руководств, рекомендаций и стандартов» (с.19), рассматривается нами не как положение, выносимое на защиту, а как описание того эмпирического материала, с которым приходится иметь дело всякому исследователю, занимающемуся проблемой встраивания систем ИИ в жизнь социума.

В положении №4 диссертант акцентирует внимание на том, что «процесс применения ИИ требует глубокого и всестороннего социально-философского осмысления, фокусирующегося на человеке» (с.19). Думается, что правильнее было бы говорить о необходимости философского, осмысления, включая не только социальную философию, но философскую антропологию, онтологию, этику... Во второй части этого положения автор утверждает, что «Исследование с теоретических позиций данной дисциплины необходимо для выявления рисков, вызываемых применением систем ИИ, для существования человека, сохранения его места и роли в современном мире» (с.19). Видимо, имеется ввиду дисциплина «социальная философия», но совсем непонятно, с теоретических позиций какой науки она исследовалась для выявления рисков внедрения ИИ.

В высказанных замечаниях фиксируется не слабость работы, а наличие недостатков в формулировках столь важных для понимания сути диссертационного исследования пунктов, как положения, выносимые на защиту и новизна. Более глубокое прочтение текста убеждает в

неудачности именно формулировок ряда положений, выносимых на защиту и некоторых пунктов новизны.

Отмечу также, что было бы уместным более подробное рассмотрение истории развития ИИ.

Данные замечания не отрицают содержательных достоинств рецензируемого исследования, перечисленных ранее. Диссертационное исследование заслуживает положительной оценки.

Результаты исследования прошли необходимую апробацию. Основные результаты диссертации обсуждены на конференциях международного и общероссийского уровня. Материалы исследования изложены в 8 научных работах, отражающих основные результаты диссертации, из них в изданиях, включенных в Перечень рецензируемых научных журналов, в которых должны быть опубликованы основные научные результаты диссертаций на соискание ученой степени кандидата наук, на соискание ученой степени доктора наук (Перечень ВАК) - 4 статьи. По тематике ИИ издана монография, в автореферате диссертации отражено ее содержание.

Все сказанное позволяет сделать вывод, что диссертационное исследование Маины Харлановны Бадмаевой «Социально-философские проблемы и принципы применения систем искусственного интеллекта», представленной на соискание ученой степени кандидата философских наук является научно-квалификационной работой, которая имеет теоретическую и практическую значимость. В диссертации предлагаются философские принципы, на которых должна строиться стратегия безопасного и эффективного применения систем ИИ, что имеет значение для развития социальной философии, соответствует требованиям, изложенным в п.п. 9-11, 13, 14 Положения о присуждении ученых степеней, утвержденного Постановлением Правительства Российской Федерации от 24 сентября 2013 года № 842. Автор диссертации, Бадмаева Маина Харлановна, заслуживает присуждения искомой ученой степени

кандидата философских наук по специальности 5.7.7. – Социальная и политическая философия.

Отзыв подготовлен заведующим кафедрой философии и методологии науки философского факультета Федерального государственного автономного образовательного учреждения высшего образования «Национальный исследовательский Томский государственный университет» доктором философских наук, профессором Черниковой Ириной Васильевной.

Отзыв обсужден и утвержден на заседании кафедры философии и методологии науки философского факультета Федерального государственного автономного образовательного учреждения высшего образования «Национальный исследовательский Томский государственный университет» «1» июня 2023 г., протокол №10.

Присутствовали на заседании 11 чел. В обсуждении приняли участие 11 чел. Результаты голосования: «за» - 11 чел., «против» - 0 чел., «воздержались» - 0 чел.

Заведующий кафедрой философии и методологии науки философского факультета Федерального государственного автономного образовательного учреждения высшего образования «Национальный исследовательский Томский государственный университет» доктор философских наук, профессор



Черникова Ирина Васильевна

Подпись заведующего кафедрой философии и методологии науки философского факультета доктора философских наук, профессора Черниковой И.В. заверяю _____.

Дата 02.06.23г.

Федеральное государственное автономное образовательное учреждение высшего образования «Национальный исследовательский Томский государственный университет»,
634050, г. Томск, пр. Ленина, 36
Тел.: +7 (3822) 533-273, e-mail: kaffinn@mail.ru
Сайт: <https://www.tsu.ru>



Подпись удостоверяю
Ведущий документовед
Андреев И.В.