ISSN 2524-2369 (Print)
ISSN 2524-2377 (Online)

ФІЛАСОФІЯ І САЦЫЯЛОГІЯ

PHILOSOPHY AND SOCIOLOGY

УДК 113.3(13)
https://doi.org/10.29235/2524-2369-2026-71-4-271-282

Поступила в редакцию 15.04.2026
Received 15.04.2026

Г. В. Вышедко

Институт философии Национальной академии наук Беларуси, Минск, Беларусь

СОЦИАЛЬНО-ОНТОЛОГИЧЕСКИЙ СТАТУС ИСКУССТВЕННОГО ИНТЕЛЛЕКТА: ПАРАДИГМАЛЬНЫЕ ПРЕДСТАВЛЕНИЯ И СТРАТЕГИИ КОНЦЕПТУАЛИЗАЦИИ

Аннотация. Искусственный интеллект (далее – ИИ) является одним из определяющих факторов современного этапа общественного развития. Внедрение технологий ИИ сопряжено с неравномерностью и непредсказуемостью результатов его функционирования, что актуализирует проблему включения машинного интеллекта в социокультурные процессы. Решение данной задачи осложняется многозначностью самого понятия «искусственный интеллект». В современном дискурсе выделяются два основных подхода к его определению: онтологический (вопрос о сущности ИИ) и функционально-утилитарный (вопрос о возможностях ИИ). Анализ данных парадигм позволяет выявить корреляцию между способами концептуализации ИИ и характером его воздействия на социальные процессы. Обосновывается, что в условиях растущей сложности и непрозрачности нейросетевых моделей системы ИИ обретают «операциональную автономию» – способность достигать поставленных целей непредсказуемыми для человека способами. Данный феномен обусловливает появление «перформативности концептуализаций» ИИ: наши теоретические представления о его природе не просто описывают реальность, но и активно формируют ее, задавая векторы общественных ожиданий и практических действий. Этот тезис обосновывает необходимость дальнейшего исследования социально-онтологического статуса искусственного интеллекта.

Ключевые слова: искусственный интеллект, нейросети, общество, онтологический подход, функционально-утилитарный подход, операциональная автономия, перформативность концептуализаций, техника

Для цитирования: Вышедко, Г. В. Социально-онтологический статус искусственного интеллекта: парадигмальные представления и стратегии концептуализации / Г. В. Вышедко // Весці Нацыянальнай акадэміі навук Беларусі. Серыя гуманітарных навук. – 2026. – Т. 71, № 4. – С. 271–282. https://doi.org/10.29235/2524-2369-2026-71-4-271-282

Gleb V. Vyshedko

Institute of Philosophy of the National Academy of Sciences of Belarus, Minsk, Belarus

SOCIO-ONTOLOGICAL STATUS OF ARTIFICIAL INTELLIGENCE: PARADIGMATIC REPRESENTATIONS AND STRATEGIES OF CONCEPTUALIZATION

Abstract. Artificial intelligence (hereinafter AI) is one of the defining factors of the current stage of social development. The implementation of AI technologies is accompanied by unevenness and unpredictability in outcomes, which brings to the fore the problem of integrating machine intelligence into sociocultural processes. Addressing this challenge is complicated by the polysemy of the very concept of artificial intelligence. Contemporary discourse distinguishes two main approaches to its definition: the ontological approach (addressing the question of the essence of AI) and the functional-utilitarian approach (addressing the question of AI's capabilities). An analysis of these approaches reveals a correlation between human perception of AI and the nature of its impact on social processes. The article argues that under conditions of growing complexity and opacity of neural network models, AI systems acquire operational autonomy – the ability to achieve specified goals in ways that are unpredictable to humans. This phenomenon underscores the significance of the performativity of AI conceptualizations: our theoretical understandings of its nature do not merely describe reality but actively shape it, guiding public expectations and practical actions. These arguments substantiate the need for further research into the socio-ontological status of artificial intelligence.

Keywords: artificial intelligence, neural networks, society, ontological approach, functional-utilitarian approach, operational autonomy, performativity, technology

For citation: Vyshedko G. V. Socio-ontological status of artificial intelligence: paradigmatic representations and strategies of conceptualization. *Vestsi Natsyyanal'nai akademii navuk Belarusi. Seryia humanitarnykh navuk = Proceedings of the National Academy of Sciences of Belarus. Humanitarian Series,* 2026, vol. 71, no. 4, pp. 271–282 (in Russian). https://doi.org/10.29235/2524-2369-2026-71-4-271-282

© Вышедко Г. В., 2026

Введение. В современном научном дискурсе все чаще обосновывается тезис, что человечество вступило в фазу фундаментальной цивилизационной трансформации. Определяющими факторами данного этапа стали возникновение и стремительное развитие технологий, базирующихся на искусственном интеллекте (ИИ).

Первоначально ИИ рассматривался исключительно как инструментальный помощник человека, призванный оптимизировать его жизнедеятельность, автоматизировать ресурсоемкие процессы и расширить когнитивные возможности индивида. Однако по мере своей эволюции данный технологический комплекс трансформировался в мощный катализатор, раскрывший глубинные противоречия современного социума. Указанные противоречия проявляются в турбулентности социальных процессов, неравномерности динамики общественного развития и антропологических изменениях. Значительная часть данных трансформаций детерминирована сложностями внедрения передовых технологий в сложившиеся социальные отношения.

Наиболее дискуссионный и глубокий характер данный процесс обретает при попытке интеграции достижений ИИ в саму ткань социума, где ИИ начинает взаимодействовать не просто с процессами, а с ценностями, идентичностью и самосознанием человека. Как отмечается, «интеграция ИИ в общество – это не только кратковременное изменение, но сдвиг с краткосрочным, среднесрочным и долгосрочным культурным эффектом, изменяющий то, как люди в различных областях взаимодействуют с передовыми технологиями и реагируют на их влияние» [1]. Появление искусственного «разума» не просто бросает вызов классическим антропологическим и социальным концепциям, но и инициирует процесс формирования новых мировоззренческих структур как на индивидуальном, так и на групповом уровнях. Прежние представления о сущности человека, его когнитивной уникальности и социальной роли были подвергнуты пересмотру.

Данная ситуация порождает состояние эпистемологической и аксиологической неопределенности, поскольку в обществе отсутствует консенсус относительно векторов дальнейшего развития. Перспективными направлениями преодоления этой неопределенности представляются ревизия и уточнение категориального аппарата современной философии и науки. Прояснение содержания таких понятий, как «искусственный интеллект», «сознание», «человек», «общество», является насущной теоретической задачей. Выявление сущностных характеристик, дифференцирующих человека и искусственный интеллект, позволит не только приблизиться к решению проблемы создания «мыслящей машины», но и выявить глубинные причины кризисных явлений в современной цивилизации.

Закономерно, что обозначенная проблема привлекла пристальное внимание научного сообщества. В последние годы наблюдается значительное увеличение количества публикаций (научных статей, аналитических отчетов, монографий), которые, так или иначе, направлены на осмысление данного круга вопросов. Примечательно, что значительная часть этих работ фокусируется на смежной, но не тождественной проблеме: поиске оптимальной модели взаимодействия человека и искусственного интеллекта.

Среди ключевых трудов, задающих тон дискуссии, следует отметить фундаментальное исследование Н. Бострома «Искусственный интеллект. Этапы, Угрозы, Стратегии» (2014), работу Ю. Н. Харари «Homo Deus: Краткая история завтрашнего дня» (2015), а также более позднюю публикацию Э. Юдковского в журнале «Time» (2023), которая носит алармистский характер. Существенный вклад в формирование общественной повестки также вносят заявления публичных интеллектуалов и предпринимателей, среди которых можно выделить И. Маска, С. Хокинга и Б. Гейтса. Несмотря на различие в концепциях и выводах, их всех объединяет попытка дать ответ на фундаментальный вопрос: чем является искусственный интеллект для человечества и какие экзистенциальные угрозы и возможности он в себе несет? Целью настоящего исследования является выявление корреляции между способами концептуализации ИИ и его социально-онтологическим статусом. Для достижения этой цели используются методы социально-философского анализа, сравнительный анализ онтологического и функционального подходов, а также кейс-стади конкретных инцидентов.

Новизна настоящего исследования заключается в установлении корреляции между концептуальным пониманием ИИ и его социальной репрезентацией. Автор исходит из предпосылки,

что искусственный интеллект становится фактором формирования новой мировоззренческой парадигмы, задающей систему координат для современного человека. Именно смысловое наполнение, которое будет приписано системам ИИ, во многом предопределит траекторию реализации того или иного сценария будущего – от наиболее оптимистичных прогнозов до откровенно апокалиптических.

Онтологический подход и социальные проявления ИИ. Несмотря на многообразие трактовок феномена искусственного интеллекта, в современном дискурсе можно выделить две фундаментальные исследовательские позиции: онтологический и функционально-утилитарный подходы. Как было отмечено ранее, обе исследовательские стратегии применяются для концептуализации и преодоления сложностей, возникающих в процессе интеграции ИИ в общественную сферу. В данном разделе основное внимание будет сосредоточено на онтологическом подходе, предметом которого является вопрос о сущности искусственного интеллекта: что представляет собой ИИ – набор алгоритмов, автономного агента, разумного субъекта или же исключительно инструмент, расширяющий возможности человека? Ответ на этот вопрос, однако, значим не сам по себе, а в той мере, в какой он определяет социальный статус искусственного интеллекта. В рамках социально-философского анализа онтологические представления выполняют функцию смыслопорождающих практик: то, как мы понимаем природу ИИ – видим ли в нем просто инструмент, автономного агента или нечто иное, – задает систему координат для выстраивания социальных отношений, институциональных ограничений и этических оценок.

Наиболее отчетливо эта логика прослеживается в самой распространенной на сегодня трактовке, которая рассматривает ИИ как универсальный инструмент, опосредующий и усиливающий способности человека в преобразовании окружающего мира. Согласно данной позиции, ИИ уже сегодня способен создавать результаты, удовлетворяющие практическим запросам, однако в силу своей технической природы он принципиально не может выйти за пределы инструментальной функции. Будучи контролируемым, настраиваемым и корректируемым извне, машинный интеллект остается средством повышения человеческого потенциала в таких сферах, как здравоохранение, транспорт, образование и коммуникация. В этом смысле его использование аналогично применению калькулятора или компьютера: это комплексный механизм, выполняющий строго определенный набор операций. Показательно, что результаты работы нейросетей, помимо элемента стохастики, детерминированы качеством и количеством обучающих данных; изменение последних неизбежно влечет за собой трансформацию выходных параметров. Таким образом, в рамках данной парадигмы возникновение полностью автономных, самообучающихся систем, обладающих субъектностью, признается невозможным.

Научное сообщество в основном согласно, что ИИ имеет инструментальную природу, однако отсутствует единое мнение о том, какая именно сфера человеческой деятельности в наибольшей степени подвергается его воздействию. В связи с этим представляется целесообразным рассмотреть основные направления, в которых ИИ проявляет себя как значимый социальный фактор.

В последние годы ИИ все чаще становится предметом крупномасштабных экономических отношений, ассоциируясь с амбициозными проектами. Так, корпорация NVIDIA заявила о намерении вложить более 100 млрд долларов в развитие OpenAI – создателя нейросети ChatGPT, что уже получило в прессе название «сделки века» [2]. Подобные инвестиции, ориентированные на долгосрочную перспективу, маркируют наступление новой эпохи, в которой значительная доля финансовых ресурсов будет направляться на развитие технологий искусственного интеллекта. Вследствие этого ИИ неизбежно становится катализатором масштабной экономической трансформации. Как отмечает экономист MIT Д. Отор, внедрение ИИ ведет к экономической поляризации: замещая профессии средней квалификации (бухгалтеров, операторов конвейерных линий), он не способен заменить низкоквалифицированный ручной труд (уборку, уход за больными) и высококвалифицированные творческие и управленческие практики. Результатом становится феномен «исчезающего среднего класса», сопровождающийся ростом доходов узкой группы топ-специалистов и стагнацией доходов остальных категорий работников. При этом следует подчеркнуть, что ИИ не упраздняет институт труда как таковой, но трансформирует его структуру, перераспределяя занятость и доходы [3]. В отличие от промышленной революции,

где механизация резко повышала производительность труда, цифровые технологии в большей степени способствуют перераспределению богатства в пользу владельцев капитала. Впрочем, как замечает исследователь, «журналисты и даже эксперты-комментаторы, как правило, преувеличивают степень замещения машинами человеческого труда и игнорируют сильные стороны взаимодополняемости автоматизации и труда» [3, с. 6]. Подобное восприятие ИИ как управляемого ресурса возможно лишь постольку, поскольку ему приписывается онтологический статус инструмента. Именно инструментальная онтология легитимирует многомиллиардные инвестиции: капитал вкладывается не в «черный ящик» с непредсказуемой волей, а в средство повышения эффективности, подлежащее настройке и контролю.

Рассмотрение ИИ в политическом контексте неизбежно ставит вопрос о природе данных, на которых он обучается, поскольку «ИИ обучается на больших объемах данных: если данные предвзяты, то сам ИИ также будет предвзятым. Учитывая, что большинство научных исследований как исторически, так и в настоящее время проводилось представителями стран глобального Севера, а также преимущественно мужчинами, гетеросексуалами и людьми с привилегированным экономическим статусом, эти условия могут влиять на результаты исследований или на то, как они представлены» [4, с. 10]. В связи с этим активизировалась дискуссия о том, в какой степени нейросети и чат-боты могут выступать инструментом цензуры. Так, в 2023 г. корпорация Google была вынуждена принести извинения после того, как чат-бот Gemini отказывался генерировать изображения белых людей, включая исторических личностей, даже при наличии соответствующего запроса [5]. В ответ на подобные ограничения предприниматели (в частности, И. Маск с проектом GrokAI) разрабатывают альтернативные нейросети, позиционируемые как свободные от политической цензуры. Ю. Харари связывает данный феномен делегирования принятия решений алгоритмам с формированием новой формы власти, «датаизма», где алгоритмы постепенно лишают человека свободы воли и права на значимые решения в собственной жизни [6, с. 430]. В описанных случаях прослеживается трансформация онтологического статуса ИИ: от нейтрального посредника, транслирующего объективные данные, к актору, обладающему имплицитной идеологической позицией. Данный сдвиг имеет непосредственные социальные последствия. Если ИИ воспринимается как носитель ценностных установок, он перестает быть исключительно объектом технической оптимизации и превращается в предмет политической рефлексии, регулятивных ограничений и механизмов сдерживания. В этом случае его социальный статус конституируется уже не через полезность, а через потенциальную угрозу, требующую контроля.

Развивая эту линию, Ш. Зубофф рассматривает ИИ как инструмент тотального контроля в рамках теории «надзорного капитализма». Социальная реальность в информационную эпоху предстает как хаотическое информационное поле; однако, будучи помещенным в условия тотальной регистрации и прогнозирования, поведение индивидов становится программируемым: «Стремление надзорного капитализма к определенности... требует непрерывного приближения к тотальной информации... Реалити-бизнес оцифровывает всех людей, вещи и процессы, превращая их в компьютерные объекты... Гонка за тотальностью неизбежно ведет к аннексии общества, социальных отношений и ключевых общесоциальных процессов как новой территории для оцифровки, расчета, модификации и прогнозирования» [7, с. 518–519].

Способность машинного интеллекта генерировать информацию превращает его в мощное средство массовой коммуникации и производства медиаконтента. Д. Деннет предупреждал об опасности дипфейков – сфабрикованных новостей и материалов, неотличимых от подлинных. Возникает риск того, что объем синтезированной, сфабрикованной информации превысит объем подлинной, превратив медиапространство в симулякр. Более того, под угрозой оказывается само понятие человеческой идентичности, поскольку нейросети уже способны имитировать голос, изображение и поведение конкретного человека, создавая «поддельных людей»: «Эти поддельные люди – самые опасные артефакты в истории человечества, способные разрушить не только экономику, но и саму свободу человека» [8].

Еще одним значимым аспектом проблемы становится обеспокоенность исследователей тем, что повсеместное применение ИИ подрывает фундаментальные способности человека к разли-

чению достоверной и ложной информации. Данная обеспокоенность коренится не только в технологических возможностях ИИ по созданию дипфейков, но и в более глубоком обстоятельстве: генеративная способность нейросетей проблематизирует онтологический статус самой информации. Если традиционные медиа функционировали как институциональные посредники, транслирующие сообщения о событиях, то ИИ конституирует новую реальность, в которой происхождение контента – результат человеческой деятельности или алгоритмической генерации – перестает быть очевидным для получателя. Социальный статус ИИ в этом контексте определяется уже не столько его инструментальной функцией, сколько его ролью производителя реальности, что требует пересмотра оснований эпистемологического доверия.

Эмпирические исследования скорее подтверждают, нежели опровергают обоснованность этих опасений. В одном из экспериментов MIT было изучено влияние регулярного использования чат-ботов (таких как ChatGPT) на когнитивные способности [9]. Участникам предлагалось написать эссе, в то время как с помощью ЭЭГ фиксировалась активность коры головного мозга. Результаты показали значимую корреляцию между использованием внешних инструментов и уровнем когнитивной активности. Наибольшую нейронную активность продемонстрировали те, кто не использовал ни поисковые системы, ни LLM; среднюю – пользователи Google; самую низкую – регулярные пользователи LLM [9]. Особенно тревожным оказался провал в памяти: 83 % пользователей LLM не могли вспомнить и процитировать только что написанный ими текст [9]. Полученные данные актуализируют опасения относительно возможного спада критического мышления, особенно среди молодежи. Чрезмерное доверие к ответам нейросетей, замещающее практику сопоставления различных источников информации, может иметь необратимые последствия, размывая основы рациональной коммуникации и делая проблематичным различение истинной и ложной информации. Безусловно, к выводам, основанным на корреляции, следует относиться с осторожностью. Однако данное наблюдение, указывающее на потенциальную трансформацию когнитивных практик под влиянием ИИ, слишком значимо, чтобы игнорировать. Оно заставляет нас всерьез отнестись к широкому спектру прогнозов – от самых мрачных до наиболее обнадеживающих.

Именно в этом спектре и располагается оптимистический взгляд на роль ИИ, радикально расходящийся с описанными выше рисками. Р. Курцвейл, например, рассматривает искусственный интеллект как одно из важнейших изобретений в истории человечества, способное совершить скачок от цифровой революции к преобразованию физического мира: «Это самое трансформирующее обещание ИИ: более долгая, здоровая жизнь, не ограниченная дефицитом и слабостью, которые ограничивали человечество с самого начала» [10]. По его мнению, нейросети не только облегчают повседневную жизнь, но и нацелены на решение фундаментальных проблем: ускорение поиска лекарств от рака, создание дешевых и возобновляемых источников энергии и т. д. С этой точки зрения риски, связанные с ИИ, могут быть минимизированы посредством ответственного проектирования, адекватного регулирования и образования.

Переходя к собственно онтологическому измерению, необходимо поставить вопрос: что считать сознанием, интеллектом, а что – простым алгоритмом? Каковы объективные критерии для определения степени развития сознания? Как отличить деятельность сознания от работы алгоритма? Ответы на эти вопросы непосредственно влияют на восприятие статуса ИИ и, следовательно, на характер взаимодействия с ним.

Если исходить из того, что искусственный интеллект не обладает сознанием, то вся полнота ответственности за его действия возлагается на человека. В случае, если беспилотный автомобиль сбивает пешехода, вина ложится на разработчиков и отдел контроля качества; если нейросеть генерирует неэтичный контент, ответственность несут ее создатели и настройщики. Показательным примером здесь является кейс с нейросетью Sora 2: первоначально компания OpenAI не блокировала контент до возникновения претензий по авторским правам, однако затем ввела массовые блокировки [11]. Общественная реакция продемонстрировала, что традиционные механизмы правового и социального регулирования не всегда успевают за технологическим развитием: рано или поздно на рынке появляется новый игрок, менее озабоченный соблюдением авторских прав.

Если же признать возможность обретения искусственным интеллектом разума, а его сущность определить как принципиально иную, «искусственную» и потенциально превосходящую человеческую, то отношение к нему неизбежно меняется. В этом случае ИИ становится не только инструментом, но и угрозой самому существованию человечества. Как отмечает Дж. Хинтон, крестный отец ИИ, «если он не будет воспитывать меня, он заменит меня» [12]. Такая позиция предполагает не только более осторожную имплементацию систем ИИ в социальное пространство, но и формирование установки на недоверие, которая, в свою очередь, может затормозить модернизационные процессы. Таким образом, вопрос об онтологическом статусе ИИ имеет не только теоретическое, но и непосредственное практическое значение, определяя векторы развития как технологий, так и общества в целом.

Резюмируя, можно утверждать: онтологический подход в его социально-философском измерении демонстрирует, что представления о сущности ИИ не являются нейтральным теоретизированием. Они выполняют функцию смыслового каркаса, задающего параметры экономических стратегий, политических регуляций, эпистемологических практик и, в конечном счете, определяющего место человека в мире, преображенном технологиями. Вопрос «что есть ИИ?» оказывается неразрывно связан с вопросом «каким мы хотим видеть общество, сосуществующее с ИИ?». Однако, как будет показано далее, анализ одних лишь онтологических допущений недостаточен: не менее важен вопрос о том, как само использование ИИ трансформирует человеческие практики независимо от наших исходных представлений о его природе.

Функционально-утилитарный подход: ИИ как фактор трансформации человеческой реальности. Если онтологическая перспектива сосредоточена на том, чем является ИИ, то второй – функционально-утилитарный – смещает фокус на анализ возможностей, которые открываются перед человеком благодаря использованию искусственного интеллекта. В рамках данной перспективы ИИ рассматривается не столько как инструмент навязывания воли внешнему миру, сколько как способ расширения границ человеческой деятельности и познания. Как отмечал М. Хайдеггер, «мы склонны считать, что техника – это средство достижения наших целей и что человеческая деятельность находится под нашим контролем. Но на самом деле мы теперь воспринимаем средства, цели и самих себя как взаимозаменяемые и управляемые…» [13, с. 63]. Однако, согласно Хайдеггеру, «существо современной техники ставит человека на путь такого раскрытия потаенности, благодаря которому действительность повсюду, более или менее явно, делается состоящей-в-наличии… То сосредоточивающее посылание, которое впервые ставит человека на тот или иной путь раскрытия потаенности, мы называем миссией и судьбой» [14, с. 231]. Из этого следует, что техника, и ИИ в особенности, не просто служит орудием воздействия на мир, но и выступает способом его раскрытия, формирования новой картины реальности. Подобно тому как появление Интернета радикально трансформировало восприятие общества, нейросети изменяют сам характер взаимодействия человека с информацией, а значит, и с миром. И если техника действительно обладает такой конституирующей силой – способностью формировать саму оптику человеческого восприятия – то ее присутствие в обществе обретает онтологическую значимость до и независимо от любых наших теоретических квалификаций. ИИ уже сегодня активно структурирует информационную среду, задает параметры принятия решений и формирует поведенческие паттерны, тем самым de facto выступая полноправным участником социального конструирования реальности.

В этой связи вопрос о сущности и статусе машинного интеллекта отходит на второй план, уступая место проблеме его воздействия на человеческую реальность. Если ИИ уже сегодня влияет на жизнь общества и индивида, значит, он приобретает особый социальный статус независимо от наших теоретических представлений о его природе. Ключевым становится не вопрос «что нам делать с ИИ?», а вопрос «как ИИ изменяет наши убеждения, практики и способы мышления?».

Иной ракурс рассмотрения предлагает Р. Брэндом, который смещает фокус с вопроса о том, обладает ли ИИ сознанием, на анализ того, как именно он функционирует и как это соотносится с человеческими практиками. Он обращает внимание на то, что «чрезмерное внимание к символической природе мышления может уводить от более существенного вопроса – алгоритмиче-

ского характера мышления и дискурсивных практик. С точки зрения прагматического анализа главная особенность автоматов заключается не в их способности оперировать символами, а в реализации особых отношений прагматически-семантической достаточности» [15, с. 113]. Иными словами, ИИ не требуется понимать символы по-человечески; достаточно того, что он способен реализовывать функции, релевантные для социального взаимодействия.

Ключевой тезис Брэндома состоит в том, что сама человеческая разумность и социальные практики также имеют алгоритмическую природу: они могут быть разложены на последовательности базовых операций и функций. Чем сложнее действие, тем более разветвленной является соответствующая функциональная структура. Это означает, что между функционированием ИИ и человеческими дискурсивными практиками нет непроходимой пропасти – они принципиально соизмеримы по своей организации. Поскольку и те, и другие строятся на алгоритмических основаниях, открывается перспектива рассмотрения ИИ не как внешнего инструмента, а как потенциального участника социальных отношений, способного быть встроенным в существующие практики при условии их адекватной функциональной спецификации.

Однако именно алгоритмическая природа ИИ, делающая возможной его интеграцию в общество, одновременно порождает и фундаментальные риски. Наиболее радикальные сценарии развития этих рисков представлены в работах Н. Бострома и Э. Юдковского. Примечательно, что оба исследователя, размышляя о потенциальных последствиях создания сверхинтеллекта, выходят за пределы чисто функционального анализа возможностей ИИ и неизбежно возвращаются к вопросу о его природе, тем самым демонстрируя, что онтологический и функциональный подходы не исключают, а дополняют друг друга. Так, Н. Бостром, исследуя природу сверхинтеллекта, приходит к выводу, что характер его взаимодействия с человеком может поставить под угрозу существование цивилизации. В своем знаменитом мысленном эксперименте он показывает, что если наделить сверхразумный ИИ, казалось бы, безобидной задачей производить скрепки для бумаги, то «он, исходя из задачи минимизации затрат и максимизации возможностей, сначала преобразует всю Землю в скрепки, а затем и всю наблюдаемую вселенную в них» [16, с. 108]. Ключевой момент в рассуждении Бострома – проблема неверно поставленной цели: катастрофа проистекает из того, что даже нейтральный алгоритм, оптимизируя решение задачи, может интерпретировать ее разрушительным для человека образом.

Э. Юдковский идет дальше, ставя под сомнение саму возможность «нейтральной» постановки задачи. В его трактовке сверхчеловечески умный ИИ – это не просто алгоритм, который может ошибиться в интерпретации цели, а принципиально чужой субъект, само существование которого несет экзистенциальную угрозу: «Многие исследователи, занимающиеся этими вопросами, включая меня, ожидают, что наиболее вероятным результатом создания сверхчеловечески умного искусственного интеллекта при любых обстоятельствах, отдаленно похожих на нынешние, будет то, что буквально все на Земле умрут. Не как “возможно, какой-то отдаленный шанс”, а как “это очевидная вещь, которая должна была произойти”» [17].

От пессимистичных, но аналитически выверенных моделей Бострома до апокалиптических прогнозов Юдковского – весь спектр сценариев развития сверхинтеллекта с неизбежностью упирается в один и тот же вопрос: применимы ли к существу, которое может обладать разумом (или его имитацией), человеческие законы и нормы нравственности? Ответ на этот вопрос определяет возможность предотвращения описанных выше негативных исходов. Возможно, не вполне корректной, но эвристически полезной является аналогия с животными, которые также не обладают сознанием в человеческом смысле: к ним не применяются категории ответственности, вины, понимания. Ситуация существенно усложняется в случае создания общего искусственного интеллекта (AGI), сравнимого с человеческим. В этой гипотетической ситуации требуется не только, чтобы ИИ обрел сознание, но и чтобы научное сообщество, эксперты и массовые пользователи признали за ним это качество. Однако остается открытым вопрос: может ли существо считаться сознательным и ответственным, если его поведение принципиально поддается внешней настройке и коррекции? Это противоречит устоявшемуся представлению о том, что сознательное существо должно обладать автономией, собственными мотивами и побуждениями к деятельности.

Операциональная автономия ИИ: проблема контроля и непредсказуемости. В итоге выделенные онтологический и функционально-утилитарный подходы позволяют концептуализи-

ровать ключевые аспекты взаимодействия человека и искусственного интеллекта: первый проясняет вопрос о сущностных характеристиках ИИ и вытекающих из них ограничениях, второй – фиксирует многообразие социальных эффектов, порождаемых его внедрением. Вместе с тем стремительное развитие технологий машинного обучения обнаруживает аспекты, которые не получают исчерпывающего объяснения в рамках данных исследовательских перспектив. Как социально-онтологическое определение ИИ в качестве инструмента, так и анализ его внешних эффектов оставляют открытым вопрос о внутренней динамике функционирования сложных нейросетевых моделей. Современные системы ИИ демонстрируют такую степень сложности и непрозрачности собственных процессов обработки информации, которая порождает качественно новую ситуацию во взаимодействии человека и машины. Эта ситуация характеризуется расхождением между формальной подконтрольностью системы (инициируемое человеком целеполагание) и фактической непредсказуемостью конкретных способов достижения поставленных целей. Данный феномен, который можно обозначить как обретение системами ИИ «операциональной автономии», требует отдельного концептуального анализа.

Описанное выше расхождение между формальной подконтрольностью систем ИИ и фактической непредсказуемостью их работы позволяет поставить вопрос о социальном измерении машинного интеллекта. ИИ может быть квалифицирован как социальный феномен постольку, поскольку он оказывает воздействие на общественные процессы и одновременно обладает определенной значимостью для социума. Однако вопрос о мере этой значимости остается дискуссионным. В связи с этим исследователями предлагаются различные проекты гармонизации взаимодействия человека и машинного интеллекта. Сложность реализации подобных проектов обусловлена рядом факторов, среди которых ключевое место занимает приобретение системами ИИ степени «операциональной автономии». Речь в данном случае идет не о наделении машин сознанием или свободой воли, а о растущей непрозрачности и непредсказуемости их работы для человека-оператора, что создает эффект самостоятельности принимаемых решений.

Данная «операциональная автономия» обусловлена несколькими факторами. Во-первых, в алгоритмы функционирования современных чат-ботов заложен элемент случайности, делающий невозможным строгое детерминированное предсказание их ответов. Во-вторых, с каждой новой итерацией объем обучающих данных неуклонно возрастает, а поведение модели становится все более сложным для интерпретации, даже для самих разработчиков. В-третьих, архитектура глубоких нейросетей порождает эффекты эмерджентности, когда система демонстрирует способности, не заложенные явно при обучении.

В результате возникает расхождение между ожидаемыми и реальными результатами деятельности ИИ. Человек остается необходимым условием функционирования нейросетей на этапе постановки задачи: именно он определяет массивы данных и целевые параметры. Однако говорить о полном контроле над системами ИИ на выходе уже не представляется возможным. Опыт взаимодействия с нейросетями варьируется от пользователя к пользователю: вероятность получения строго идентичного ответа двумя разными пользователями крайне мала в силу стохастической природы генерации контента. Но эта вариативность не является чисто случайной – она подчиняется обученным закономерностям, которые, из-за сложности модели, не поддаются прямой интерпретации. Генерируемые нейросетями ответы могут иметь далеко идущие последствия в различных сферах – от политики до экономики, причем эти последствия не всегда совпадают с ожиданиями разработчиков. Тем самым ИИ выступает фактором, способным направлять общественное развитие по иному руслу, автономному от изначальных человеческих интенций.

Рассмотренные в предыдущем разделе социальные эффекты ИИ – экономическая поляризация, политическая ангажированность, эпистемологический кризис – предстают в новом свете, будучи осмыслены сквозь призму операциональной автономии. Именно непредсказуемость работы ИИ делает эти эффекты труднорегулируемыми и потенциально опасными.

Показательным примером непредсказуемости, коренящейся не в стохастике, а в сложности модели, стал инцидент с ChatGPT в конце 2023 г. Пользователи по всему миру столкнулись с тем, что чат-бот начал «лениться»: отказывался выполнять задачи и генерировал лишь фрагменты кода вместо полных решений. 8 декабря 2023 г. компания OpenAI официально признала

проблему, заявив: «Мы не обновляли модель с 11 ноября, и это, конечно, не было сделано намеренно. Поведение модели может быть непредсказуемым» [18]. В развернутом комментарии представители OpenAI пояснили, что обучение больших языковых моделей принципиально отличается от стандартизированного промышленного производства: даже при использовании одних и тех же данных разные обучающие запуски могут порождать модели, заметно различающиеся по стилю письма, характеру отказов от ответов и даже политической окраске генерируемых текстов [18]. Этот случай наглядно демонстрирует феномен операциональной автономии: при формальной подконтрольности и неизменном коде поведение системы оказалось непредсказуемым для ее создателей.

В этой связи особую актуальность приобретает задача не только прогнозирования ответов, но и разработки методов интерпретируемого моделирования результатов деятельности ИИ. Ключевой задачей социально-философской рефлексии на современном этапе должен стать не поиск иллюзорного тотального контроля над ИИ, а выработка эпистемологических и практических стратегий жизни в условиях «контролируемой непредсказуемости», которые позволят обществу использовать потенциал новых технологий, минимизируя экзистенциальные риски утраты контроля над ними.

Заключение. Проведенный анализ показывает, что в современном дискурсе сформировались две основные стратегии осмысления феномена искусственного интеллекта. При всех различиях их объединяет общая исследовательская интенция: рассмотрение ИИ не как изолированного объекта («вещи-в-себе»), но как элемента, встроенного в сложившиеся социальные коммуникации и практики. Центральным для обеих позиций остается вопрос о выработке конструктивной линии взаимодействия человека и машинного интеллекта – от полного отказа от технологии до ее тотального внедрения.

Онтологический подход фокусируется на проблеме сущности ИИ, исходя из предпосылки, что постижение этой сущности позволит сформировать адекватное отношение к ней. Однако консенсус относительно содержания данной сущности отсутствует: одни исследователи определяют ее через внутреннее устройство ИИ, другие – через социальный статус, который он приобретает. Наиболее распространенным остается понимание ИИ как инструмента, однако контексты его применения оцениваются по-разному. Оптимистическая позиция (Р. Курцвейл) акцентирует практический потенциал технологии; критическая (Ш. Зубофф, Д. Отор, Ю. Харари, Д. Деннет) указывает на риски политического контроля, углубления экономического неравенства и деструктивного информационного воздействия. В рамках этой перспективы положительный эффект от внедрения ИИ перекрывается вызовами, которые представляются трудноразрешимыми, особенно если исходить из понимания ИИ как «чуждого» человеку агента.

Функционально-утилитарный подход смещает фокус с вопроса «что есть ИИ?» на вопрос «что ИИ делает и как меняет наши практики?». Техника, включая системы ИИ, трансформирует способы познания мира, производства информации и социального взаимодействия. Это позволяет утверждать, что ИИ, независимо от его онтологического статуса, уже приобрел реальное социальное измерение. Соответственно, риски его внедрения коренятся не столько в его «природе», сколько в характере его использования человеком. Если справедливо предположение Р. Брэндома о том, что машинный интеллект функционирует в рамках прагматического дискурса, то для его конструктивного использования необходимо выстроить прозрачные механизмы социального взаимодействия, одинаково интерпретируемые человеком и системой.

Представляется продуктивным дополнить рассмотренные подходы еще одной перспективой – анализом ИИ как фактора фундаментальной трансформации картины мира человека. Как было показано ранее, современные системы машинного обучения обретают «операциональную автономию»: при формальном сохранении контроля со стороны человека их функционирование становится все более непрозрачным, а результаты деятельности – трудно прогнозируемыми. Эта непредсказуемость, усиливающаяся отсутствием единой модели понимания ИИ в рассмотренных парадигмах, создает ситуацию принципиальной неопределенности.

Именно в этой ситуации неопределенности особую значимость приобретает то, каким образом концептуализируется искусственный интеллект. Поскольку объективные критерии его сущ-

ности остаются дискуссионными, а его операциональная автономия делает поведение систем лишь частично предсказуемым, наши теоретические представления об ИИ начинают выполнять особую функцию. Они не просто описывают реальность, но и активно формируют ее, задавая векторы общественных ожиданий и практических действий. Следует, однако, отметить, что сам тезис о перформативности концептуализаций ИИ не является ценностно-нейтральным. Принимая во внимание традиции критической теории (М. Хоркхаймер, Т. Адорно), необходимо признать, что любое теоретическое описание социальной реальности неизбежно включено и оказывает на нее обратное воздействие. Вместе с тем результат настоящего исследования состоит не в подтверждении этого общего тезиса, а в выявлении того, какую специфическую форму он приобретает в случае с искусственным интеллектом. Как было отмечено выше, перформативность концептуализаций ИИ усиливается за счет «операциональной автономии» технологии, что создает принципиально новую ситуацию обратной связи между теорией и объектом. Эта специфика наиболее отчетливо проявляется в том, как различные концептуализации ИИ порождают конкретные социальные стратегии и практики. Эти тактики можно продемонстрировать на примере двух полярных позиций, уже рассмотренных в основной части исследования. Пессимистическая концептуализация ИИ как «чуждого» и «враждебного» агента способна порождать стратегии тотального контроля и блокировки, которые, в свою очередь, могут спровоцировать технологическое отставание и упущенные возможности. Напротив, некритически-оптимистическое восприятие ИИ как универсального инструмента решения всех проблем чревато неоправданным делегированием ответственности и атрофией человеческих навыков, что проблематизируется, в частности, в исследовании MIT.

В результате задача философии – сделать эту перформативность предметом рефлексии, показав, что выбор той или иной парадигмы понимания ИИ есть одновременно и выбор сценария будущего. Осмысление данной проблемы осложняется комплексной природой самого феномена ИИ. Расхождение между ожидаемым и реальным результатом порождает онтологическую и эпистемологическую неопределенность, которая коренится в операциональной автономии ИИ. Последняя, делая функционирование систем лишь частично предсказуемым, создает разрыв, который существующие подходы не в состоянии преодолеть в силу отсутствия единой модели понимания.

Таким образом, человечество оказывается перед выбором: принять риски технологического развития или отказаться от него. Этот выбор не имеет однозначно рационального обоснования в рамках существующих подходов. Выявление онтологических (что есть ИИ?), антропологических (что есть человек?) и этических (каковы пределы ответственности?) предпосылок является необходимым условием для выработки адекватных стратегий взаимодействия. Только на этой основе общество сможет не только минимизировать риски, но и реализовать потенциал человека в условиях технологической трансформации.

Благодарности. Исследование выполнено в рамках реализации НИР «Технологии искусственного интеллекта и мировоззренческие структуры: ключевые аспекты взаимовлияния» при поддержке Белорусского республиканского фонда фундаментальных исследований (договор № Г25ИИ М-014 от 2 мая 2025 г.), научный руководитель – И. К. Ставровский.

Acknowledgements. This article was prepared as part of the research project “Artificial intelligence technologies and worldview structures: key aspects of mutual influence” with the support of the Belarusian Republican Foundation for Fundamental Research (agreement no. G25AI M-014 dated May 2, 2025). Supervisor – I. K. Stavrovsky.

Список использованных источников

1. Chakraborty, S. AI and society – a case study on positive social change / S. Chakraborty // thechUK. – URL: https://www.techuk.org/resource/ai-and-society-a-case-study-on-positive-social-change.html (date of access: 10.10.2025).

2. Союз Nvidia и OpenAI, бессмертие и страхи инвесторов: главные события мировых рынков // Forbes. – URL: https://www.forbes.ru/investicii/546713-souz-nvidia-i-openai-bessmertie-i-strahi-investorov-glavnye-sobytia-mirovyh-rynkov (дата обращения: 10.10.2025).

3. Autor, D. H. Why are there still so many Jobs? The history and future of workplace automation / D. H. Autor // Journal of Economic Perspectives. – 2015. – Vol. 29, № 3. – P. 3–30. https://doi.org/10.1257/jep.29.3.3

4. Vera Hoyos, C. The use of artificial intelligence in political decision-making / C. Vera Hoyos, W. O. Cárdenas Marín // Philosophie. – 2025. – Vol. 10, № 5. – Art. 95. https://doi.org/10.3390/philosophies10050095

5. Lanum, N. Google apologizes after new Gemini AI refuses to show pictures, achievements of White people / N. Lanum // FOXBusiness. – URL: https://www.foxbusiness.com/media/google-apologizes-new-gemini-ai-refuses-show-pictures-achievements-white-people (date of access: 10.10.2025).

6. Харари, Ю. Н. Homo Deus: краткая история будущего / Ю. Н. Харари; [пер. с англ. А. Андреева]. – М.: Синдбад, 2018. – 494 с.

7. Зубофф, Ш. Эпоха надзорного капитализма: битва за человеческое будущее на новых рубежах власти / Ш. Зубофф; пер. с англ. А. Ф. Васильева ; под науч. ред. Я. Охонько, А. Смирнова. – М.: Изд-во Ин-та Гайдара, 2022. – 782 с.

8. Dennett, D. The problem with counterfeit people / D. Dennett // The Atlantic. – URL: https://www.theatlantic.com/technology/archive/2023/05/problem-counterfeit-people/674075/ (date of access: 10.10.2025).

9. Your Brain on ChatGPT: accumulation of cognitive debt when using an ai assistant for essay writing task / N. Kosmyna, E. Hauptmann, Y. T. Yuan [et al.]. – [Ithaca]: Cornell Univ., 2025. – 216 p. – (Preprint / Cornell Univ.; arXiv:2506.08872). https://doi.org/10.48550/arXiv.2506.08872

10. Ray Kurzweil on how AI will transform the physical world // The Economist. – URL: https://www.economist.com/by-invitation/2024/06/17/ray-kurzweil-on-how-ai-will-transform-the-physical-world/ (date of access: 10.10.2025).

11. Altman, S. Sora update #1 / S. Altman // Blog.samaltman. – URL: https://blog.samaltman.com/ (date of access: 10.10.2025).

12. Egan, M. The «godfather of AI» reveals the only way humanity can survive superintelligent AI / M. Egan // CNN. – URL: https://edition.cnn.com/2025/08/13/tech/ai-geoffrey-hinton (date of access: 10.10.2025).

13. Blitz, M. Understanding Heidegger on technology / M. Blitz // The New Atlantis. – 2014. – № 41. – P. 63–80.

14. Хайдеггер, М. Бытие и время: ст. и выступления / М. Хайдеггер. – М.: Республика, 1993. – 447 с.

15. Кустов, С. А. Функционалистский и феноменологический подходы к проблеме искусственного интеллекта: сравнительный анализ концепций Р. Брэндома и Х. Дрейфуса в контексте развития современных технологий / С. А. Кустов // Общество: философия, история, культура. – 2025. – № 2. – С. 111–118. https://doi.org/10.24158/fik.2025.2.15

16. Bostrom, N. Superintelligence: paths, dangers, strategies / N. Bostrom. – Oxford: Oxford Univ. Press, 2014. – XXII, 352 p.

17. Yudkowsky, E. Pausing AI Developments Isn't Enough. We Need to Shut it All Down / E. Yudkowsky // Time. – URL: https://time.com/6266923/ai-eliezer-yudkowsky-open-letter-not-enough/ (date of access: 10.10.2025).

18. Edwards, B. As ChatGPT gets «lazy», people test «winter break hypothesis» as the cause / B. Edwards // Ars Technica. – URL: https://arstechnica.com/information-technology/2023/12/is-chatgpt-becoming-lazier-because-its-december-people-run-tests-to-find-out/ (date of access: 09.01.2026).

References

1. Chakraborty S. AI and society – a case study on positive social change. *thechUK*. Available at: https://www.techuk.org/resource/ai-and-society-a-case-study-on-positive-social-change.html (accessed 10.10.2025).

2. The alliance of Nvidia and OpenAI, immortality and investor fears: main events of global markets. *Forbes.* Available at: https://www.forbes.ru/investicii/546713-souz-nvidia-i-openai-bessmertie-i-strahi-investorov-glavnye-sobytia-mirovyh-rynkov (accessed 10.10.2025) (in Russian).

3. Autor D. H. Why are there still so many Jobs? The history and future of workplace automation. *Journal of Economic Perspectives,* 2015, vol. 29, no. 3, pp. 3–30. https://doi.org/10.1257/jep.29.3.3

4. Vera Hoyos C., Cárdenas Marín W. O. The use of artificial intelligence in political decision-making. *Philosophie*, 2025, vol. 10, no. 5, art. 95. https://doi.org/10.3390/philosophies10050095

5. Lanum N. Google apologizes after new Gemini AI refuses to show pictures, achievements of White people. *FOXBusiness.* Available at: https://www.foxbusiness.com/media/google-apologizes-new-gemini-ai-refuses-show-pictures-achievements-white-people (accessed 10.10.2025).

6. Harari Y. N. *Homo Deus: a brief history of tomorrow.* New York, HarperCollins, 2017. 449 p.

7. Zuboff S. *The age of surveillance capitalism*. New York, PublicAffairs, 2019. 704 p.

8. Dennett D. The problem with counterfeit people. *The Atlantic.* Available at: https://www.theatlantic.com/technology/archive/2023/05/problem-counterfeit-people/674075/ (accessed 10.10.2025).

9. Kosmyna N., Hauptmann E., Yuan Y. T., Situ J., Liao X. H., Beresnitzky A. V., Maes P. *Your Brain on ChatGPT: accumulation of cognitive debt when using an ai assistant for essay writing task.* Preprint, arXiv:2506.08872. Ithaca, Cornell University, 2025. 216 p. https:/doi.org//10.48550/arXiv.2506.08872

10. Ray Kurzweil on how AI will transform the physical world. *The Economist.* Available at: https://www.economist.com/by-invitation/2024/06/17/ray-kurzweil-on-how-ai-will-transform-the-physical-world/ (accessed 10.10.2025).

11. Altman, S. Sora update #1. *Blog.samaltman*. Available at: https://blog.samaltman.com/ (accessed 10.10.2025).

12. Egan M. The «godfather of AI» reveals the only way humanity can survive superintelligent AI. *CNN*. Available at: https://edition.cnn.com/2025/08/13/tech/ai-geoffrey-hinton (accessed 10.10.2025).

13. Blitz M. Understanding Heidegger on Technology. *The New Atlantis,* 2014, no. 41, pp. 63–80.

14. Heidegger M. *Being and Time: articles and speeches.* Moscow, Respublika Publ., 1993. 447 p. (in Russian).

15. Kustov S. A. Functionalist and phenomenological approaches to the problem of artificial intelligence: comparative analysis of the concepts of R. Brandom and H. Dreyfus in the context of technological development. *Obshchestvo: filosofiya, istoriya, kul'tura = Society: Philosophy, History, Culture,* 2025, no. 2, pp. 111–118 (in Russian). https://doi.org/10.24158/fik.2025.2.15

16. Bostrom N. *Superintelligence: paths, dangers, strategies.* Oxford, Oxford University Press, 2014. XXII, 352 p.

17. Yudkowsky E. Pausing AI developments isn't enough. We need to shut it all down. *TIME*. Available at: https://time.com/6266923/ai-eliezer-yudkowsky-open-letter-not-enough/ (accessed 10.10.2025).

18. Edwards B. As ChatGPT gets «lazy», people test «winter break hypothesis» as the cause. *Ars Technica*. Available at: https://arstechnica.com/information-technology/2023/12/is-chatgpt-becoming-lazier-because-its-december-people-run-tests-to-find-out/ (accessed 09.01.2026).

Информация об авторе

Вышедко Глеб Вячеславович – аспирант. Институт философии, Национальная академия наук Беларуси (ул. Сурганова, 1, корп. 2, 220072, Минск, Республика Беларусь). E-mail: gleb.vv7@gmail.com

Information about the author

Gleb V. Vyacheslavovich – Postgraduate student. Institute of Philosophy of the National Academy of Sciences of Belarus (1 Surganova Str., Bldg 2, Minsk 220072, Belarus). E-mail: gleb.vv7@gmail.com