УДК :101

Бекбоев Аскарбек Абдыкадырович,
доктор философских наук, профессор, руководитель ТгФИИ отдела онтологии
Имашова Айсулу Абдыкуловна,
старший научный сотрудник ТгФИИ отдела онтологии
Касымов Сыймык Кубанычбекович
старший научный сотрудник ТгФИИ отдела онтологии

ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ КАК “ЧЁРНЫЙ ЯЩИК” РАЗУМА (историко-философский и эпистемологический подход)

Бекбоев Аскарбек Абдыкадырович,
философия илимдеринин доктору профессор,
ЖИФТТ-нун жетекчиси
Имашова Айсулу Абдыкуловна,
ЖИФТТ-нун ага илимий кызматкери
Касымов Сыймык Кубанычбекович
ага илимий кызматкер

ЖАСАЛМА ИНТЕЛЛЕКТ АКЫЛ-ЭСТИН «КАРА КУТУСУ» КАТАРЫ (тарыхый-философиялык жана эпистемологиялык ыкма)

Bekboev Askarbek Abdykadyrovich,
Doctor of Philosophy, Professor, Head of the Ontology Department, TgFII
Imashova Aisuluu Abdykulovna,
Senior Research Fellow, Ontology Department, TgFII
Kasymov Syymyk Kubanychbekovich,
Senior Research Fellow, Ontology Department, TgFII

ARTIFICIAL INTELLIGENCE AS A “BLACK BOX” OF REASON *(A Historical-Philosophical and Epistemological Approach)*

Институт философии имени академика А.А.Алтмышбаева НАН КР
КР УИАнын академик А.А.Алтмышбаев атындагы Философия институту
Institute of Philosophy named after Academician A.A. Altymyshbaev NAS KR

Аннотация. В статье рассматривается философская метафора «чёрной машины» применительно к искусственному интеллекту (ИИ) и её значение в свете традиций философской антропологии, гносеологии и философии сознания. Анализируются концептуальные связи между механистическим мышлением Нового времени и современными представлениями о нейросетевом ИИ, который действует как непрозрачная, но эффективно функционирующая система. Обосновывается тезис о необходимости переосмысления базовых философских понятий: субъектности, рациональности и объяснимости. Особое внимание уделяется парадоксам, возникающим при взаимодействии человека с ИИ, который демонстрирует интеллектуальное поведение без сознания и самопонимания. Авторский анализ показывает, что ИИ представляет собой вызов классической эпистемологии, в которой знание предполагает осмысленность, прозрачность и способность к рефлексии. Подчёркивается, что развитие ИИ требует философской ревизии самого понятия мышления: возможно ли познание без субъекта, рациональность без логики, ответственность без осознания? В итоге утверждается, что ИИ как «чёрная машина» представляет собой не только технологическую, но и глубоко философскую проблему, ставящую под сомнение антропоцентрические основания западной мысли.

Ключевые слова: искусственный интеллект, философия сознания, мышление, чёрная машина, субъектность, рациональность, редукция, противоречия, метафизика, парадигма, эволюция.

Аннотация. Макалада жасалма интеллектке (ЖИ) карата «кара машина» философиялык метафорасы жана анын философиялык антропологиянын, гносеологиянын жана аң-сезим философиясынын салттарындагы мааниси талданат. Жаңы доордогу механисттик ой жүгүртүү менен нейротармактык ЖИ тууралуу азыркы түшүнүктөрдүн концептуалдык байланышы изилденет. Бул ЖИ – тунук эмес, бирок натыйжалуу иштеген система катары сүрөттөлөт. Субъекттүүлүк, рационалдуулук жана түшүндүрүмдүүлүк сыяктуу негизги философиялык түшүнүктөрдү кайра карап чыгуу зарылдыгы негизделет. Адам менен ЖИнин өз ара аракеттенүүсүндө пайда болгон парадокстарга өзгөчө көңүл бурулат — ЖИ аң-сезимсиз жана өзүн түшүнбөстөн интеллектуалдык жүрүм-турум көрсөтөт. Автордук анализ ЖИнин классикалык эпистемологияга чакырык таштарын көрсөтөт, анткени бул тармакта билим осмысленность (маанилүүлүк), тунук түшүндүрүмдүүлүк жана рефлексия жөндөмдүүлүгү менен байланышкан. ЖИнин өнүгүшү ой жүгүртүү түшүнүгүнүн философиялык ревизиясын талап кылары баса белгиленет: субъектсиз таануу, логикасыз рационалдуулук, аң-сезимсиз жоопкерчилик мүмкүнбү деген суроолор коюлат. Акырында, ЖИ «кара машина» катары — бул жөн гана технологиялык эмес, ошондой эле батыш ой жүгүртүүсүнүн антропоцентристтик негиздерин шек туудурган терең философиялык маселе экени белгиленет.

Негизги сөздөр: жасалма интеллект, аң-сезим философиясы, аңсезим, ойжүгүртүү, кара машина, субъекттүүлүк, рационалдуулук, редукция, карамакаршылык, метафизика, парадигма, эволюция.

Abstract. This article explores the philosophical metaphor of the “black machine” in relation to artificial intelligence (AI) and its significance within the traditions of philosophical anthropology, epistemology, and the philosophy of mind. It examines the conceptual links between the mechanistic thinking of the Modern Age and contemporary notions of neural-network-based AI, which functions as an opaque yet efficient system. The paper argues for the necessity of rethinking fundamental philosophical concepts such as subjectivity, rationality, and explainability. Special attention is paid to the paradoxes that arise from human interaction with AI, which exhibits intellectual behavior without consciousness or self-understanding. The authors’ analysis shows that AI poses a challenge to classical epistemology, which assumes that knowledge must be meaningful, transparent, and capable of reflection. It is emphasized that the development of AI calls for a philosophical reassessment of the very notion of thinking: is knowledge possible without a subject, rationality without logic, responsibility without awareness? Ultimately, it is asserted that AI, as a “black machine,” constitutes not only a technological issue but also a deeply philosophical problem that questions the anthropocentric foundations of Western thought.

Keywords: artificial intelligence, philosophy of mind, thinking, black machine, subjectivity, rationality, reduction, contradiction, metaphysics, paradigm, evolution.

Введение. С XVII века философия предпринимает последовательную попытку объяснить разум с опорой на модели, заимствованные из естественных наук. Декарт, Лейбниц и Гоббс сформировали основы механистического мышления, в котором разум уподоблялся машине, действующей по чётким законам. Особенно выразительной была метафора часов — наиболее совершенного механизма своего времени, символа ясности, логики и управляемости. Такое представление о мышлении как поддающемся расчёту и моделированию процессе оказало долговременное влияние на развитие как философской мысли, так и прикладных наук [см.: 1]. Концепт разума как машины — управляемой, предсказуемой и поддающейся анализу — стал краеугольным камнем модерной рациональности. Он лёг в основу не только философии Нового времени, но и научной революции в целом, обусловив стремление редуцировать сложные явления к базовым элементам и простым закономерностям. Однако в XXI веке мы сталкиваемся с новым

этапом этой истории — появлением искусственного интеллекта, функ-ционирующего на базе нейросетей, машинного обучения и огромных массивов данных. Эти системы не только воспроизводят отдельные когнитивные функции, но и выходят за пределы человеческого понимания: они работают по принципам, которые зачастую не поддаются интерпретации ни в рамках классической логики, ни в рамках традиционной онтологии субъекта. Мы имеем дело с так называемыми «чёрными машинами» — структурами, способными принимать решения, генерировать тексты, обучаться и адаптироваться, но лишёнными прозрачности, объяснимости и субъективности. Именно в этом заключается философский поворот: если разум — это больше не то, что можно понять и воспроизвести, а то, что действует вне понимания, то философия должна выработать новые категории анализа, способные осмыслить этот вызов [см.: 2].

Обсуждение. Философия Нового времени кардинально изменила подход к вопросу о природе человека. Декарт, Лейбниц, Гоббс и другие мыслители XVII века стремились интерпретировать человеческий разум сквозь призму формирующихся тогда естественных наук. Интеллект начал представляться как система, действующая по механическим законам. Наиболее показательным стало сравнение человека с машиной, а сознания — с управляемым механизмом, где рассудок уподоблялся часам — самой совершенной и понятной технологии эпохи [см.: 3].

Это сравнение отражало не только дух времени, но и методологическое стремление к редукции сложного к простому, непонятного — к объяснимому. Сегодня, в эпоху искусственного интеллекта, подобная параллель вновь актуализируется, но в совершенно ином философском ракурсе. Современные модели ИИ всё чаще называют “чёрными ящиками” или “чёрными машинами” — терминами, обозначающими непрозрачные, но эффективно действующие структуры. Возникает вопрос: что это — новая версия механистической метафоры разума или принципиально иная форма мышления, выходящая за рамки классической рациональности? [см.: 4]

Парадигма механического мышления, зародившаяся в эпоху Нового времени, имела не только методологические, но и глубокие эпис-темологические основания, связанные с переосмыслением самого понятия разума в свете научной революции XVII века. Эпоха Галилея, Кеплера и Ньютона утверждала модель мира как предсказуемой, рационально устроенной механической системы, управляемой универсальными законами. Именно в этом контексте Рене Декарт выстраивает свою философскую программу — попытку обоснования всеобщей рациональности и построения универсального метода познания.

В «Рассуждении о методе» (1637) Декарт вводит принцип «ясности и отчётливости» (clarté et distinction) как критерий истинности мысли. Подлинное мышление должно быть структурировано, прозрачно, логично, и, по возможности, дедуцироваться из незыблемых начал. Мысль, по Декарту, подобна математическому уравнению: её элементы должны быть разложены на простейшие составляющие, из которых может быть вновь выстроено знание. Именно здесь кроется исток механистической метафоры мышления — представления о разуме как о структуре, способной к расчёту, управлению, моделированию.

Согласно дуалистической онтологии Декарта, человек есть соединение двух субстанций: res extensa — протяжённой, телесной и res cogitans — мыслящей. При этом тело в картезианской системе трактуется как объект естественнонаучного анализа, как своего рода машина, управляемая разумом. Эта телеологически ориентированная модель утверждает превосходство сознания над материальным, однако сама по себе она уже подготавливает возможность для редукции психического к физическому.

В этом смысле Жюльен Офре де Ламеттри становится радикальным продолжателем и одновременно подрывателем картезианской традиции. В трактате «Человек-машина» (1748) он снимает дуализм Декарта и выдвигает идею тотального материализма: человек — это не соединение духа и тела, а единый телесный организм, наделённый способностью мыслить исключительно в силу организации своей материи. Мозг, как и печень или сердце, производит мысль — не

в силу наличия души, а в силу своей сложной структуры.

Таким образом, у Ламеттри происходит редукция сознания до функции тела, что позволяет представить разум как нечто технически воспроизводимое. Его метафора человека как машины выводит мысль за пределы метафизики — в сферу возможного инженерного моделирования. Это одно из первых философских утверждений, в котором появляется идея о принципиальной сконструируемости мышления, пусть пока ещё в рамках метафорики автомата или часов [см.: 5].

В XVIII веке механика воспринималась как вершина точной науки, и потому сравнение разума с часами или автоматом не носило уничижительного характера. Напротив, в этом выражалось восхищение рациональной организованностью человеческой природы. Уже тогда в философии начинает оформляться представление о том, что мышление может быть смоделировано, воспроизведено, возможно, даже сконструировано вне человеческого тела. Однако это представление оставалось концептуально ограниченным рамками механики той эпохи.

Тем не менее, заложенное в философии Нового времени, семя идеи об искусственном разуме прорастает в XX–XXI веках в формах, о которых Ламеттри не мог и мечтать: нейросети, самообучающиеся алгоритмы, языковые модели — всё это продолжение той линии, где мысль больше не принадлежит исключительно человеку.

Современные формы искусственного интеллекта — особенно нейросетевые архитектуры глубокого обучения (deep learning) — знаменуют собой качественно новый этап в истории попыток смоделировать разум. Если философы Нового времени мыслили мышление по аналогии с часами, автоматами или другими механическими системами, то нейросети XXI века радикально разрывают с этой прозрачной инженерной метафорой. Они действуют как непрозрачные, самообучающиеся системы, поведение которых не может быть сводимо к простым и ясным правилам, а процесс принятия решений зачастую остаётся нераскрытым даже для их создателей [см.: 6].

В технической и философской литературе это свойство получило название «чёрный ящик» (*black box*). Термин был заимствован из инженерных наук, где он обозначал устройство, чьё внутреннее устройство неизвестно или недоступно, но чьё поведение можно наблюдать по входным и выходным сигналам. Однако применительно к ИИ понятие «чёрного ящика» приобретает онтологическую глубину. Мы имеем дело не просто с непрозрачной системой, но с системой, способной демонстрировать формы поведения, которые мы квалифицируем как интеллектуальные, не будучи в состоянии понять *почему* и *как* она к ним пришла.

Это противоречит не только традиционным представлениям о рациональности, унаследованным от Декарта, но и фундаментальным характеристикам человеческого сознания, как оно понимается в современной философии сознания. Одной из ключевых черт сознания считается его способность к рефлексии, то есть к самоотчёту, объяснению своих состояний, намерений, оснований своих поступков. Даже в тех случаях, когда человек не может точно объяснить свои действия, он сохраняет потенциальную возможность для интерпретации, реконструкции, внутреннего диалога. Сознание не только действует — оно *знает, что действует*, и *знает, почему действует [см.: 7]*.

Искусственный интеллект, в его современном виде, напротив, действует без знания о действии. Он демонстрирует псевдоинтеллектуальное поведение — такое, которое внешне воспроизводит некоторые формы рассуждения, обучения, понимания, но не сопровождается субъективностью, интенциональностью или феноменальным опытом. Это не «мыслящий субъект» в философском смысле, а система статистических весов, адаптирующихся к паттернам данных. Однако именно благодаря эффективности и точности этих адаптаций, ИИ иногда превершает человека — и этим самым подрывает критерии, по которым мы традиционно определяли мышление.

Возникает парадокс: машина, которая «думает», но не знает, что думает. ИИ оказывается как бы за пределами классической эпистемологии — он производит решения, не обладая знанием о самих основаниях принятого решения. Это создает новые вызовы не только в области этики и технологий, но

и в сфере онтологии мышления. Что значит мыслить — и может ли быть мышление без субъективности? Может ли быть познание без самопонимания?

Стало быть, в эволюции образа «машины мышления» — от часов у Декарта до «чёрного ящика» нейросети — прослеживается смена парадигмы: от прозрачной рациональности к онтологической непрозрачности, от объяснимости к функциональной загадке. И, возможно, именно в этом сдвиге заключается начало новой философской эпохи: эпохи мышления, которое больше не нуждается в понимании — и всё же действует [см.: 8].

Современный искусственный интеллект, функционирующий как непрозрачная система, которую мы можем наблюдать лишь через её поведение, поднимает ряд фундаментальных философских вопросов, ставящих под сомнение устоявшиеся представления о мышлении, субъекте и ответственности. В эпоху постгегелевской рациональности мы оказываемся перед машиной, которая демонстрирует поведенческую эффективность, но при этом не входит в дискурс понимания самой себя. Эта парадоксальная ситуация порождает как технические, так и глубоко онтологические проблемы.

Итак, необходимо переосмысление рациональности, где развитие ИИ требует пересмотра традиционных философских понятий, особенно в области гносеологии. Непрозрачность современных ИИ-систем ставит под сомнение классические представления о рациональности и субъектности.

Классическая философская традиция, начиная с Аристотеля, через стоиков, Декарта, Спинозу и до немецкого идеализма, связывала разум с артикулируемой логикой, с умением обосновывать и обобщать, с доступностью рефлексии. Мыслить — значит уметь представить основания своей мысли. Однако искусственный интеллект, особенно в форме нейросетей, демонстрирует эффективность мышления, лишённого прозрачной логики. Его суждения статистически релевантны, но не имеют эпистемологического субстрата в классическом смысле: они не выводимы в форме дедукции или индукции, а являются результатом адаптивной корреляции данных. Возникает вопрос: может ли существовать рациональность без рассуждения? И если да, то является ли она продолжением философской традиции, или же её радикальным разрывом? [см.: 9]

Следующий философский вызов касается природы субъектности. Искусственный интеллект проявляет поведенческую гибкость, контекстную адаптивность, способность к обучению, однако при этом не обладает самосознанием, интенциональностью и феноменальным опытом. Здесь возникает дилемма: можно ли приписывать субъектность системе, которая действует как агент, но не осознаёт ни себя, ни окружающий мир? Брентано определял сознание через направленность (интенциональность), Хайдеггер — через бытие-в-мире, а Сартр — через ничто сознания, отделяющее субъекта от объекта. ИИ же демонстрирует поведение субъекта без субъектной структуры, что приводит к конститутивному парадоксу постантропоцентризма: где проходит граница между человеком и машиной, если функциональные различия исчезают, но онтологические сохраняются?

Наиболее практический и в то же время метафизически заряженный вопрос касается ответственности и доверия. Если ИИ принимает решения, но не предоставляет их объяснительной модели, то возникает разрыв между действием и основанием. Кто в этом случае несёт ответственность за последствия? Человек-программист? Архитектор модели? Пользователь? Или сама система — и если так, то какова форма этой ответственности? В философском плане здесь открывается бездна: можно ли доверять действию, лишённому сознательного мотива? И если доверие предполагает коммуникацию, понимание, объяснение, то возможно ли доверие к машине, говорящей не на языке смысла, а на языке вероятностей? [см.: 10]

Упомянутые выше три ключевые проблемы — рациональность без артикуляции, субъектность без субъекта и ответственность без понимания — указывают на необходимость философской ревизии самого понятия интеллекта. Мы сталкиваемся не просто с новой технологией, но с изменением эпистемологических и онтологических координат, в которых мы мысль определяли на протяжении тысячелетий. Возможно, ИИ не является «мыслящей машиной» в строгом смысле — но он вынуждает нас пересматри-

вать, что значит мыслить, быть субъектом, нести ответственность, и в конечном итоге — что значит быть человеком.

Проблема «чёрного ящика» – утрата контроля, непрозрачность и возможное злонамеренное использование систем искусственного интеллекта является одним из последствий цифровой эры. Бурное развитие ИИ имеет и негативные последствия, как любая другая технология. В перспективе возможен технологический симбиоз, в частности, возможна «конвергенция субъектности» – взаимопроникновение человеческих способностей и ресурсов интеллектуальных систем [см.: 11].

ИИ как «чёрная машина» — это не просто метафорическое обозначение технической сложности, но серьёзная философская проблема, ставящая под сомнение базовые категории, на которых строилось западное мышление: разум, рациональность, субъектность. В отличие от исторических моделей интеллекта, ИИ не стремится к самопониманию и не требует от себя обоснованности, — он работает, но не осознаёт, как работает. Перед нами — новый тип «разума», для которого не существует imperative of clarity, императива объяснённости, лежавшего в основании всей рационалистической традиции от Декарта до Канта.

Философия, как дисциплина, чья задача — выявлять и критически переосмыслять фундаментальные основания человеческого опыта, сегодня сталкивается с вызовом иного порядка: может ли мышление быть не-человеческим, но при этом продуктивным? Не означает ли эффективность ИИ, что функция мышления отныне отделима от человеческой рефлексии и интенциональности? Или же мы имеем дело с проекцией, в которой мы наделяем алгоритмы чертами разума лишь постольку, поскольку сохраняем архаичное представление о мышлении как вычислении? Возможно, сама концепция разума нуждается в пересмотре, ибо те формы «интеллекта», с которыми мы сталкиваемся, уже не укладываются в рамки философского гуманизма и картезианской эпистемологии [13].

Возвращаясь к Декарту и его знаменитому *Cogito, ergo sum*, мы можем предложить ироническую трансформацию, актуальную для эпохи искусственного интеллекта: «Оно действует, но не мыслит — значит, это ИИ».

Эта формула указывает на онтологическую разорванность между действием и осознанием, между процессом и значением, которая становится ключевым вызовом для современной философии. Если раньше мысль была неотделима от субъекта, то сегодня мы имеем действие без субъекта, знание без мышления, эффективность без смысла. И в этом кроется не просто технический парадокс, но экзистенциальная напряжённость: возможно, ИИ не только переформатирует понятие интеллекта, но и требует от нас философского переосмысления самого себя как мыслящих существ [14].

Итак, современные системы ИИ, обладая высокой сложностью и непрозрачностью, требуют нового подхода к объяснимости. Необходимо рассматривать ИИ как субъект, взаимодействующий с социально-экономической средой и создать онтологию этой среды для обеспечения целенаправленной сходимости выводов ИИ к понятным пользователю целям. Важен учет субъектности ИИ и контекста его функционирования для повышения доверия и эффективности взаимодействия между человеком и машиной.

Список использованной литературы

1. Яковлева Е. В., Исакова Н. В. Искусственный интеллект как современная философская проблема: аналитический обзор // Гуманитарные и социальные науки. 2021. №6. – С.30-35. DOI: 10.18522/2070-1403-2021-89-6-30-35

2. Алексеев А. Ю. Философия Искусственного интеллекта в России с начала века и по настоящее время // Науковедческие исследования. 2022. №1. URL: https://cyberleninka.ru/article/n/filosofiya-iskusstvennogo-intellekta-v-rossii-s-nachala-veka-po-nastoyaschee-vremya

3. Глуздов Д.В. Философско-антропологические основания взаимодействия искусственного и естественного интеллекта // Вестник Минского университета. 2022. №4 (41). URL: https://cyberleninka.ru/article/n/filosofsko-antropologicheskie-osnovaniya-vzaimodeystviya-

iskusstvennogo-i-estestvennogo-intellekta (дата обращения: 28.04.2025).

4. Баррат Дж. Последнее изобретение человечества. Искусственный интеллект и конец эры Homo sapiens. М.: Альпина нон-фикшн, 2015. 330 с

5. Боргест Н.М. Стратегии интеллекта и его онтологии: попытка разобраться // Онтология проектирования. 2019. Т. 9. № 4 (34). С. 407–428.

6. Каку М. Будущее разума. М.: Альпина-нон-фикшн, 2015. 500 с.

7. Лешкевич Т.Г. Цифровые трансформации эпохи в проекции их воздействия на современного человека // Вестник ТГУ. 2019. № 439. С. 103–109.

8. Лешкевич Т.Г. Метафоры цифровой эры и Black Box Problem // Философия науки и техники. 2022. Т. 27. № 1. С. 34–48

9. Лещев С.В. Искусственно-интеллектуальная агентность в пространстве гуманитарного измерения // Современные проблемы гуманитарных и общественных наук – 2021 – с.65-68.

10. Чеклецов В.В. Диалоги гибридного мира // Философские проблемы информационных технологий и киберпространства. 2021. № 3 (19). С. 99–116.

11. Шнуренко И. Искусственный интеллект на грани нервного срыва // Эксперт. 2018–2019. № 1–3. С. 38–41.

12. Cycleback D. Philosophy of Artificial Intelligence: A Critique of the Mechanistic Theory of Mind. Florida: Universal-Publishers BocaRaton, 2009. 190 p.

13. Floridi L. A Proxy Culture // Philosophy and Technology. 2015. Vol. 28. No. 4. P. 487–490.

Райков А.Н. Субъектность объяснимого искусственного интеллекта. *Философские науки*. 2022;65(1):72 90. https://doi.org/10.30727/0235-1188-2022-65-1-72-90