УДК 004.8:17

Специфика искусственного интеллекта и человеческий интеллект: философские аспекты

Гилязова Ольга Сергеевна

Кандидат философских наук,
доцент Центра развития универсальных компетенций,
Уральский федеральный университет,
620062, Российская Федерация, Екатеринбург, ул. Мира, 19;
e-mail: olga_gilyazova@mail.ru

Замощанская Анна Николаевна

Преподаватель,
ведущий менеджер Центра развития универсальных компетенций,
Уральский федеральный университет,
620062, Российская Федерация, Екатеринбург, ул. Мира, 19;
e-mail: a.n.kolobaeva@urfu.ru

Аннотация

Растущие возможности искусственного интеллекта (ИИ) и его наступление на прерогативы человека вызывают беспокойное внимание к теме ИИ со стороны общественности и научного сообщества. В своей статье мы, опираясь на исследования зарубежных и отечественных философов, сфокусируемся на отдельных философских аспектах специфики ИИ сравнительно с человеческим интеллектом. Демонстрируем, как эволюция от наивно вычислительной/функционалистской концепции интеллекта к интерактивной/контекстуальной, связывает ИИ с робототехникой и заодно актуализирует проблему развития ИИ в качестве эмоционального и социального ИИ. Выявляем три основных различия между человеческим интеллектом и ИИ: способность быть эмоциональным, способность быть социальным, способность быть абдуктивным. Без преодоления этих различий ИИ не сможет стать явным, а тем более полным этическим агентом, а, следовательно, и гармоничным партнером человечеству. В заключении подчеркиваем огромный (как позитивный, так и негативный) потенциал ИИ.

Для цитирования в научных исследованиях

Гилязова О.С., Замощанская А.Н. Специфика искусственного интеллекта и человеческий интеллект: философские аспекты // Контекст и рефлексия: философия о мире и человеке. 2024. Том 13. № 12А. С. 24-32.

Ключевые слова

Искусственный агент, искусственный интеллект, робототехника, машинная этика, философия науки, этический агент.

Введение

До сих пор нет единого мнения о том, что такое интеллект, но данный факт не мешает исследователям говорить об искусственном интеллекте (термин, обязанный своим рождением Дартмутской конференции, организованной американским ученым Д. Маккарти в 1956 г.). Согласно Сообщению Европейской комиссии, дефиницией которой мы будем пользоваться в качестве рабочей, искусственный интеллект (ИИ) «относится к системам, которые демонстрируют разумное поведение, анализируя окружающую среду и предпринимая действия – с некоторой степенью автономии – для достижения конкретных целей» [European-Commission, 2018].

Ожидается, что в XXI веке ИИ приобретет то же значение, что и электричество в XX веке. Данная отрасль переходит к долгожданному рубежу – общему искусственному интеллекту (ОИИ) (Artificial General Intelligence (AGI)), который, в отличие от ограниченного («узкого») ИИ, специализирующегося на отдельных конкретных задачах, может генерализировано применяться к нескольким областям и задачам и даже составлять конкуренцию человеческому интеллекту в целом.

Именно разработка ОИИ стала амбициозной целью стартапа OpenAI, прославившегося на сегодняшний момент в области создания генеративного ИИ (см.: GPT-4, ChatGPT, DALL-E, OpenAI Five и т.д.). Данная технология, способная в ответ на короткие подсказки писать адекватные программы, прозу и стихи, рисовать картины в заданном стиле и выполнять множество промежуточных задач, связанных с искусством и творчеством, стала тревожным сигналом для программистов, писателей, художников. Опасение она вызывает и у тех, кто привык работать по алгоритму, шаблону, скриптам: юристов, терапевтов, переводчиков, трейдеров, сметчиков, копирайтеров, компиляторов и т.п. Конечно, модели генеративного ИИ пока еще не вышли на уровень автономных работников, но в отдельных отраслях уже вполне эффективно справляются с ролью ремесленников и полноценных ассистентов.

Так что если еще не так давно ученые обсуждали уязвимость рутинного труда: физического – из-за автоматизации производства, интеллектуального – из-за компьютеризации, то развитие генеративного ИИ подрывает один из последних оплотов человеческой занятости и уникальности – творческий труд.

Неудивительно, что в 2019 году Белый дом выступил с инициативой American AI Initiative, в которой сформулировано всеобъемлющее видение лидерства США в области технологий. Китай тратит все больше средств на исследования и разработки в области ИИ. Европа также увеличивает свои инвестиции. В России тоже осознают важность разработок в данной сфере [Галикеева, Фархиева, 2021].

Своеобразие современных технологий в том, что они являются зеркалом для человечества, где отражением становимся не мы сами, а наши представления о себе и мире. Развитие ИИ заставляет нас задуматься о природе собственного интеллекта, о его потенциале и границах. Это многомерная тема, занимавшая умы философов с древних времен. Мы же сосредоточимся на отдельных ее моментах.

Основная часть

С первоначальных шагов по созданию ИИ в этой области сосуществовали два основных подхода: 1) традиционный ИИ (также называемый символическим ИИ, логическим ИИ, «старым добрым ИИ» (Good Old Fashioned AI (GOFAI)) или классическим ИИ) и 2)

коннекционизм (также называемый нейронными сетями) [Jordanous, 2020].

Объединяет два данных подхода представление, что связь между разумом и мозгом напоминает взаимосвязь между управлением программой и компьютером. Философскую основу этого представления можно найти у Т. Гоббса (по которому рассуждение – это «не более чем расчет»), у Г. Лейбница (пытавшегося создать логическое исчисление всех человеческих идей), у Д. Юма (желавшего свести восприятие к «атомарным впечатлениям») и отчасти у И. Канта (по которому опыт контролируется формальными правилами). Таким образом, высшей целью ИИ было воспроизвести человеческий интеллект в вычислительной системе [Boden, 1990; Brooks, 1990].

В конце 1980-х годов несколько исследователей выступили за совершенно новый подход к ИИ, отстаивающий необходимость наличия тела для проявления настоящего интеллекта – ему нужно воспринимать, двигаться, выживать и иметь дело с миром. Этот подход позже назвали «концепцией воплощенного интеллекта». Применительно к робототехнике, воплощенный ИИ формулируется как антропомиметический принцип Пфайфера: достижение человеческого интеллекта требует человеческого опыта в восприятии и взаимодействии с окружающей средой. В расширенном варианте: «опыт собаки вызовет интеллект, подобный собачьему, опыт автомобильного автопилота, как ожидается, будет генерировать интеллект автомобиля и так далее» [Potkonjak, 2020, p. 5].

Как следствие – в обыденном сознании робот часто ассоциируется с ИИ и наоборот, хотя робот как физическая машина с датчиками и исполнительными механизмами может и не нуждаться в интеллекте для выполнения определенных задач, а ИИ как программа не обязательно должен быть физическим. Однако их взаимосвязь обогащает обоих: ИИ, благодаря воплощению в тело робота, становится все более способным воспринимать, действовать и формировать свою среду, а также участвовать в социальном взаимодействии, и, в свою очередь, ИИ позволяет роботам учиться и совершенствоваться.

Целью воплощенного ИИ стало создание интеллектуальных агентов, то есть сущностей, которые воспринимают окружающую среду и действуют в соответствии с ней. Взамен создания интеллекта «сверху вниз» (т.е. зависимого от нисходящих инструкций) усилия были направлены на создание интеллекта «снизу вверх» (адаптивного к среде) [Russell, 2016]. Это привело к изучению адаптивного поведения (в основном на основе этологии (науки о поведении животных)) и к переориентации подходов к определению интеллекта. Если традиционно преобладало *внутреннее* понимание интеллекта (в греческой философии – от Анаксагора до Платона, а затем от неоплатонизма к идеализму (как классическому, так и современному)), согласно которому интеллект внутренне присущ субъекту, независимо от контекста или взаимодействия с другими субъектами и средой, то в современных дискуссиях набирает обороты *контекстуальный* подход, согласно которому интеллект зависит от взаимодействия между субъектом и окружающей средой [Farisco, Evers, Salles, 2020].

Значимый вклад в этот подход был внесен Р. Бруксом, настаивающем на преимуществах т.н. «минимального познания», возникающего при взаимодействии с окружающей средой, с использованием относительно простых нейронных адаптивных контроллеров [Brooks, 1990]. Но не все готовы пренебрегать потенциалом высокоуровневого интеллекта. Так что одним из основных мотивов современных исследований является рассмотрение возможных принципов, алгоритмов и схем реализации, которые могут преодолеть парадокс Моравека, т.е. разрыв между низкоуровневым (включающим сенсорное восприятие, физический контроль и формирование поведенческих движений) и высокоуровневым (включающим логические

выводы, планирование и овладение языком) уровнями познания. Это позволит ИИ справляться с неопределенностью в нашей повседневной жизни.

Однако такая попытка может не достичь успеха, т.к. «эти два уровня не имеют одного и того же метрического пространства, необходимого для плотного взаимодействия между нисходящими и восходящими процессами» [Tani, White, 2022, p. 82]. Для решения этой проблемы многообещающими представляются схемы глубокого обучения. Например, обучение с подкреплением может осуществляться в рамках не одного искусственного агента (ИА), а целой их когорты: ведь связь между мозгами, такая как телепатия, и копирование изученной нейронной сети, например при трансплантации мозга, которые технически и этически сложны для людей, для них довольно просты. Ведь «в отличие от реальных форм жизни на Земле, искусственные агенты могут выполнять ламарковский способ копирования выученного поведения через Интернет с тысячами других агентов в любом месте» [Doya, Taniguchi, 2019, p. 95]. Еще одним перспективным подходом является эволюционная робототехника: Д. Тани и Д. Уайт предлагают развивать способности роботов постепенно, с переходом от уровня к уровню (в соответствии с фундаментальными теориями развития ребенка) [Tani, White, 2022].

Феномен ИИ доказывает, что интеллект (в этом его специфика) не является исключительно биологическим, а тем более узко гуманоидным (и тем подрывает самомнение вида Homo sapiens), а «выходит за рамки химической основы возникновения, будь то на основе кремния или углерода» [Boltuc, 2022, p. 9]. Переход от наивно вычислительной/функционалистской концепции интеллекта к интерактивной и контекстуальной не отменяет разницу между определением интеллекта в исследованиях ИИ и в биологических науках, подчеркивающих важность эмоциональности, т.е. «способности мозга придавать значение внутренним представлениям или внешним входным сигналам» [Farisco, Evers, Salles, 2020, p. 2417]. К тому же эмоциональная вовлеченность и сочувствие часто действуют как моральный компас и источник понимания, обеспечивая уникальный доступ к потребностям других. Конечно, ИИ может распознавать эмоциональное состояние людей, но это не значит, что он осознает эти эмоции или сам испытывает их с тем, чтобы использовать их для заботы по отношению к другим.

В этом смысле эмоциональный интеллект оказывается неразрывно связанным с социальным интеллектом. И тот, и другой у ИА пока еще тоже не дотягивают до уровня человека. Отсюда – асимметрия между способностью современных ИА решать сложные интеллектуальные задачи и их способностью двигаться, улавливать ситуативный контекст и читать намерения людей. Перед нами проявление парадокса Моравека, обусловленного тем, что, в отличие от умения решать сложные интеллектуальные задачи, физическая ловкость и умение понимать намерения окружающих развивались под влиянием естественного отбора на протяжении миллионов лет. Так что те «примитивные» навыки, которые люди принимают как нечто должное, оказывается совсем нелегко запрограммировать в ИИ. Как отмечают Т.Дж. Прескотт и Д.М. Робийярд, «этот дисбаланс, который может сбивать с толку пользователей, будет снижаться по мере улучшения контекстной чувствительности социального интеллекта роботов» [Prescott, Robillard, 2020, p. 3]. Это особенно важно для социальной робототехники как наиболее перспективной области исследований, посвященной социально компетентным роботам, основная задача которых – налаживание естественного взаимодействия с людьми.

Таким образом, возникает необходимость развивать ИИ в качестве «социального ИИ», который не должен сводиться к «искусственному социальному интеллекту». В когнитивных науках и этологии исходят из положения, что человеческий интеллект является по своей сути социальным, т. к. он возник из необходимости решать социальные проблемы и конфликты.

Показательно, что с годами исследователи отказались трактовать социальный интеллект в качестве «Макиавеллистского интеллекта», склонного к обману и манипуляциям, установив, что, в отличие от интеллекта других приматов, человеческий интеллект в бóльшей мере зависит от сотрудничества [Damiano, Dumouchel, 2018]. Человеческие социальные компетенции, а также кооперативный характер человеческого интеллекта должны стать образцом для развития аналогичных характеристик и способностей у ИИ.

Притом, лежащий в основе самых передовых ИИ взгляд на интеллект как на интерактивный процесс со средой рискует свести интеллект к реакции на воздействие окружающей среды. Такая интерпретация оспаривается современной нейробиологией, согласно которой мозг является не просто реактивным, а спонтанно, внутренне активным и способен строить свои модели окружающего мира, предвосхищающие реальный опыт [Farisco, Laureys, Evers, 2017]. Это важно в силу того, что внутренняя активность мозга обеспечивает выдающуюся человеческую способность – к абдуктивным рассуждениям, т.е. к противоречивым выводам.

Таким образом, есть три основополагающих различия между человеческим интеллектом и ИИ: 1) способность к эмоциональной вовлеченности; 2) способность быть социальным; 3) способность быть абдуктивным, т.е. интерпретировать поступающие стимулы не только на основе фактического опыта, но и в более широких рамках, которые иногда могут оправдывать противоречивые выводы.

Эти три существующих в настоящее время различия между человеческим интеллектом и ИИ актуальны с этически-практической точки зрения: без преодоления этих различий ИИ не может стать явным (а тем более полным) «этическим агентом», способным осуществлять свой моральный выбор и объяснять его [Boddington, 2017]. Более того, он окажется уязвимым для манипуляций и контекстуально наивным [Butkus, 2020].

Здесь мы следуем классификации Дж. Мура, выделяющего четыре основных типа агентов, в зависимости от их автономности в этическом смысле [Moor, 2006]:

1. *Агенты этического воздействия*. Так как их действия имеют лишь косвенное этическое воздействие, то они являются «этическими» в самом слабом смысле.

2. *Неявные этические агенты*. Эти агенты не способны к дифференциации этичного и неэтичного поведения, но их конструкция напрямую определяется соображениями безопасности или критической надежности, чтобы избежать неэтичного поведения. Это, например, мобильные банковские приложения.

3. *Явные этические агенты*. Это агенты, которые могут не только идентифицировать и обрабатывать этическую информацию о различных ситуациях, но и выносить четкие этические суждения и руководствуются ими в своем поведении.

Основная часть описываемых в литературе стратегий создания этих агентов базируется на нормативных этических теориях, таких как деонтологические (этика долга), утверждающие, что нравственное поведение должно основываться главным образом на принципах (поэтому другое их название – «принциплистские»), и телеологические (от греч. «телос» – «конец» или «цель») или иначе консеквенциалистские, подразделяющиеся на ситуационизм и утилитаризм и подчеркивающие важность возможных и ожидаемых последствий при принятии этических решений. В большинстве моральных дилемм, применяемых для разработки и проверки настоящей и будущей этики систем ИИ, выбирают между абсолютизмом первых и релятивизмом вторых.

4. *Полные этические агенты*. Помимо способности выносить этические суждения, они обладают теми особенностями (такими как «сознание, интенциональность и свобода воли» ([Moor, 2006, p. 20]), которые обычно приписываются людям.

Как видим, вопросы машинной этики актуальны именно для третьего типа этических агентов. В чем же разница между людьми и ИА как этическими агентами? Ответ на этот вопрос позволит понять, к чему должны быть способны ИА, чтобы получить право считаться полными этическими агентами. Хотя, как аргументирует Э. Фирт, их создание само по себе не гарантирует, что они будут иметь человеческую мораль [Firt, 2024].

По мнению А. Матиаса, разница – в способности человека выражать несогласие (здесь понимается как способность отказываться (на основе собственных рациональных соображений и ценностей) действовать так, как предписано внешними силами), что является маркером его способности быть самоуправляемым, автономным субъектом. Истинная моральная автономия ИА включает в себя способность ИА не подчиняться алгоритму или ценностям своего создателя и выбирать творческий и даже морально «плохой» образ действий [Matthias, 2020].

Согласно лекции Д. Эдмондса «Машины-убийцы: надо ли роботу читать Канта?», прочитанной в Москве в 2016 г., для того чтобы считать ИА полным этическим агентом, необходимо наличие у него способности к моральному выбору, а также способности иметь намерение – то есть того, без чего немыслима моральная ответственность, а значит, и моральная автономность агента. Более того, «этическая» программа, управляющая агентом, должна иметь механизм, определяющий вину [Эдмондс, 2016]. Но проблема с включением в ИИ понятий вины и ответственности необычайно сложна.

Вопрос даже не в том, какой именно этический кодекс (деонтологический, телеологический или их сочетание) встроить в ИИ. Исследования показывают, что люди хотели бы, чтобы ИИ принимал морально справедливые решения в принципе, но они также хотят, чтобы ИИ отклонился от этого морального пути, если моральное действие потребует принесения в жертву их собственной жизни или жизни членов их семьи. Более того, существуют индивидуальные и культурные различия в том, что считается морально справедливым и правильным. Например, может так выйти, что беспилотный автомобиль, сконструированный в Индии, скорее сохранит жизнь корове, чем человеку.

Даже три закона робототехники А. Азимова, несмотря на их внешнюю очевидность и универсальность, морально уязвимы: нельзя изменить эти три закона, особенно в части первого закона, но можно изменить определение понятия «человек», что актуально (и опасно) не только в свете трансгуманистического будущего [Tsvyk, Tsvyk, Pavlova, 2023], но и в свете характерного для нынешнего политического и военного противостояния расчеловечивания противников. Так что размышления об этике роботов (или шире – машинной этики) заставляют задуматься о человеческой этике.

По мнению П. Болтуса, ИИ стоит вооружать не только техническими знаниями, но и гуманитарными (не ограничиваясь этикой), без которых ИИ не сможет быть совместим с человеческим интеллектом (не сводимым к когнитивным способностям) и не сможет быть внедрен в качестве потенциального партнёра людей «в более широкую человеческую цивилизацию — культуру, ценности, традиции и смыслы, которыми мы живем, — что будет означать огромный провал в общении» [Boltuc, 2022, p. 13].

Заключение

Подводя итог, можно отметить, что развитие ИИ как один из ведущих трендов революции софта обладает огромным потенциалом как для решения, так и для создания проблем человечеству. Модели генеративного ИИ уже сейчас посягают на положение «креативного»

класса, воплощенный (обладающий телом) ИИ пытается заменить человека в помогающих профессиях (медицина, социальное обслуживание и т.п.), раздвигая пределы тех миссий, для которых он изначально создавался и которые обозначаются как 3D – dull, dirty, dangerous (скучная, грязная и опасная работа).

В любом случае, социально полезные инновации чреваты т.н. орфанными рисками, трудно измеряемыми и легко упускаемыми из виду: угроза конфиденциальности и автономности, а также неприкосновенности частной жизни (со стороны бота, собирающего и делящегося потенциально конфиденциальной информацией; робота, оснащенного передовыми системами наблюдения), расхождение между этическими или идеологическими взглядами (в борьбе с преступностью или цифровым активизмом и т.п.). Не говоря уже о более явных угрозах – сокращения рабочих мест; использования в вооружении. Отдельно стоит упомянуть проблему технологической предвзятости, в результате усвоенного от людей поведения, которое бывает довольно дискриминационным. Данная проблема не тривиальна и ее последствия не безобидны, учитывая готовность людей делегировать ИИ часть своих полномочий по принятию решений в медицине, праве, а в перспективе – и выход ИИ на арену политических решений.

Все это лишний раз подчеркивает важность рассмотрения специфики ИИ (его сходства и отличия сравнительно с человеческим интеллектом), приобретения им не только когнитивного, но и эмоционального и социального измерения с тем, чтобы в качестве явного (а в перспективе даже полного) морального агента суметь стать партнером (полезным и эффективным, словами OpenAI) человечеству и планете в целом.

Библиография

1. Галикеева Н.Н., Фархиева С.А. О национальной стратегии развития искусственного интеллекта до 2030 года в РФ и Федеральном проекте «Искусственный интеллект» // Современная школа России. Вопросы модернизации. 2021. № 3-1 (36). С. 186-188.
2. Эдмондс Д. Машины-убийцы: надо ли роботу читать Канта? URL: https://old.inliberty.ru/blog/2328-Mashinyubiycy-nado-li-robotu-chitat-Kanta
3. Boddington P. Towards a code of ethics for artificial intelligence. Cham, Switzerland: Springer, 2017. 124 p. DOI: https://doi.org/10.1007/978-3-319-60648-4
4. Boden M.A. (Ed.) The Philosophy of Artificial Intelligence. Oxford: Oxford University Press, 1990. 452 p.
5. Boltuc P. Transhumanities as the Pinnacle and a Bridge // Humanities. 2022. 11 (1): 27. DOI: https://doi.org/10.3390/h11010027
6. Brooks R.A. Elephants don´t play chess // Robotics and Autonomous Systems. 1990. 6 (1-2). Pp. 3-15.
7. Butkus M.A. The Human Side of Artificial Intelligence // Science and Engineering Ethics. 2020. Vol. 26. Pp. 2427-2437. DOI: https://doi.org/10.1007/s11948-020-00239-9
8. Damiano L., Dumouchel P. Anthropomorphism in Human-Robot Co-evolution // Frontiers in Psychology. 2018. 9: 468. DOI: 10.3389/fpsyg.2018.00468
9. Doya K., Taniguchi T. Toward evolutionary and developmental intelligence // Current Opinion in Behavioral Sciences. 2019. Vol. 29. Pp. 91-96. DOI: https://doi.org/10.1016/j.cobeha.2019.04.006
10. European-Commission. Communication from the commission. Artificial Intelligence for Europe. Brussels, 2018. URL: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM%3A2018%3A237%3AFIN
11. Farisco M., Evers K., Salles A. Towards Establishing Criteria for the Ethical Analysis of Artificial Intelligence // Science and Engineering Ethics. 2020. Vol. 26. Pp. 2413-2425. DOI: https://doi.org/10.1007/s11948-020-00238-w
12. Farisco M., Laureys S., Evers K. The intrinsic activity of the brain and its relation to levels and disorders of consciousness // Mind & Matter. 2017. Vol. 15(2). Pp. 197-219.
13. Firt E. What makes full artificial agents morally different // AI & Society. 2024. DOI: https://doi.org/10.1007/s00146-024-01867-6
14. Jordanous A. Intelligence without Representation: A Historical Perspective // Systems. 2020. 8. No. 3: 31. DOI: https://doi.org/10.3390/systems8030031
15. Matthias A. Dignity and Dissent in Humans and Non-humans // Science and Engineering Ethics. 2020. 26. Pp. 2497-2510. DOI: https://doi.org/10.1007/s11948-020-00245-x

16. Moor J.H. The nature, importance, and difficulty of machine ethics // IEEE Intelligent Systems. 2006. Vol. 21. Iss. 4. Pp. 18-21. DOI: https://doi.org/10.1109/MIS.2006.80
17. Potkonjak V. Is Artificial Man Still Far Away: Anthropomimetic Robots Versus Robomimetic Humans // Robotics. 2020. 9(3): 57. DOI: https://doi.org/10.3390/robotics9030057s
18. Prescott T.J., Robillard J.M. Are friends electric? The benefits and risks of human-robot relationships // iScience. 2020. 24(1): 101993. DOI: https://doi.org/10.1016/j.isci.2020.101993
19. Russell S. Rationality and intelligence: A brief update // Müller V.C. (Ed.). Fundamental issues of artificial intelligence. Synthese Library, vol. 376. Cham: Springer, 2016. Pp. 7-28.
20. Tani J., White J. Cognitive neurorobotics and self in the shared world, a focused review of ongoing research // Adaptive Behavior. 2022. 30(1). Pp. 81-100. DOI: 10.1177/1059712320962158
21. Tsvyk V.A., Tsvyk I.V., Pavlova T.P. The Problematic Area of Philosophical Discourses on the Application of Artificial Intelligence Systems in Society // Вестник Российского университета дружбы народов. Серия: Философия. 2023. Т. 27. № 4. С. 928-939. DOI: https://doi.org/10.22363/2313-2302-2023-27-4-928-939

The Specificity of Artificial Intelligence and Human Intelligence: Philosophical Aspects

Ol'ga S. Gilyazova

PhD in Philosophical Sciences,
Associate Professor, Center for the Development of Universal Competencies,
Ural Federal University,
620062, 19, Mira str., Yekaterinburg, Russian Federation;
e-mail: olga_gilyazova@mail.ru

Anna N. Zamoshchanskaya

Lecturer,
Leading Manager, Center for the Development of Universal Competencies,
Ural Federal University,
620062, 19, Mira str., Yekaterinburg, Russian Federation;
e-mail: a.n.kolobaeva@urfu.ru

Abstract

The growing capabilities of artificial intelligence (AI) and its encroachment on human prerogatives have drawn anxious attention to the topic of AI from both the public and the scientific community. In this article, based on the research of foreign and domestic philosophers, we focus on specific philosophical aspects of the specificity of AI compared to human intelligence. We demonstrate how the evolution from a naive computational/functionalist concept of intelligence to an interactive/contextual one connects AI with robotics and simultaneously actualizes the problem of developing AI as emotional and social AI. We identify three main differences between human intelligence and AI: the ability to be emotional, the ability to be social, and the ability to be abductive. Without overcoming these differences, AI cannot become an explicit, let alone a full-fledged ethical agent, and thus a harmonious partner for humanity. In conclusion, we emphasize the enormous (both positive and negative) potential of AI.

For citation

Gilyazova O.S., Zamoshchanskaya A.N. (2024) Spetsifika iskusstvennogo intellekta i chelovecheskiy intellekt: filosofskie aspekty [The Specificity of Artificial Intelligence and Human Intelligence: Philosophical Aspects]. *Kontekst i refleksiya: filosofiya o mire i cheloveke* [Context and Reflection: Philosophy of the World and Human Being], 13 (12A), pp. 24-32.

Keywords

Artificial agent, artificial intelligence, robotics, machine ethics, philosophy of science, ethical agent.

References

1. Boddington P. (2017) *Towards a code of ethics for artificial intelligence*. Cham, Switzerland: Springer. 124 p. DOI: https://doi.org/10.1007/978-3-319-60648-4
2. Boden M.A. (Ed.) (1990) *The Philosophy of Artificial Intelligence*. Oxford: Oxford University Press. 452 p.
3. Boltuc P. (2022) Transhumanities as the Pinnacle and a Bridge. *Humanities*, 11 (1): 27. DOI: https://doi.org/10.3390/h11010027
4. Brooks R.A. (1990) Elephants don´t play chess. *Robotics and Autonomous Systems*, 6 (1-2), 3-15.
5. Butkus M.A. (2020) The Human Side of Artificial Intelligence. *Science and Engineering Ethics*, 26, 2427-2437. DOI: https://doi.org/10.1007/s11948-020-00239-9
6. Damiano L., Dumouchel P. (2018) Anthropomorphism in Human-Robot Co-evolution. *Frontiers in Psychology*, 9: 468. DOI: 10.3389/fpsyg.2018.00468
7. Doya K., Taniguchi T. (2019) Toward evolutionary and developmental intelligence. *Current Opinion in Behavioral Sciences*, vol. 29, 91-96. DOI: https://doi.org/10.1016/j.cobeha.2019.04.006
8. Edmonds D. (2016) *Mashiny-ubiytsy: nado li robotu chitat' Kanta?* [Killer Machines: Should a Robot Read Kant?]. Available at: https://old.inliberty.ru/blog/2328-Mashinyubiycy-nado-li-robotu-chitat-Kanta
9. European-Commission (2018) *Communication from the commission. Artificial Intelligence for Europe*. Brussels. Available at: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM%3A2018%3A237%3AFIN
10. Farisco M., Evers K., Salles A. (2020) Towards Establishing Criteria for the Ethical Analysis of Artificial Intelligence. *Science and Engineering Ethics*, 26, 2413-2425. DOI: https://doi.org/10.1007/s11948-020-00238-w
11. Farisco M., Laureys S., Evers K. (2017) The intrinsic activity of the brain and its relation to levels and disorders of consciousness. *Mind & Matter*, 15(2), 197-219.
12. Firt E. (2024) What makes full artificial agents morally different. *AI & Society*. DOI: https://doi.org/10.1007/s00146-024-01867-6
13. Galikeeva N., Farhieva S. (2021) On the National Strategy for the Development of Artificial Intelligence Until 2030 in the Russian Federation and the Federal Project "Artificial Intelligence". *Modern school of Russia. Modernization issues*, 3-1 (36), 186-188.
14. Jordanous A. (2020) Intelligence without Representation: A Historical Perspective. *Systems*, 8, no. 3: 31. DOI: https://doi.org/10.3390/systems8030031
15. Matthias A. (2020) Dignity and Dissent in Humans and Non-humans. *Science and Engineering Ethics*, 26, 2497-2510. DOI: https://doi.org/10.1007/s11948-020-00245-x
16. Moor J.H. (2006) The nature, importance, and difficulty of machine ethics. *IEEE Intelligent Systems*, 21 (4), 18-21. DOI: https://doi.org/10.1109/MIS.2006.80
17. Potkonjak V. (2020) Is Artificial Man Still Far Away: Anthropomimetic Robots Versus Robomimetic Humans. *Robotics*, 9(3): 57. DOI: https://doi.org/10.3390/robotics9030057s
18. Prescott T.J., Robillard J.M. (2020) Are friends electric? The benefits and risks of human-robot relationships. *iScience*, 24(1): 101993. DOI: https://doi.org/10.1016/j.isci.2020.101993
19. Russell S. (2016) *Rationality and Intelligence: A Brief Update*. In: Müller V.C. (eds) Fundamental Issues of Artificial Intelligence. Synthese Library, vol. 376. Cham: Springer, pp. 7-28. DOI: https://doi.org/10.1007/978-3-319-26485-1_2
20. Tani J., White J. (2022) Cognitive neurorobotics and self in the shared world, a focused review of ongoing research. *Adaptive Behavior*, 30(1), 81-100. DOI: 10.1177/1059712320962158
21. Tsvyk V.A., Tsvyk I.V., Pavlova T.P. (2023) The problematic area of philosophical discourses on the application of artificial intelligence systems in society. *RUDN Journal of Philosophy*, 27 (4), 928-939. DOI: 10.22363/2313-2302-2023-27-4-928-939