

УДК 378.126

СОЗНАНИЕ И РАЗУМ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА: ЗАДАЧИ, ПРОБЛЕМЫ И ЭКЗИСТЕНЦИАЛЬНЫЕ УГРОЗЫ

Суслов Алексей Викторович^{1,2},

канд. филос. наук, доцент,
e-mail: suslov.aleksei@mail.ru,

Серегин Алексей Анатольевич¹,

e-mail: tihamath@yandex.ru,

¹Российский государственный социальный университет, г. Москва, Россия

²Московский университет имени С.Ю. Витте, г. Москва, Россия

В статье рассматриваются гносеологические, этико-философские и культурно-мировоззренческие аспекты развития искусственного интеллекта (ИИ). Работа знакомит читателей с современными тенденциями эволюции искусственных интеллектуальных систем, а также с позицией авторов об экзистенциальных вызовах и угрозах, которые может нести человечеству эта технология. Опираясь на идеи классической рационалистической традиции о единстве априорного и апостериорного видов знания, авторы проводят корреляцию между процессом человеческого познания и функционированием систем искусственного интеллекта на основе самообучающихся программ. Это приводит к выводу о принципиальной невозможности выхода искусственного интеллекта за пределы изначально заложенной алгоритмической модели (априорного знания) и обретения им феноменального сознания. Размышляя над одним из самых обсуждаемых (как в популярной, так и в научной литературе) вопросов о том, сможет ли ИИ превзойти и даже заменить самого человека, авторы делают акцент на уникальность экзистенциального бытия и невозможность как самим человеком, так и сверхмощным ИИ разрешить фундаментальные вопросы человеческого существования. В статье показано, что этико-антропологические вопросы и проблемы базируются на принципиальной идее наличия свободы воли и поэтому являются несопоставимыми с проблематикой создания и функционирования искусственного интеллекта. Научная новизна заключается в предлагаемой авторами концепции, согласно которой в современную эпоху – преобладания в интеллектуальной традиции позитивизма, сциентизма и технократизма – будущие возможности искусственного интеллекта и связанные с ним антропологические и социальные ожидания неизбежно будут оставаться в плоскости алгоритмизированно-рациональных координат и в принципе не могут быть переместимы в сферу иррационально-экзистенциальной проблематики. В этом и находит своё адекватное выражение диалектика огромных возможностей и непреодолимых ограничений теории и практики искусственного интеллекта.

Ключевые слова: сознание, искусственный интеллект, нейронные сети, априорное знание, апостериорное знание, экзистенциальные угрозы

CONSCIOUSNESS AND MIND OF ARTIFICIAL INTELLIGENCE: TASKS, PROBLEMS AND EXISTENTIAL THREATS

Suslov A.V.^{1,2},

candidate of philosophical sciences, associate professor,
e-mail: suslov.aleksei@mail.ru,

Seregin A.A.¹,

e-mail: tihamath@yandex.ru,

¹Russian State Social University, Moscow, Russia

²Moscow Witte University, Moscow, Russia

The article examines the epistemological, ethical, philosophical, cultural and ideological aspects of the development of artificial intelligence (AI). The work introduces readers to current trends in the development of artificial

intelligence systems, as well as the authors' position on the existential challenges and risks that this technology may pose for humanity. Based on the ideas of the classical rationalist tradition about the unity of a priori (pre-experimental) and a posteriori (experienced) types of knowledge, the authors correlate between the process of human cognition and the functioning of artificial intelligence systems based on self-learning programs. This leads to the conclusion that it is fundamentally impossible for artificial intelligence to go beyond the limits of the originally laid algorithmic model (a priori knowledge) and acquire phenomenal consciousness. Reflecting on one of the most discussed issues (both in popular and scientific literature) about whether AI can surpass and even replace humans themselves, the authors emphasize the uniqueness of existential being and the impossibility of both humans and super-powered AI to solve the fundamental issues of human existence. The article demonstrates that ethical and anthropological issues and problems are based on the fundamental idea of having free will and therefore are not comparable to the problems of AI creating and functioning. The scientific novelty lies in the statement proposed by the authors, according to which in the modern era – the predominance of positivism, scientism and technocratism in the intellectual tradition – the future possibilities of artificial intelligence and related anthropological and social expectations will inevitably remain in the area of algorithmized rational coordinates and, essentially, cannot be moved into the sphere of irrational existential issues. This is where the dialectic of the enormous possibilities and insurmountable limitations of the theory and practice of artificial intelligence finds its adequate expression.

Keywords: consciousness, artificial intelligence, neural networks, a priori knowledge, a posteriori knowledge, existential threats

DOI 10.21777/2500-2112-2024-4-104-112

1. Постановка проблемы (введение)

Актуальность темы исследования определяется возрастающей ролью цифровых технологий и их значительным влиянием на жизнь современного человека и общества. На заре своего рождения искусственный интеллект был способен лишь на простейшие логические операции, такие, как, например, последовательный подбор букв для составления осмысленных слов. В 1997 году шахматный суперкомпьютер “Deep Blue” обыграл чемпиона мира Гарри Каспарова, а в 2016 компьютерная программа “AlphaGo” выиграла матч по игре «го» у корейского профессионала 9 дана Ли Седоля. Сегодня искусственный интеллект повсеместно используется в нашей жизни: беспилотный транспорт, навигация, системы распознавания лиц и распознавания речи, банковское дело и поисковые системы. В медицине искусственный интеллект применяется в диагностике и исследованиях заболеваний, а в сервисах электронной почты отсеивает спам. Получивший широкое распространение в странах Запада, чат GPT уже активно используется в научной сфере для обработки экспериментальных данных, обобщения статистических таблиц, подготовки текстовых обзоров. В этой статье авторы задаются вопросом, на что в перспективе способен искусственный интеллект? Сможет ли он по примеру биологических существ самостоятельно эволюционировать? Способен ли искусственный интеллект превзойти интеллект человека и обрести сознание? Какие в связи с этим риски и угрозы ожидают человечество?

Автор термина «искусственный интеллект» (ИИ), американский информатик Джон Маккарти, отвечая на вопрос: «Пытается ли ИИ достичь уровня интеллекта человека?», сказал, что конечная цель – создать алгоритмы, которые смогут решать проблемы и выполнять задачи в реальном мире так же хорошо, как и люди. Вот уже 70 лет в научном мире тема ИИ вызывает перманентный, пролонгированный интерес. Проблемы, связанные с ИИ, его развитием и возможностями, охватывают широкий круг научных дисциплин, как естественных (кибернетика, математика, физика, информатика и даже химия), так и гуманитарных (философия, социология, психология и антропология). Рассмотрим несколько примеров научных работ по теме искусственного интеллекта.

В статье И.М. Меликова и И.С. Журинского «Искусственный интеллект как угроза смысловой детерминанте личности» [1] авторы рассматривают проблему места личности в эпоху цифрового общества и участия искусственного интеллекта в интеллектуальной деятельности человека. Отмечается особая роль ИИ как помощника человека в процессе обработки большого объема информации. Также

в статье приведены интересные выводы, например, о невозможности бессмысленной деятельности и утрате стремления искать смысл в условиях неограниченной информации. К сожалению, в работе не прослеживается влияние ИИ непосредственно на человека.

В статье «Злонамеренное использование искусственного интеллекта: новые угрозы для международной информационно-психологической безопасности и пути их нейтрализации» [2] автор Е.Н. Пашенцев исследует риски информационной безопасности, вызванные внедрением искусственного интеллекта в различные области социума. Автор пишет, что в результате внедрения ИИ возможна не только массовая безработица, но и утрата частичного и даже полного контроля за результатами его применения. Это довольно спорное утверждение. Тем более, что и сам автор пишет далее, что в основной массе такие опасения обусловлены угрозой от злонамеренного использования искусственного интеллекта людьми.

Являясь частью самой быстро развивающейся индустрии информационных технологий, искусственный интеллект постоянно совершенствуется и прогрессирует. Так что, несмотря на большое количество работ, посвященных теме искусственного интеллекта, перед обществом встают новые вопросы, требующие их постоянного изучения, анализа и решения.

Цель исследования – разработать концепцию диалектики возможностей и непреодолимых ограничений теории и практики искусственного интеллекта на основе анализа тенденций и перспектив развития технологий искусственного интеллекта в широком культурно-мировоззренческом и этико-философском контексте.

Задачи исследования:

- установить содержание понятия «искусственный интеллект»;
- проанализировать основные этапы эволюции систем искусственного интеллекта;
- проанализировать основные концепции искусственного интеллекта в современной научной литературе;
- сопоставить процесс познания как деятельности человека по обретению новых знаний на основе чувственного опыта и рационального мышления с технологией функционирования самообучающихся компьютерных программ;
- дать попытку ответа на один из самых обсуждаемых вопросов о том, сможет ли в перспективе искусственный интеллект стать системой, потенциально способной превзойти человеческий разум.

Методами исследования являются: индуктивное обобщение, дедуктивные выводы, умозаключения по аналогии, мысленный эксперимент и идеализация.

2. Анализ и оценка возможностей искусственного интеллекта

Первые работы по созданию машинного интеллекта начались в середине прошлого века. Так, например, нейрофизиолог У. Мак-Калок и математик В. Питтс в 1943 году опубликовали свою статью «Логическое исчисление идей, относящихся к нервной активности» [3], заложив концепцию нейронных сетей. В 1949 году в работе «Организация поведения как нейропсихологическая теория» [4] канадским физиологом и нейропсихологом Дональдом Хеббом были описаны принципы обучения нейронов. Следующим шагом стал вопрос о критериях оценки разумности машинного интеллекта. В статье «Вычислительные машины и интеллект» [5], вышедшей в 1950 году, американский математик Алан Тьюринг утверждал, что если машина в слепом тесте может успешно притвориться человеком, отвечая на вопросы наблюдателя, то её безусловно можно считать разумной. «Нам кажется, что метод вопросов и ответов пригоден для того, чтобы охватить почти любую область человеческой деятельности, какую мы захотим ввести в рассмотрение ... и мы будем считать, что наилучшая стратегия для машины состоит в том, чтобы давать ответы, которые в соответствующей обстановке дал бы человек».

Поворотным моментом в дальнейшем развитии искусственного интеллекта стал «Дартмутский семинар». Ученые, собравшиеся в Дартмуте, обсуждали различные подходы к созданию машин, которые способны были бы выполнять задачи, требующие интеллектуальных усилий, аналогичных человеческим. На семинаре присутствовали знаменитые ученые: основоположник кибернетики Норберт Ви-

нер, ученый-когнитивист, основатель первой лаборатории искусственного интеллекта Марвин Мински, инженер и математик Клод Шеннон. Их идеи послужили основой для дальнейших исследований в области ИИ. В результате успешных экспериментов и реализации первых алгоритмов, обрабатывающих информацию, искусственный интеллект начал развиваться как отдельная область науки.

Первые достижения были довольно обнадеживающими. В 1950-х и 1960-х годах учёные разрабатывали программы, способные решать математические задачи, играть в шахматы и даже вести простые диалоги. Стали доступны первые логические машины и системы на основе правил, имевшие успех в решении различных прикладных задач.

Вместе с тем, многие ожидания не были оправданы. В 1970-х и 1980-х годах исследования ИИ переживали периоды стагнации, когда финансирование сократилось, а интерес к новой технологии снизился. Многие разработанные алгоритмы не могли эффективно масштабироваться, а недостаток вычислительных ресурсов тормозил их развитие.

Тем не менее, исследования в области теории вероятностей и математической статистики продолжались, что дало новый импульс развитию технологий ИИ уже в XXI веке. Применение глубокого обучения, больших данных и мощных вычислительных возможностей способствовало тому, что ИИ начал активно использоваться в различных сферах: от медицины и финансов до автоматизации производственных процессов.

В ходе дальнейших исследований стало очевидно, что создание «универсального интеллекта», способного к самосознанию и пониманию человеческих эмоций, остается крайне сложной задачей. Это привело к обсуждениям этических вопросов, связанных с ИИ, его влиянием на общество и возможностью создания безопасных и ответственных систем.

Сегодня, несмотря на гигантский скачок в развитии информационных технологий, стоит отметить, что критерии оценки разумности машины, сформулированные Тьюрингом более 70 лет назад, представляются авторам релевантными и сейчас. Действительно, притворство, обман, хитрость и прочие поведенческие категории характерны только для живых биологических существ, обладающих разумом, инстинктом, сознанием. Что важно, конкретно наличие сознания у искусственного интеллекта, станет, во-первых, абсолютным критерием превосходства разума ИИ над разумом человека. И, во-вторых, поскольку именно мозг порождает сознание, будет решён один из главных вопросов проблемы «сознание – тело» – является ли сознание физическим?

Исходя из цели исследования – анализ тенденций и перспектив развития ИИ, можно поставить три ключевых вопроса, ответы на которые позволят понять, на что будет способен искусственный интеллект, если он сможет:

- 1) сравняться и превзойти интеллект человека;
- 2) самостоятельно эволюционировать;
- 3) представлять экзистенциальные угрозы человечеству.

Чтобы решить возможные в будущем проблемы, связанные с развитием ИИ, необходимо, во-первых, разобраться и понять, что собой представляет искусственный интеллект, какова его природа; во-вторых, увязать сугубо технические начала ИИ с гуманитарным знанием человека.

Понятие «интеллект» довольно однозначно трактуется словарями. Так, например, в толковом Словаре русского языка С.И. Ожегова¹ интеллект – это «Мыслительная способность, умственное начало у человека, определяющее деятельность его». Представление о ключевых характеристиках интеллекта, его качественного и количественного аспектов даёт определение современного французского философа Андре Конт-Спонвиля²: «Большая или меньшая способность решать проблемы, иначе говоря, способность понимать новое и сложное». Другими словами, интеллект, во-первых, является результатом процесса эволюции живой природы, во-вторых, поддаётся измерению (хотя и сугубо относительному).

Существует достаточное количество определений понятия искусственного интеллекта, что обусловлено его онтологическим дуализмом – разумная (в идеале), материальная (не биологическая) си-

¹ Толковый словарь русского языка С.И. Ожегова. – Москва: Мир образования, 2018. – С. 384.

² Философский словарь Конта-Спонвиля. – Москва: Eterna, 2012. – С. 223.

стема. Сегодня искусственным интеллектом принято называть системы, способные решать нерутинные задачи. Чтобы не провоцировать разночтения в понимании данной работы, авторы будут отталкиваться от известного определения искусственного интеллекта В. Барра и Э. Фейгенбаума³: «Искусственный интеллект – это область информатики, которая занимается разработкой интеллектуальных компьютерных систем, т.е. систем, обладающих возможностями, которые мы традиционно связываем с человеческим разумом: понимание языка, обучение, способность рассуждать, решать проблемы и т.д.».

История эволюции ИИ на сегодняшний день состоит из трёх этапов. Первый – это этап становления и начального развития искусственного интеллекта с 50-х годов прошлого века. С 80-х годов начинается второй, так называемый, этап машинного обучения. Третий этап – глубокое обучение. Его развитие мы можем наблюдать в настоящее время.

Первые искусственные интеллектуальные системы решали поставленные перед ними задачи методом простейшего перебора (подбора) возможных вариантов решений. И делали это с успехом с учётом высокой, относительно арифметических способностей человека, производительности – скорости тактовой частоты центрального процессора. Например, первый однокристалльный микропроцессор Intel 4004, вышедший в ноябре 1971 года, работал на частоте 108 кГц и проводил 92 600 логических операций в секунду.

3. Искусственный интеллект: разум или алгоритм

3.1. Машинное и глубокое обучение

Машинное и глубокое обучение являются частью ИИ и между собой отличаются тем, что в глубоком обучении используются гораздо более сложные алгоритмы и продвинутые математические модели. Машинное обучение строится на процессах выведения общих правил из пула примеров данных. Таким образом, алгоритм «учится» решать задачу на основе введённой информации. Простой пример машинного обучения – типизация фотографий: в алгоритм загружаются несколько тысяч фотографий собак и кошек с целью научить его отличать их по биологическому виду. Каждый раз, когда программа ошибается, алгоритм корректируется. После некоторого количества циклов получается обученный алгоритм, способный отличать собак и кошек.

Глубокое обучение – это подвид машинного обучения с использованием, как правило, нейронных сетей. Искусственная нейронная сеть (ИНС) представляет собой множество последовательных слоёв, каждый из которых состоит из нейронов и выполняет свою, конкретную функцию. Аналогия с нейронами головного мозга, разумеется, условна, но уместна, поскольку нейрон – это электрически возбудимая клетка, предназначенная для приёма, обработки, хранения и передачи информации посредством электрических импульсов. Обучение программы происходит методом «поощрения» и «наказания». Наказание подразумевает корректировку работы алгоритма. Точность логических выводов, т.е. результат и эффективность работы ИНС напрямую зависят от использованных в обучении примеров данных. Большой объём обеспечит высокую точность, но усложнит анализ и увеличит вероятность «переобучения» ИНС.

Интеллект человека можно описать тремя неразрывно связанными категориями – это созерцание, рассудок и разум. Не вызывает сомнения, что с помощью электронно-механических датчиков мы можем научить ИИ созерцанию, т.е. по-своему видеть, слышать, осязать и т.д. окружающий мир. Также ИИ можно наделить специфическим рассуждением, потому что многие феномены встраиваются в некую логику посредством наблюдения. Например, яблоко с дерева всегда падает вниз, а не летит вверх, – это простейшее рассуждение, сделанное на основе наблюдения, в котором нет особого интеллекта. Но искусственный интеллект невозможно наделить разумом. Он не сможет улавливать закономерности мира и выдвигать гипотезы. Человек, как высокоорганизованное живое существо, обладает определёнными свойствами, которые не могут быть присущи ИИ. Нейронные сети можно учить, но невозможно научить их додумываться и выдвигать гипотезы.

³ *Барр А., Фейгенбаум Э.* Справочник по искусственному интеллекту. Т. 1. – California: Butterworth-Heinemann, 2014. – 424 с.; *Джексон П.* Введение в экспертные системы. – 3-е изд. – Москва [и др.]: Вильямс, 2001. – 622 с.

3.2. Априорное и апостериорное в искусственном интеллекте

В настоящее время одной из широко обсуждаемых проблем является дальнейшая эволюция ИИС, в частности, вопрос о том, смогут ли роботы превзойти человека по своим интеллектуальным способностям или даже соединиться с человеческим телом, усилив его естественные физические и умственные способности [6]. Эти размышления приводят нас к вопросу о природе и возможностях человеческого познания, которые традиционно в европейской философии решались в терминах априорного и апостериорного знания.

A priori – это знание, предшествующее опыту, оно изначальное и не зависит от него. *A posteriori* – знание, получаемое из опыта.

Понятие априорного (не термин) знания появляется у Платона в «учении о припоминании» в диалоге Менон. Согласно Платону [7], это внутреннее, наследуемое и неотъемлемое человеческому уму знание. В схоластике средних веков априорное определялось в ипостаси познания вещей как действий из их причин, то, что является их предпосылкой. Апостериорное же – познание вещей как причин из действий. Так, Фома Аквинский в споре об универсалиях изложил их тройственную природу: «до вещей», в «вещах» и «после вещей». Знание «до вещей» не что иное, как идеи будущих вещей, а «после вещей» – это результат человеческого мышления: абстрагирование и обобщение [8].

Начиная с XVII века идея априорного знания стала довольно широко использоваться в гносеологии, прежде всего в немецкой. Для Р. Декарта эта идея выражена в учении об интеллектуальной интуиции и «чистом разуме», которые предшествуют опыту [9]. Чистый разум – это знание абсолютное и неизменное, которое накладывает на данные опыта определённые идеи и схемы. Примером такого знания является способность принять очевидное без доказательства, сделать логический вывод из исходных суждений и т.д.

В гносеологии И. Канта понятия априорного и апостериорного знания являются ключевыми. Процесс познания, в результате которого и рождается апостериорное знание, есть синтез чувственности и рассудка и представляет собой наложение доопытных форм чувственности и категорий рассудка на материал ощущений. В «Критике чистого разума» И. Кант пишет: «Но хотя всякое наше познание и начинается с опыта, отсюда вовсе не следует, что оно целиком происходит из опыта. Вполне возможно, что даже наше опытное знание складывается из того, что мы воспринимаем посредством впечатлений, и из того, что наша собственная познавательная способность (только побуждаемая чувственными впечатлениями) даёт от себя самой, причем это добавление мы отличаем от основного чувственного материала лишь тогда, когда продолжительное упражнение обращает на него наше внимание и делает нас способными к обособлению его» [10].

Если проецировать эти философские понятия на искусственный интеллект, то программа, заложенная в память машины, – это знание априорное, т.е. доопытное, абсолютное и неизменное. Искусственные интеллектуальные системы первого уровня функционировали исключительно на основе априорного знания. Алгоритм корректировки обучения – это знание, полученное в опыте и обусловленное им. То есть ИИ получает способность к самообучению и самостоятельному использованию в своем функционировании нового знания. Именно самообучающиеся программы лежат в основе современных разработок в области искусственного интеллекта, на которые их создатели возлагают большие надежды.

Одна из идей гносеологии И. Канта состоит в том, что в априорных формах познаваем лишь мир явлений. Мир «вещей в себе» (внешний мир) остается недоступен, ибо человек не может выйти за пределы эмпирического опыта и не может воспринимать реальность вне априорных форм (то есть вне субъект-объектных, пространственно-временных и т.д. отношений). Логично предположить, что и искусственный интеллект сможет функционировать и «развиваться» на основе заложенной изначально алгоритмической модели, будучи не способным выйти за её пределы.

Наряду с доопытными формами чувственности и рассудка априорный характер носят также и нравственные идеи, ибо человек есть изначально нравственное существо, что является тайной, достойной удивления. Вспомним слова, написанные на могиле И. Канта: «Звёздное небо надо мной и моральный закон во мне». Зададимся вопросом, способен ли будет эволюционирующий ИИ к нравственному действию и нравственной оценке и какие моральные идеи и принципы следует заложить при его раз-

работке? Уже на стадии программирования ИИС возникают вопросы, связанные с действием системы в критической ситуации и упирающиеся в моральные дилеммы, не имеющие решения. Так, в случае с беспилотным автомобилем возникает вопрос о том, можно ли спасти несколько человек ценой жизни одного человека (так называемая «проблема вагонетки»), оправдано ли свернуть с дороги ради спасения пешехода или возможно проехать через сплошную линию. Подходы к решению этих вопросов могут отличаться в рамках разных этических доктрин (абсолютистской, утилитарной, прагматической).

Обратим внимание, что этические вопросы и проблемы базируются на принципиальной идее наличия свободы воли и поэтому являются несопоставимыми с проблематикой создания и функционирования ИИ. Проблема свободы воли или свободного нравственного выбора между как минимум двумя или большим количеством возможностей часто характеризуется философским знанием в качестве тайны, так как принципиально непостижимо, какая из возможностей будет реализована в свободном выборе личности. Об этом же знаменитые слова Ф. Достоевского о том, что «человек есть тайна. Ее надо разгадать, и ежели будешь ее разгадывать всю жизнь, то не говори, что потерял время; я занимаюсь этой тайной, ибо хочу быть человеком» [11]. Что касается искусственного интеллекта и связанных с ним технологий, то при всех его нынешних и будущих, поражающих воображение возможностях, он остается сугубо алгоритмизированным объектом. Будучи способным сочетать в себе априорное и апостериорное начало, будучи способным к самообучению и даже к саморазвитию, он остается в смысловом поле координат, несовместимым ни со свободой экзистенциального выбора, ни с тайного человеческого бытия.

4. Современные дискуссии о сознании и разуме искусственного интеллекта

Искусственный интеллект – это область исследований, некоторое поле деятельности, где создаются в основном алгоритмы и в меньшей степени киберфизические системы, которые симулируют интеллектуальные решения или интеллектуальную деятельность человека. Но искусственный интеллект никогда не будет испытывать те же потребности, что и человек. У него нет человеческого тела, а, значит, и возможности так же чувствовать окружающий мир.

Стоит особо выделить факт отсутствия сознания и, как следствие, неспособность сознательного опыта себя и мира у искусственного интеллекта. У человека категории разума как способа мыслительной деятельности и феноменального сознания как субъективного переживания явлений внешнего мира неотделимы. Так что для преодоления проблемы превосходства искусственному интеллекту необходимо не только превзойти человеческий разум, но и обрести собственное сознание.

В ноябре 2021 года группа учёных – биолог-теоретик, профессор Университета Пенсильвании С. Кауффман, доцент кафедры компьютерных наук и инженерии Болонского университета А. Роли и системный эволюционный биолог, бывший директор Института эволюции и когнитивных исследований имени Конрада Лоренца Й. Йегер опубликовали свою работу «Как организмы познают мир: фундаментальные ограничения на искусственный общий интеллект» [12]. В ней авторы представили свою «теорию относительности интеллекта». Общий вывод этой теории состоит в том, что искусственный интеллект не может быть опасен для человека, поскольку, в отличие от естественного интеллекта, он очень ограничен – лишен так называемых аффордансов. Аффорданс – это способность интеллектуальных агентов, то есть людей и животных, находить и использовать конкретные свойства мира в зависимости от того, насколько эти свойства полезны или вредны для достижения целей интеллектуальных агентов. Проще говоря, физические свойства, скажем, булыжника, обеспечивают очевидную возможность использовать его в качестве молотка, ну или прочного строительного материала.

Существует ещё одна проблема построения искусственного интеллекта. Для некоторых задач невозможно построить универсальный алгоритм, который бы их решал. Дело в том, что человеку самому трудно понять, как именно надо поступать в том или ином экстренном случае. И он не может показать алгоритмам примеры правильного и неверного поведения.

Одна из главных задач данного исследования состояла в том, чтобы показать объективную несостоятельность популярных сегодня утверждений о будущем всемогуществе искусственного интеллекта, о том, что вскоре ИИ превзойдёт разум человека и станет венцом творения.

Даже если искусственный разум будет создан, непонятно, насколько он будет полезен человеку? Механические машины давно стали заменять человека в сфере рутинного, машинообразного труда. Но стремление делегировать электронным помощникам разные задачи привело человека к тому, что он начал терять присущие ему способности. Таким образом, машины постепенно вытесняют человека из его привычной среды. Такая новая техногенная среда, которая значительную часть интеллектуальных решений примет на себя, будет потенциально опасной для человека. Это следует из того, что машины допускают ошибки, которые могут приводить к авариям и техногенным катастрофам.

В научной литературе ИИ принято разделять на узкий или слабый, способный решать одну определённую задачу, например, беспилотное управление, распознавание лиц или речи, и общий или сильный – будущее искусственного интеллекта. Утверждается, что если слабый ИИ (настоящее) способен, условно, обыграть в шахматы человека, то сильный ИИ (предстоящее) превзойдёт человека в любой когнитивной задаче. Это не совсем так. Во-первых, сильный искусственный интеллект – это просто многозадачность. Напомним, что первая многозадачность была реализована в 1964 году (OS Multics). Во-вторых, очень важную роль в интеллектуальной деятельности человека играет когнитивная система, то есть система познания, которая отсутствует у ИИС, либо имеет иную природу.

Если развивать тему когнитивных способностей и навыков, стоит упомянуть социально-когнитивную теорию. В её основе лежит утверждение, что социальные взаимодействия оказывают влияние на поведение субъекта. Таким образом, наблюдая за другими поведенческими образцами, индивид получает опыт, на котором может строить свои модели поведения. Другими словами, люди учат людей. Машины не могут и не смогут учить друг друга подражать себе подобным.

Ещё одна распространённая в научной литературе концепция развития ИИ заключается в том, что его эволюция может пойти по пути конструктивного или деструктивного для человека развития [6]. То есть, либо искусственный интеллект будет работать на благо человека, либо во зло. Действительно, очень важно, чтобы прорывные решения, открывающие поистине безграничные возможности, работали не во вред человеку. Однако, учитывая, что любой ИИ – это программа, созданная человеком, и алгоритмы его обучения корректируются тоже человеком, то, следовательно, деструктивным характером ИИ может наделить только человек. Так что все, кто говорит об опасности и деструктивности искусственного интеллекта, забывают о том, насколько деструктивен естественный интеллект человека.

Заключение

На основе проведённого исследования можно сделать следующие выводы. Внедрение технологий ИИ несёт как положительные изменения, так и определённые вызовы и риски, влияние которых на человека и общество ещё предстоит оценить. ИИС (искусственные интеллектуальные системы) способны обрабатывать в единицу времени количество данных, несопоставимое с возможностями человеческого мозга. Поэтому машина побеждает человека в интеллектуальных видах спорта и в практической деятельности, где речь идет о переборе вариантов, а не о творчестве, интуиции и озарении. ИИС не смогут заменить человека, поскольку не способны использовать ситуационные знания, рассуждать метафорически и лишены здравого смысла. Для создания ИИ, мыслительные способности которого были бы выше или хотя бы одинаковы с такими же, как у человека, искусственный интеллект должен не только обладать более совершенным (или сравнимым с человеческим) разумом, но и обладать собственным сознанием.

Технические устройства на основе ИИ можно понимать как алгоритмизированные объекты, неспособные к метафизической практике, к этической рефлексии и в этом смысле не могут быть тождественны человеку, имеющему уникальную духовно-экзистенциальную природу. Отвечая на вопрос о том, удержит ли человек под контролем искусственный интеллект, можно предположить, что да, если только саморазрушительные возможности человека не превысят его собственный инстинкт самосохранения.

Список литературы

1. Меликов И.М., Журинский И.С. Искусственный интеллект как угроза смысловой детерминанте личности // V Международная научно-практическая конференция: сборник статей. – Москва: РГСУ, 2021. – С. 38–42.
2. Пащенко Е.Н. Злонамеренное использование искусственного интеллекта: Новые угрозы для международной информационно-психологической безопасности и пути их нейтрализации // Государственное управление. Электронный вестник. – 2019. – № 76. – С. 279–300. – URL: <https://doi.org/10.24411/2070-1381-2019-10013> (дата обращения: 10.01.2024). – Текст: электронный.
3. McCulloch W.S., Pitts W. A logical calculus of the ideas immanent in nervous activity // Bull. Math. Biophys. – 1943. – Vol. 5. – P. 115–133.
4. Хебб О.Б. Организация поведения как нейропсихологическая теория. – New York: Wiley J & Sons, Inc., 1949. – 279 с.
5. Тьюринг А.М. Вычислительная техника и интеллект // Разум: ежеквартальный журнал. Обзор психологии и философии. – 1950. – Vol. 59, No. 236. – С. 433–460.
6. Милкова Э.Г. Развитие искусственного интеллекта и мифы, связанные с ним // Universum: технические науки: электронный научный журнал. – 2021. – № 6 (87). – URL: <https://7universum.com/ru/tech/archive/item/11992> (дата обращения: 10.07.2024). – Текст: электронный.
7. Платон. Собрание сочинений: в 4 т.: пер. с др.-греч. – Москва: Мысль, 1990–1994. – Т. 1.
8. Вербеке Г. Фома Аквинский и проблемы его времени. – Левен; Гаага, 1976.
9. Декарт Р. Происхождение философии: сочинения: в 2 т. – Москва: Мысль, 1989–1994. – Т. 1.
10. Кант И. Критика чистого разума: пер. с англ. – Москва: Академический проект, 2018. – 567 с.
11. Достоевский Ф.М. Полное собрание сочинений в пятнадцати томах. – Санкт-Петербург: Изд-во Наука. Ленинградское отделение, 1966. – Т. 15.
12. Роли А., Джеггер Дж., Кауффман С.А. Как организмы познают мир: фундаментальные ограничения на искусственный интеллект. – URL: <https://osf.io/yfmt3/> (дата обращения: 25.09.2024). – Текст: электронный.

References

1. Melikov I.M., Zhurinskij I.S. Iskusstvennyj intellekt kak ugroza smyslovoj determinante lichnosti // V Mezhdunarodnaya nauchno-prakticheskaya konferenciya: sbornik statej. – Moskva: RGSU, 2021. – S. 38–42.
2. Pashencev E.N. Zlonamerennoe ispol'zovanie iskusstvennogo intellekta: Novye ugrozy dlya mezhdunarodnoj informacionno-psihologicheskoy bezopasnosti i puti ih nejtralizacii // Gosudarstvennoe upravlenie. Elektronnyj vestnik. – 2019. – № 76. – S. 279–300. – URL: <https://doi.org/10.24411/2070-1381-2019-10013> (data obrashcheniya: 10.01.2024). – Tekst: elektronnyj.
3. McCulloch W.S., Pitts W. A logical calculus of the ideas immanent in nervous activity // Bull. Math. Biophys. – 1943. – Vol. 5. – P. 115–133.
4. Hebb O.B. Organizaciya povedeniya kak nejropsihologicheskaya teoriya. – New York: Wiley J & Sons, Inc., 1949. – 279 s.
5. T'yuring A.M. Vychislitel'naya tekhnika i intellekt // Razum: ezhekvartal'nyj zhurnal. Obzor psihologii i filosofii. – 1950. – Vol. 59, No. 236. – S. 433–460.
6. Milkova E.G. Razvitie iskusstvennogo intellekta i mify, svyazannye s nim // Universum: tekhnicheskie nauki: elektronnyj nauchnyj zhurnal. – 2021. – № 6 (87). – URL: <https://7universum.com/ru/tech/archive/item/11992> (data obrashcheniya: 10.07.2024). – Tekst: elektronnyj.
7. Platon. Sbranie sochinenij: v 4 t.: per. s dr.-grech. – Moskva: Mysl', 1990–1994. – T. 1.
8. Verbeke G. Foma Akvinskij i problemy ego vremeni. – Leven; Gaaga, 1976.
9. Dekart R. Proiskhozhdenie filosofii: sochineniya: v 2 t. – Moskva: Mysl', 1989–1994. – T. 1.
10. Kant I. Kritika chistogo razuma: per. s angl. – Moskva: Akademicheskij proekt, 2018. – 567 s.
11. Dostoevskij F.M. Polnoe sobranie sochinenij v pyatnadcati tomah. – Sankt-Peterburg: Izd-vo Nauka. Leningradskoe otdelenie, 1966. – T. 15.
12. Roli A., Dzheger Dzh., Kauffman S.A. Kak organizmy poznayut mir: fundamental'nye ogranicheniya na iskusstvennyj intellekt. – URL: <https://osf.io/yfmt3/> (data obrashcheniya: 25.09.2024). – Tekst: elektronnyj.