

Дормидошина Дарья Андреевна,
Технический директор, АО «ЦКБ «Дейтон»
Dormidoshina Daria A.,
Technical Director, JSC «CDO «Deyton»

**ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ:
ОТ НАУЧНОЙ ФАНТАСТИКИ К РЕАЛЬНОСТИ
ARTIFICIAL INTELLIGENCE:
FROM SCIENCE FICTION TO REALITY**

Аннотация. Цель исследования – анализ исторического развития искусственного интеллекта, его философских основ и социально-этических последствий внедрения в современное общество. В работе исследуются следующие ключевые аспекты: исторический контекст, философские проблемы, социально-экономическое влияние, будущие вызовы. В статье искусственный интеллект рассматривается как технология, которая не только трансформирует общество, но и ставит перед человечеством сложные философские и этические вопросы. В статье акцентируется внимание на том, что искусственный интеллект представляет собой не только технологический инструмент, но и феномен, нуждающийся в серьезном философском анализе. Научная новизна статьи заключается в систематизации и расширении существующих знаний об искусственном интеллекте через призму философии, истории и социологии, а также в постановке новых исследовательских задач, актуальных для эпохи стремительного технологического прогресса. Тема особенно значима в контексте стремительного развития технологий и их проникновения во все сферы жизни. Статья предлагает комплексный взгляд на искусственный интеллект, объединяющий историю, философию и прогнозы на будущее.

Abstract. The purpose of this study is to analyze the historical development of artificial intelligence, its philosophical foundations, and the socio-ethical implications of its implementation in modern society. The paper examines the following key aspects: historical context, philosophical issues, socio-economic impact, and future challenges. This article examines artificial intelligence as a technology that not only transforms society but also poses complex philosophical and ethical questions for humanity. It emphasizes that artificial intelligence is not only a technological tool but also a phenomenon requiring serious philosophical analysis. The scientific novelty of this article lies in its systematization and expansion of existing knowledge on artificial intelligence through the prism of philosophy, history, and sociology, as well as in its formulation of new research problems relevant to the era of rapid technological progress. This topic is particularly significant in the context of the rapid development of technology and its penetration into all spheres of life. The article offers a comprehensive view of artificial intelligence, combining history, philosophy, and forecasts for the future.

Ключевые слова: Искусственный интеллект, нейронная сеть, философия, история, машинное сознание, роботизация, технологии, машины.

Keywords: Artificial intelligence, neural network, philosophy, history, machine consciousness, robotics, technology, machines.

Введение

Актуальность изучения искусственного интеллекта в философском, историческом и социально-этическом аспектах обусловлена стремительным развитием технологий и их растущим влиянием на все сферы человеческой жизни. Созданные человеком машины уже способны выполнять всеобъемлющие виды трудоемкой работы. Тем не менее, во многих случаях, движимые требованиями более высокой производительности и, возможно, просто



любопытством, люди пытались внедрить человеческий интеллект в машины, что составляет изначальную мотивацию искусственного интеллекта. Исследования искусственного интеллекта ведутся уже более 65 лет и достигли впечатляющих успехов с точки зрения теоретического изучения и реальных приложений. Искусственный интеллект используется практически везде и считается основным навыком будущего. Прогнозируется, что рынок искусственного интеллекта вырастет до 190 миллиардов долларов в 2025-2026 годах при среднегодовом темпе роста (CAGR) более 36% в период с 2018 по 2026годы [15]. Существует множество определений искусственного интеллекта. В тесте Тьюринга искусственный интеллект определяется как способность машин общаться с людьми (используя электронные устройства вывода), не раскрывая при этом, что они не являются людьми. Марвин Мински, один из пионеров искусственного интеллекта, определил его как возможность машинам выполнять действия, требующие человеческого интеллекта. Символическая школа полагает, что искусственный интеллект – это операция символов, а самые примитивные символы соответствуют физическим сущностям. Хотя описания искусственного интеллекта различны, широко распространено мнение, что ядро искусственного интеллекта – это исследовательские теории, методы, технологии и приложения для моделирования и расширения человеческого интеллекта. В настоящее время концепция искусственного интеллекта оказывает все более глубокое влияние на жизнь человека. Как и роли паровых двигателей в эпоху пара, генераторов в эпоху электричества, компьютеров в эпоху информации, так и искусственный интеллект является опорой технологий в современную эпоху и в будущее.

В настоящее время в развитых и быстроразвивающихся странах интенсивно ведутся исследования полноценных систем искусственного интеллекта. Это стимулирует создание высокоэффективных автоматизированных систем и внедрение роботов (роботизированных систем) в образование, промышленность, сферу услуг и быта, повседневную жизнь. В литературе интенсивно обсуждаются как позитивные, так и негативные последствия этих процессов в ближайшем будущем. Роботизация основных функций человека в производственной сфере и в системе социально значимых услуг может привести к массовой безработице.

На основе анализа современных тенденций роботизации и внедрения компьютерного зрения обсуждается предположение, что роботы, оснащенные универсальным человекоподобным интеллектом и искусственным суперинтеллектом, уже в нашем столетии смогут составить серьезную конкуренцию человеку и даже сделать его дисфункциональным во многих сферах жизнедеятельности. Это должно привести к тому, что проблемы роботизации и компьютеризации всех сфер общественной и индивидуальной жизни станут важнейшим предметом философского дискурса середины и второй половины XXI в.

Наблюдается, что философия искусственного интеллекта – это область, которая охватывает этические, философские и экзистенциальные последствия развития искусственного интеллекта и растущего его присутствия в обществе. Она включает в себя изучение фундаментальных концепций, таких как интеллект, знания и искусственный интеллект, и того, как они влияют на человеческое существование [17]. Развитие искусственного интеллекта поднимает важные вопросы об отношениях между людьми и машинами, будущем человечества и необходимости этических стандартов и механизмов управления [11]. Философские основы искусственного интеллекта имеют важное значение для руководства разработками и использованием искусственного интеллекта в Обществе 5.0 (социально-экономическая и культурная стратегия развития общества, основанная на использовании цифровых технологий во всех сферах жизни). Это будущее общество, как представляется, будет гармоничным сосуществованием людей и искусственного интеллекта, где технологии служат для расширения человеческих возможностей и улучшения общего



благополучия общества, сохраняя при этом человеческие ценности и достоинство. Философия искусственного интеллекта играет решающую роль в обеспечении того, чтобы системы искусственного интеллекта разрабатывались ответственным и этичным образом, в соответствии с ценностями и стремлениями общества и вносили вклад в лучшее будущее для всех [24]. Некоторые из ключевых философских основ искусственного интеллекта включают изучение дуализма разума и тела, природы интеллекта и знаний, ограничений искусственного интеллекта и его влияния на человеческое существование, отношений между людьми и машинами, этических последствий искусственного интеллекта и необходимости этических стандартов и механизмов управления [11]. Эти философские основы помогают обеспечить основу для размышлений о фундаментальных вопросах и проблемах, которые возникают в связи с растущим присутствием и влиянием искусственного интеллекта в нашей жизни. Критическая оценка приложений искусственного интеллекта и их последствий чрезвычайно важна. Этого можно достичь посредством междисциплинарного диалога и исследований, чтобы лучше понять влияние искусственного интеллекта на человечество и мир в целом. Именно это делает глубокое понимание философии искусственного интеллекта очень важным, поскольку оно обеспечивает основу для оценки этических и моральных принципов. Изучая философские основы, можно получить более глубокое понимание влияния искусственного интеллекта на человеческое и социальное существование и предположить, что технологии искусственного интеллекта разрабатываются ответственным и этичным образом, который соответствует ценностям и стремлениям общества [23, 18].

В статье рассматривается комплекс взаимосвязанных проблем, возникающих в связи с развитием искусственного интеллекта. Основное внимание уделено фундаментальным противоречиям между технологическим прогрессом и человеческими ценностями. Ключевая проблема заключается в несоответствии между стремительным развитием технологий искусственного интеллекта и недостаточной проработанностью их философско-этических оснований. В статье показано, как это противоречие проявляется в нескольких аспектах. Во-первых, это вопрос о природе искусственного интеллекта: может ли он обладать сознанием или лишь имитирует когнитивные процессы. Во-вторых, возникает проблема ответственности - кто должен отвечать за решения, принимаемые автономными системами.

Важной составляющей исследования также является анализ этических дилемм, связанных с искусственным интеллектом, и проблема будущего человечества в условиях стремительного развития искусственного интеллекта.

Искусственный интеллект: от истоков до современности

Мир, в котором мы живем сегодня, во многом напоминает Страну чудес, похожую на ту, которую британский математик Чарльз Лютвидж Доджсон, более известный под именем Льюис Кэрролл, описал в своих знаменитых романах. Технологии, такие как распознавание изображений, умные колонки и беспилотные автомобили, стали возможны благодаря искусственному интеллекту. Искусственный интеллект определяется как способность системы интерпретировать внешние данные, извлекать из них знания и использовать их для достижения целей через гибкую адаптацию. Хотя искусственный интеллект как академическая дисциплина зародился в 1950-х годах, его практическое применение долгое время оставалось ограниченным. Сегодня, благодаря развитию больших данных и увеличению вычислительной мощности, искусственный интеллект стал неотъемлемой частью бизнеса и общественных дискуссий.

Период до 1956 года считается инкубационным периодом искусственного интеллекта. Ученые и инженеры пытались автоматизировать умственную работу. В 1936 году математик Алан Тьюринг предложил математическую модель идеального компьютера, которая легла в основу современных вычислительных машин. В 1943 году нейрофизиологи Уоррен



МакКаллок и Уолтер Питтс создали первую модель искусственной нейронной сети (модель М-П), имитирующей работу биологических нейронов. Эту модель можно считать самой ранней искусственной нейронной сетью. В 1949 году Дональд Хебб предложил механизм обучения, основанный на нейропсихологии, который стал предшественником современных методов машинного обучения [20]. «Правило обучения Хебба» – это принцип, согласно которому изменение весов между модельными нейронами увеличивается, если они активируются одновременно, и уменьшается, если они активируются по отдельности. Это одна из ранних концепций машинного обучения, которая похожа на человеческое познания. Однако прогресс в этой области замедлился в 1969 году, когда Марвин Мински и Сеймур Паперт доказали, что компьютеры того времени не обладали достаточной мощностью для реализации сложных нейронных сетей. Тем не менее, уже в 1952 году ученый IBM Артур Сэмюэл разработал программу для игры в шашки, которая могла обучаться на основе текущих позиций, что стало одним из первых примеров эволюционных вычислений [25].

Термин «искусственный интеллект» был официально введен в 1956 году на Дартмутской конференции в Дартмутском колледже в Нью-Гемпшире, организованной Марвином Мински (американский учёный в области искусственного интеллекта, сооснователь Лаборатории искусственного интеллекта в Массачусетском технологическом институте) и Джоном Маккарти (американский информатик, автор термина «искусственный интеллект», основоположник функционального программирования, лауреат премии Тьюринга за огромный вклад в область исследований искусственного интеллекта). Это событие объединило ведущих ученых, таких как Натаниэль Rochester (главный архитектор IBM 701, первого серийного научного компьютера, и прототипа его первой коммерческой версии, IBM 702. Он написал первый ассемблер и участвовал в создании области искусственного интеллекта) и Клод Шеннон (американский инженер, криптоаналитик и математик, считается «отцом информационного века»), и положило начало новой области исследований, направленной на создание машин, способных имитировать человеческий интеллект. С тех пор исследования искусственного интеллекта принесли много замечательных достижений, включая машинное обучение, доказательство теорем, распознавание образов, решение проблем, экспертные системы и обработку естественного языка.

За конференцией в Дартмуте последовал период почти двух десятилетий, в течение которых в области искусственного интеллекта был достигнут значительный успех. В 1960-х годах появились первые значимые достижения в области искусственного интеллекта. Программа ELIZA, разработанная Джозефом Вайценбаумом в Массачусетском технологическом институте, стала первым виртуальным собеседником, способным имитировать диалог терапевта с пациентом [21]. ELIZA – первая в истории компьютерных наук программа, общающаяся с человеком на естественном языке. Эта программа – важный этап в разработке технологий искусственного интеллекта.

Другой важной вехой стала программа General Problem Solver, созданная лауреатом Нобелевской премии Гербертом Саймоном и учеными корпорации RAND Клиффом Шоу и Алленом Ньюэллом, которая могла решать логические задачи, такие как Ханойские башни. Благодаря этим успешным историям, исследования в области искусственного интеллекта получили весомое финансирование, что способствовало появлению всё большего количества проектов в этой области.

Однако в 1970-х годах искусственный интеллект столкнулся с критикой – Конгресс Соединенных штатов выразил критику по поводу значительных затрат на исследования в области искусственного интеллекта. В том же году британский математик Джеймс Лайтхилл в своем отчете подверг сомнению оптимистичные прогнозы исследователей, что привело к сокращению финансирования и началу периода, известного как «зима искусственного



интеллекта». Он высказал мнение, что здравый смысл всегда будет выше способностей машин с искусственным интеллектом. Британское правительство прекращает финансовую поддержку исследований искусственного интеллекта во всех университетах, кроме трех (Эдинбург, Сассекс и Эссекс), и правительство США вскоре последовало примеру Великобритании. Несмотря на это, уже в 1980-х годах японское правительство начало активно инвестировать в исследования искусственного интеллекта, что стимулировало возобновление интереса к этой области. Управление перспективных исследовательских проектов Министерства обороны США вслед за Японией также отреагировало увеличением финансирования, но, к сожалению, никаких дальнейших успехов не было достигнуто в это время.

Одной из причин начального отсутствия прогресса в сфере искусственного интеллекта и значительного расхождения между реальностью и ожиданиями является подход, которым ранние системы, такие как ELIZA и General Problem Solver, основанные на экспертных системах – наборах правил, которые пытались формализовать человеческий интеллект. Хотя такие системы успешно справлялись с задачами, поддающимися формализации (например, шахматы), они оказались неэффективны в более сложных областях, таких как распознавание изображений или обработка естественного языка. Например, экспертную систему нельзя легко обучить распознавать лица или даже различать изображения еды или, например, животного. Для таких задач необходимо, чтобы система могла правильно интерпретировать внешние данные, учиться на таких данных и использовать эти знания для достижения конкретных целей и задач посредством гибкой адаптации – характеристик, которые определяют искусственный интеллект. Поскольку экспертные системы не обладают этими характеристиками, технически говоря, они не являются настоящим искусственным интеллектом. Статистические методы достижения настоящего искусственного интеллекта обсуждались еще в 1940-х годах, как было отмечено выше, когда Дональд Хебб разработал теорию обучения, которая воспроизводит процесс нейронов в человеческом мозге. Это привело к пониманию, что настоящий искусственный интеллект требует более гибких подходов, таких как искусственные нейронные сети.

Искусственные нейронные сети, впервые предложенные в 1940-х годах, вернулись в виде глубокого обучения. В 2015 году программа AlphaGo, разработанная Google, победила чемпиона мира в игре Го, что стало важным прорывом. Глубокое обучение, основанное на нейронных сетях, стало основой для современных приложений искусственного интеллекта. Они являются основой алгоритмов распознавания изображений, используемых Facebook¹, алгоритмов распознавания речи, которые питают умные колонки и беспилотные автомобили.

Следует подчеркнуть, что в Новое время, особенно в XX – начале XXI века, в Европе, Японии и США стремительное развитие математических и инженерно-технических знаний привело к тому, что выполнение отдельных операций и целых систем деятельности, которые на протяжении тысячелетий выполнялись людьми, стало осуществляться с помощью отдельных технических устройств – автоматов и роботов. По открытой статистике в середине второго десятилетия XXI в. в мире разрабатывалось более 100 проектов создания роботизированных систем, в том числе в Японии, США, Южной Корее, Германии, Китае, Швеции, России, Турции, Австрии, Испании, Австралии, Канаде, Иране, Италии [1].

«Искусственный интеллект» стало модным словом почти в каждом аспекте нашей жизни [28]. Применение искусственного интеллекта получило большое внимание и в литературе. Области исследований искусственного интеллекта включают системы и инженерии, науку о мозге, психологию, когнитивную науку, математику, информатику и многие другие области. Сегодня искусственный интеллект проникает во все сферы жизни, от здравоохранения до энергетики.

¹ Запрещено на территории РФ



Тот факт, что в ближайшем будущем системы искусственного интеллекта будут все больше входить в нашу повседневную жизнь, поднимает вопрос о том, необходимо ли регулирование, и если да, то в какой форме. Хотя искусственный интеллект по своей сути объективен и беспристрастен, это не означает, что системы, основанные на искусственном интеллекте, не могут быть предвзятыми.

Аналогично тому, как автоматизация производственных процессов привела к потере рабочих мест, растущее использование искусственного интеллекта приведет к снижению потребности в служащих и даже в высококвалифицированных профессиональных рабочих местах. Например, инструменты распознавания изображений уже превосходят врачей в выявлении рака кожи, а в юридической профессии технологии электронного обнаружения сократили потребность в больших командах юристов и помощников юристов для изучения миллионов документов. Конечно, в прошлом наблюдались значительные сдвиги на рынках труда (например, в контексте промышленной революции 1820-1840 гг.), но не очевидно, будут ли обязательно созданы новые рабочие места в других областях для размещения этих сотрудников. Это связано как с количеством возможных новых рабочих мест (которое может быть намного меньше количества потерянных рабочих мест), так и с требуемым уровнем квалификации.

Следует также обратить внимание на научно-фантастическую литературу, созданную в конце XX века и в первых десятилетиях XXI века, где авторы поднимают вопросы о потенциальной угрозе, которую представляют роботы для человечества. Если технические устройства смогут полностью воспроизводить интеллектуальные процессы и функции человека, это может привести к тому, что люди утратят свою значимость в различных сферах и станут излишними, уступая роботам в ряде аспектов.

Интересно, что подобно тому, как художественная литература может рассматриваться как отправная точка искусственного интеллекта, ее также можно использовать, чтобы получить представление о том, как может выглядеть мир с большей безработицей. В фантастическом романе «Лавина» («Snow Crash»), опубликованном американским писателем Нилом Стивенсоном, описывается мир, в котором люди проводят свою физическую жизнь в складских помещениях, окруженных техническим оборудованием, в то время как их фактическая жизнь происходит в трехмерном мире, называемом Метавселенной, где они появляются в виде трехмерных аватаров. Как бы фантастично ни звучал этот сценарий, последние достижения в области обработки изображений виртуальной реальности в сочетании с прошлым успехом виртуальных миров (и тем фактом, что более высокая безработица приводит к меньшему располагаемому доходу) делают альтернативные формы развлечений менее доступными и делают этот сценарий далеким от утопии.

Приведенное выше исследование ясно показывает, что искусственный интеллект уже стал частью нашей повседневной жизни, и его влияние будет только расти, как Интернет или социальные сети в прошлом. При этом искусственный интеллект меняет не только личную жизнь людей, но и способы взаимодействия компаний с клиентами и партнерами. И вопрос заключается не в том, будет ли искусственный интеллект играть роль в будущем, а в том, какую именно роль он будет играть и как люди смогут сосуществовать с этими технологиями. Решения, которые будут приниматься в ближайшие годы, определят, как искусственный интеллект повлияет на общество и экономику.

Философские основы искусственного интеллекта: знание, сознание и ответственность

Философия и искусственный интеллект с момента зарождения последнего находятся в тесной взаимосвязи, что порождает необходимость их систематического изучения. Утверждения о том, что философия является близким и надежным союзником исследователей искусственного интеллекта или, что большая часть искусственного интеллекта строится на



работах философов, часто встречаются как в философской, так и в научной литературе. Суть этих подходов заключается в центральной роли, которую философия играет по отношению к искусственному интеллекту: философия оказывает влияние на искусственный интеллект как с исторической, так и с методологической точки зрения.

Философия играет важную роль для искусственного интеллекта в прояснении его целей и методов, в тоже время искусственный интеллект предлагает философии мощные инструменты для ответа на несколько различных вопросов. Однако систематическая структура все еще отсутствует и несколько вкладов сосредоточены исключительно только на частичных аспектах проблемы, несмотря на различные попытки исчерпывающего исследования сложной природы этого взаимодействия.

До появления вычислительных машин вопросы о природе мышления и менталитета были преимущественно предметом философских изысканий. Одним из наиболее влиятельных мыслителей в этой области был Рене Декарт (1596-1650), которого часто называют «отцом современной философии». Декарт придерживался дуалистической концепции, согласно которой разум и тело представляют собой взаимоисключающие категории. Он утверждал, что разум – это способность мыслить, доступ к которой возможен через интроспекцию, то есть внутреннее восприятие собственных психических состояний.

Идеи Декарта доминировали вплоть до XX века, когда развитие цифровых компьютеров привело к поиску более научного и объективного понимания мышления. Важнейший вклад в эту область внес Алан Тьюринг (1912-1954) – британский математик, логик и криптограф, оказавший существенное влияние на развитие информатики. Тьюринг исследовал границы вычислимости, предположив, что задачи, которые могут быть решены с помощью логических правил, также могут быть решены с использованием машин, предназначенных для решения проблем.

Искусственный интеллект можно рассматривать как научную дисциплину, а также как устройства, способные моделировать человеческое мышление. Современный искусственный интеллект охватывает широкий спектр вопросов, включая теорию восприятия, методы игры в шахматы и шашки, доказательство математических теорем и диагностику заболеваний. Исследователи искусственного интеллекта применяют его методы в различных областях, а ученые из других дисциплин находят в нем основу для систематизации интеллектуальных задач. Таким образом, ИИ представляет собой универсальную область знаний.

С момента появления компьютеров и развития информационных технологий философы активно изучали вопросы, связанные с искусственным интеллектом. В середине и второй половине XX в. значительный вклад в понимание сходства компьютера и человека внесли такие зарубежные и российские исследователи, как Д. Маккарти [22], А. Тьюринг [27], Дж. фон Нейман [12], Н. Винер [6], Дж. Сёрль [26], Э. В. Ильенков [10] и др. Именно в работах Д. Маккарти появился термин «искусственный интеллект» и было дано его определение: свойство интеллектуальных систем выполнять творческие функции, которые традиционно считаются прерогативой человека [22].

Один из способов понять философию искусственного интеллекта заключается в том, что она в основном имеет дело с тремя кантовскими вопросами: Что такое искусственный интеллект? Что может делать искусственный интеллект? Каким должен быть искусственный интеллект? Традиционно философия искусственного интеллекта имеет дело с несколькими выбранными моментами, где философы нашли, что сказать об искусственном интеллекте, например, о тезисе, что познание – это вычисление или, что компьютеры могут иметь значимые символы. Обзор этих моментов и соответствующих авторов приведет к фрагментарному обсуждению, которое никогда не достигнет картины общего проекта.



Отношения между машинами и людьми – это еще одна фундаментальная философская основа, требующая максимального внимания. Растущая сложность и возможности систем искусственного интеллекта, особенно в областях, где они полностью превосходят людей, делают отношения чрезвычайно чувствительными. Эта производительность машин принципиально зажигает вопросы для обоснования роли людей в мире, где доминируют машины. Это заставляет нас, людей, думать о наших человеческих ценностях, их важности и уместности в управлении миром. Такие ценности, как человеческое творчество и эмпатия, не только подвергаются сомнению, но и требуют реального обоснования их важности сегодня. Это открывает различные вопросы об этике искусственного интеллекта.

Знание и интеллект проявляют очень чувствительную связь, которая требует критического философского анализа. Эта связь обеспечивает еще одну философскую основу искусственного интеллекта для понимания смысла или возможностей искусственного интеллекта обрабатывать огромные объемы информации, не имея возможности знать что-либо. Таким образом она порождает вопросы типа «Что значит для искусственного интеллекта действительно «знать» что-то»? Это также очень сложная исследовательская задача, особенно проблема представления знаний – как знания могут быть представлены в компьютере. В то время как некоторые философы все еще утверждают, что системы, управляемые искусственным интеллектом, никогда не смогут по-настоящему представлять знания в том же смысле, что и люди, другие полагают, что аспект представления знаний в искусственном интеллекте может быть развит в реальном мире истинного понимания и рассуждения. Также пока остается под вопросом, как искусственный интеллект может превзойти людей без четкого истинного понимания представления знаний.

Другим философским основанием искусственного интеллекта является вопрос сознания. Этот вопрос заставляет некоторых философов верить, что искусственный интеллект никогда не сможет быть сознательным. Другие философы утверждают, что вполне возможно создать сознание в машинах. В то время как люди еще не обосновали источник своего сознания, нельзя исключить тот факт, что машины не могут быть сознательными. Этот загадочный феномен сознания был предметом длительных споров философов на протяжении столетий, и эти споры еще не закончены. Что усложняет спор, так это нечеткое различие и связь между сознанием и интеллектом.

Одной из ключевых проблем является связь между знанием и интеллектом. Искусственный интеллект способен обрабатывать огромные объемы информации, но может ли он действительно «знать» что-то? Этот вопрос остается открытым и вызывает споры среди философов. Некоторые утверждают, что системы искусственного интеллекта никогда не смогут представлять знания так, как это делают люди, в то время как другие считают, что развитие представления знаний в искусственном интеллекте может привести к настоящему пониманию и рассуждению [8].

Вопрос ответственности за решения, принимаемые искусственным интеллектом, также вызывает серьезные дискуссии. Даже современные системы искусственного интеллекта, не обладающие сознанием, способны совершать действия с неоднозначными последствиями. Согласно классификации Дж. Сёрля, выделяются «сильный» и «слабый» искусственный интеллект. Если «слабый» искусственный интеллект действует в рамках заданной программы, то ответственность за его действия лежит на человеке, который создал и запрограммировал систему. Однако в случае создания «сильного» искусственного интеллекта, способного к самостоятельному мышлению, вопрос ответственности становится гораздо более сложным и может касаться выживания человечества [16]. Некоторые ученые даже высказывают предположения, что наделение искусственного интеллекта субъектностью может быть способом снять ответственность с человека в случае негативных последствий.



Термины «слабый» и «сильный» искусственный интеллект, введенные Дж. Сёрлем, помогают разграничить подходы к изучению искусственного интеллекта. «Слабый» искусственный интеллект рассматривает компьютер как инструмент для симуляции разумного поведения, тогда как «сильный» искусственный интеллект предполагает, что правильно запрограммированный компьютер может обладать сознанием и когнитивными состояниями. Эти концепции продолжают влиять на философские и научные дискуссии о природе искусственного интеллекта [16].

С развитием технологий философия искусственного интеллекта ставит новые вопросы, которые становятся все более актуальными. В настоящее время ни у кого не вызывает сомнений, что вокруг человека увеличивается количество роботов, а также роботизированных и автоматизированных систем. В будущем это число продолжит расти, например, благодаря внедрению беспилотных такси, роботов-уборщиков, курьеров, домашних питомцев, продавцов и компаньонов. Поэтому философия искусственного интеллекта начинает ставить вопросы, которые окажутся актуальными в ближайшее время. Например, будут ли у искусственных существ права, как будут строиться их отношения с людьми и можно ли считать автономные машины личностями? Эти вопросы важны как для «сильного», так и для «слабого» искусственного интеллекта. Хотя большинство исследователей сходятся во мнении, что системы искусственного интеллекта не могут быть личностями, сам факт постановки таких вопросов способствует развитию философского осмысления проблем искусственного интеллекта.

Философия искусственного интеллекта представляет собой активно развивающееся направление в философии, находящееся под воздействием современных научных достижений и технологий. Она активно откликается на новшества в области своего изучения – искусственном интеллекте. В этом контексте исследователи отмечают возникновение более узкой ветви данной философии – философии нейронных сетей, которые составляют основу искусственного интеллекта [7]. Искусственные нейронные сети создаются по аналогии с естественными, воспроизводя и моделируя процессы, имеющие место в человеческом мозге, особенно в нейронах. Их функционирование организовано на базе искусственного интеллекта и в значительной мере отражает человеческое мышление, однако по-прежнему не приводит к разработке «сильного» искусственного интеллекта. В последние годы сложность искусственных нейронных сетей значительно увеличивается, а логика принимаемых ими решений становится всё более комплексной.

Философия искусственного интеллекта – это динамично развивающаяся область, которая ставит перед человечеством сложные вопросы о природе знания, сознания и ответственности. Развитие искусственного интеллекта требует не только технологических инноваций, но и глубокого философского анализа, который поможет нам понять, как взаимодействовать с искусственным разумом в будущем.

Заключение

Искусственный интеллект – одна из самых обсуждаемых тем в современном мире. Его влияние на общество, экономику и культуру вызывает как оптимизм, так и опасения. Рэймонд Курцвейл из Google считает, что искусственный интеллект может стать инструментом для улучшения человеческого интеллекта, в то время как Илон Маск предупреждает о потенциальных рисках, включая возможность Третьей мировой войны. Независимо от того, какой сценарий реализуется, очевидно, что искусственный интеллект ставит перед нами уникальные этические, правовые и философские проблемы, которые требуют немедленного решения.

Одной из ключевых этических проблем, связанных с искусственным интеллектом, является проблема вагонетки – мысленный эксперимент, в котором человек должен выбрать



между бездействием, приводящим к гибели многих, и активным действием, которое спасает большинство, но жертвует меньшинством. В контексте беспилотных автомобилей эта проблема становится реальной: машины и их программисты должны будут принимать решения, которые могут повлиять на жизнь людей.

Регулирование искусственного интеллекта – еще одна сложная задача. Технологии развиваются настолько быстро, что даже эксперты не всегда успевают за ними, не говоря уже о политиках. Марк Цукерберг и другие лидеры технологической индустрии призывают к созданию регуляторных рамок, но как регулировать то, что постоянно меняется? Одним из возможных решений может быть подход, предложенный судьей Верховного суда США Поттером Стюартом в 1964 году. Он определил непристойность, сказав: «Я узнаю это, когда увижу». Этот подход может быть применен и к искусственному интеллекту, где регулирование должно быть достаточно гибким, чтобы допускать будущие инновации, но достаточно точным, чтобы избежать злоупотреблений.

Искусственный интеллект также поднимает ряд фундаментальных философских вопросов. Среди них – природа сознания, свобода воли, выравнивание ценностей и проблема предвзятости. Например, если искусственный интеллект принимает решения на основе данных, которые содержат предвзятость, это может привести к дискриминации и несправедливости.

Философы, ученые и инженеры должны работать вместе, чтобы решить эти проблемы. Важно, чтобы технологии искусственного интеллекта разрабатывались и внедрялись таким образом, чтобы они приносили пользу обществу, а не создавали новые проблемы. Это требует междисциплинарного подхода, где философия играет ключевую роль в осмыслении последствий технологического прогресса.

Внедрение искусственного интеллекта в повседневную жизнь имеет глубокие социальные и психологические последствия. Социологам предстоит изучить, как искусственный интеллект повлияет на рынок труда и проблему безработицы. Уже сейчас автоматизация приводит к сокращению рабочих мест в некоторых отраслях, и этот процесс будет только ускоряться.

Психологи, в свою очередь, должны исследовать, как искусственный интеллект влияет на самооценку и идентичность человека. Например, взаимодействие с умными помощниками и роботами может изменить наше представление о себе и своих возможностях. Эти изменения могут быть как положительными, так и отрицательными, и важно понимать их природу, чтобы минимизировать негативные последствия.

Философия искусственного интеллекта становится все более важным направлением современных гуманитарных исследований. Она объединяет в себе вопросы этики, философии сознания, социальной философии и других дисциплин. В каждом из этих направлений есть своя специфика, но все они связаны с осмыслением того, как искусственный интеллект меняет нашу жизнь и общество.

Например, философы исследуют, может ли искусственный интеллект обладать сознанием или свободой воли. Эти вопросы не только теоретические, но и имеют практическое значение. Если искусственный интеллект когда-нибудь достигнет уровня, сравнимого с человеческим сознанием, это потребует пересмотра многих этических и правовых норм.

Искусственный интеллект – это не просто технология, это явление, которое меняет наше общество, культуру и даже наше понимание самих себя. Этические, философские и социальные проблемы, связанные с искусственным интеллектом, требуют глубокого осмысления и междисциплинарного подхода. Регулирование искусственного интеллекта должно быть гибким, чтобы учитывать быстрые изменения в технологиях, но при этом достаточно строгим, чтобы предотвратить злоупотребления.



Философия искусственного интеллекта становится важным инструментом для понимания этих изменений. Она помогает нам не только осмыслить текущие проблемы, но и предвидеть будущие вызовы. В конечном итоге, успешное внедрение искусственного интеллекта в общество зависит от того, насколько хорошо мы сможем решить эти вопросы и направить технологии на благо человечества.

Список литературы:

1. Бабкина В. А., Создание и реализация дистанционных образовательных программ: опыт университетов стран мира и России: коллективная монография / С.Н. Колоцей, Е.М. Ключева, А.С. Леменкова, Т.В. Починок, А.С. Ясюченя и др.; отв. ред.: Е.Л. Кабахидзе, А.Ю. Сувинова, Джеймс Хайнс. – Москва: МГПУ, 2018. – 164 с.
2. Баррат Дж. Последнее изобретение человечества: Искусственный интеллект и конец эры Homo sapiens. – М.: Альпина нон-фикшн, 2015. – 304 с.
3. Беликова Е.К. Основные вопросы философии искусственного интеллекта // Философия и культура. – 2024. – №1. URL: https://nbpublish.com/library_read_article.php?id=69543 (дата обращения: 03.12.2025)
4. Бриньолфсон Э. Вторая эра машин. Работа, прогресс и процветание в эпоху новейших технологий / Макафи Э. – М.: АСТ, 2017. – 384 с. (28)
5. Брокман Д. Что мы думаем о машинах, которые думают: Ведущие мировые учёные об искусственном интеллекте. – М.: Альпина нон-фикшн, 2017. – 548 с.
6. Винер Н. Кибернетика, или Управление и связь в животном и машине. М.: Советское радио, 1968. – 328 с.
7. Воробьев А. В., История философии нейронных сетей как ядра искусственного интеллекта / Кудинов В. А. // Проблемы онто-гносеологического обоснования математических и естественных наук. – Курск: Курский государственный университет, 2021, №12. – С. 17–27
8. Георгиу Т. С. Философия автоматизации и искусственного интеллекта: от мифологического Талоса до будущих киборгов // Вестник Московского государственного областного университета. Серия: Философские науки. – М. обл.: Московский государственный областной университет, 2022, № 1. – С. 68–75.
9. Дубровский Д. И. Фетиш искусственного интеллекта / Ефимов А. Р., Лепский В. Е., Славин Б. Б. // Философские науки. – М.: Академия гуманитарных исследований, 2022, № 1. – С. 44–71.
10. Ильенков Э. В. Философия и культура. М.: Политиздат, 1991. – 464 с.
11. Мамина Р. И. Искусственный интеллект в оптике философской методологии: образовательный трек / Почебут С. Н. // ДИСКУРС. – СПб.: СПбГЭТУ «ЛЭТИ», 2022. № 1. – С. 64-81
12. Нейман Дж. фон. Вычислительная машина и мозг. М.: АСТ, 2022. – 192 с.
13. Ракитов А. И. Философия, роботы, автоматы и зримое будущее // Философия и общество – Волгоград: Учитель, 2019, №3. – С. 35-48
14. Толкачев С.А., Роботизация как направление неоиндустриализации (на примере США) / Кулаков А.Д. // Мир новой экономики М.: Финансовый университет при Правительстве Российской Федерации – , 2016, № 2. С. 79–87
15. Artificial Intelligence (AI) Market // MarketsandMarkets – 2024 – URL: <https://www.marketsandmarkets.com/Market-Reports/artificial-intelligence-market-74851580.html> (дата обращения: 21.12.2025)
16. Bishop J. M. Artificial Intelligence Is Stupid and Causal Reasoning Will Not Fix It // Frontiers in Psychology, Switzerland – 2021, №11. P. 1-18



17. Bruneault F. AI Ethics: how can information ethics provide a framework to avoid usual conceptual pitfalls? An Overview / Laflamme A.S. // AI & SOCIETY, Springer-Verlag London Ltd., part of Springer Nature – 2020, №36 (3). P. 1-10.
18. Grosz B. J. A Century Long Commitment to Assessing Artificial Intelligence and its Impact on Society / Stone P. // Communications of the ACM, New York – 2018, №12. P. 68-73
19. Haenlein M. A Brief History of Artificial Intelligence: On the Past, Present, and Future of Artificial Intelligence / Kaplan A. // CALIFORNIA MANAGEMENT REVIEW, University of California –, 2019, №61 (4). P. 5-14
20. Hebb D. O. The Organization of Behavior A Neuropsychological Theory. New York: JOHN WILEY & SONS, Inc. London CHAPMAN & HALL, Limited, 1949. – 365 p.
21. Killgrove K. “ELIZA” the world's 1st chatbot, was just resurrected from 60-year-old computer code // Live Science – 2025 – URL: <https://www.livescience.com/technology/eliza-the-worlds-1st-chatbot-was-just-resurrected-from-60-year-old-computer-code> (дата обращения: 01.01.2026)
22. McCarthy J. Recursive Functions of Symbolic Expressions and Their Computation by Machine, Part I // Communications of the ACM, New York – 1960. – №4. P. 184–195.
23. Nascimento A.M. Artificial intelligence and industry 4.0: The next frontier in organizations // SciELO Brazil – 2018 – URL: <https://www.scielo.br/j/bar/a/XTxRc4gcytpcBjPnhTWpBPf/?lang=en> (дата обращения: 21.11.2025)
24. Omohundro S. Rational artificial intelligence for the greater good // The frontiers collection, Berlin – 2013. – P. 161-179.
25. Samuel A. Early Machine Learning and Checkers Program // Robots Authority – 2024 – URL: <https://robotsauthority.com/arthur-samuel-early-machine-learning-and-checkers-program/> (дата обращения: 25.12.2025)
26. Searle J. R. Minds, brains and programs // Behavioral and Brain Sciences, Cambridge University Press –1980. – № 3. P. 415–431.
27. Turing A. M. Computing Machinery and Intelligence // Mind, New Series, Oxford University Press – 1950. – № 236. P. 433–460.
28. Yuchen Jiang Quo vadis artificial intelligence / Xiang Li, Hao Luo, Shen Yin, Okayay Kaynak // Springer Nature Link, – 2022. – URL: <https://link.springer.com/article/10.1007/s44163-022-00022-8> (дата обращения: 04.12.2025) .

