

НЕКОТОРЫЕ АСПЕКТЫ РАССМОТРЕНИЯ СОЗДАНИЯ НАУКИ ОБ ИСКУССТВЕННОМ ИНТЕЛЛЕКТЕ



<https://doi.org/10.24412/2181-1784-2025-24-12-19>

Науменко Олег Александрович,

кандидат философских наук,
доцент кафедры социальных наук и технологий
НИТУ МИСИС
Москва, Россия

***Аннотация.** В статье рассматриваются ключевые философские, методологические и этические аспекты, связанные с формированием искусственного интеллекта (ИИ) как самостоятельной научной дисциплины. Анализируются исторические этапы развития ИИ, современные вызовы и перспективы, а также проблемы, возникающие при попытках создания "сильного" ИИ, обладающего сознанием и способностью к самообучению. Особое внимание уделяется вопросам определения критериев "интеллектуальности" машин, проблемам ответственности за действия ИИ и возможному влиянию ИИ на будущее человечества.*

***Ключевые слова.** Искусственный интеллект, наука об ИИ, философия ИИ, этика ИИ, сознание, самообучение, ответственность, методология ИИ.*

SOME ASPECTS OF CONSIDERING THE CREATION OF ARTIFICIAL INTELLIGENCE SCIENCE

Oleg A. Naumenko,

PhD, Associate Professor of the
Social Sciences and Technology Department,
NUST MISIS
Moscow, Russia

***Abstract.** The article examines the key philosophical, methodological and ethical aspects related to the formation of artificial intelligence (AI) as an independent scientific discipline. The article analyzes the historical stages of AI development, modern challenges and prospects, as well as the problems that arise when trying to create a "strong" AI with consciousness and the ability to self-learn. Special attention is paid to the issues of defining criteria for the "intelligence" of*

machines, the problems of responsibility for AI actions and the possible impact of AI on the future of mankind.

Keywords. *Artificial intelligence, AI science, philosophy and ethics of AI, consciousness, self-learning, responsibility, AI methodology.*

Под «искусственным интеллектом» в первую очередь подразумевается «метафорическое понятие для обозначения системы созданных людьми средств, воспроизводящих определенные функции человеческого мышления». Само понятие «искусственный интеллект» по-разному трактуется в литературе. Однако все же можно выделить два главных значения этого понятия: 1) актуальное, важнейшее междисциплинарное, комплексное исследование данного феномена; 2) моделирование и имитация человеческого интеллекта, его психических и мыслительных процессов в компьютере, а также создание «электронного мозга», технического устройства, который в принципе будет превосходить человеческий интеллект. В литературе обычно называются следующие характерные черты компьютерной системы, «искусственного интеллекта» [1, С. 5-11]:

- способность решать сложные задачи, хранить большие объемы информации и работать с ними;
- наличие в них собственной внутренней модели внешнего мира; эта модель обеспечивает индивидуальность, относительную самостоятельность системы в оценке ситуации, возможность семантической и прагматической интерпретации запросов к системе;
- способность пополнения имеющихся знаний;
- способность к дедуктивному выводу, то есть к генерации информации, которая в явном виде не содержится в системе; это качество позволяет системе конструировать информационную структуру с новой семантикой и практической направленностью;
- умение оперировать в ситуациях, связанных с различными аспектами нечеткости, включая «понимание» естественного языка;
- способность к диалоговому взаимодействию с человеком;
- способность к адаптации.

Естественно было бы предположить, что создание науки об искусственном интеллекте станет насущной проблемой. Но и здесь первые мысли по этому поводу уже относились к эпохе Нового времени.

На саму возможность мыслить о понятии «искусственный интеллект» огромное влияние оказало рождение механического материализма, которое начинается с работы уже упоминаемого нами великого француза Р.Декарта

«Рассуждение о методе» (1637 г.) и сразу вслед за этим работы Т.Гоббса «Человеческая природа» (1640 г.). Декарт предположил, что животное – некий сложный механизм, тем самым сформулировав механистическую теорию.

И тут важно понимать, чем отличается именно механистический материализм, взгляды которого запечатлены в работах Аристотеля, и последующей диалектики Г.Гегеля, диалектического и исторического материализма в лице Л.Фейербаха, К.Маркса, Ф.Энгельса, В.Ленина. Дело в том, что механистический материализм направлен на механистическое происхождение организмов, в то время как античный материализм направлен на механистическое происхождение природы, а диалектический и исторический материализм относится к проявлениям механизма в обществе.

Поэтому понятно, что без понимания механистичности в организмах не могла идти речь о понимании искусственного интеллекта даже в самом примитивном смысле, а наличие механистичности природы и общества выходят за рамки области об искусственном интеллекте, и строго говоря не являются необходимыми предпосылками.

Также интересно отметить технологические предпосылки к возникновению науки об искусственном интеллекте. Так, в 1623 г. В.Шикард построил первую механическую цифровую вычислительную машину, за которой последовали машины Б.Паскаля (1643 г.) и Г.Лейбница (1671 г.). Лейбниц также был первым, кто описал современную двоичную систему счисления, хотя до него этой системой периодически увлекались многие великие ученые. Также выдающийся немецкий математик и философ мысленный эксперимент, описанный им в «Монадологии». Он предлагает представить машину размером с мельницу («мельница Лейбница»), которая симулирует чувства, мысли и восприятие. И если зайти внутрь такой машины, то ни одна из движущихся частей, обеспечивающих её работу, не будет являться сознанием или объяснением восприятия. Так Лейбниц пытался показать, что одних физических состояний недостаточно для функционирования сознания.

В России тоже проводились некоторые изыскания в области изучения возможности некоторых проявлений искусственного интеллекта. В 1832 г. коллежский советник С.Н.Корсаков выдвинул принцип разработки научных методов и устройств для усиления возможностей разума и предложил серию «интеллектуальных машин», в конструкции которых, впервые в истории информатики, применил перфорированные карты [2]. В XIX веке Ч.Бэббидж

и А.Лавлейс работали над программируемой механической вычислительной машиной [3].

Непосредственно рождение науки об искусственном интеллекте приходится на начало XX в. В 1910-1913 гг. Б.Рассел и А.Н.Уайтхэд опубликовали работу «Принципы математики», которая произвела революцию в формальной логике. В 1941 г. К.Цузе построил первый работающий программно-управляемый компьютер. В дальнейшем У.Маккалок и У.Питтс в 1943 г. опубликовали работу «Логическое исчисление идей, присущих нервной деятельности», который заложил основы нейронных сетей. Они предложили понятие «искусственной нейронной сети». Д.Хебб в своей работе «Организация поведения» (1949 г.) описал основные принципы обучения нейронов. Несколько лет спустя эти идеи были развиты американским нейрофизиологом Ф.Розенблаттом. Он предложил схему устройства, которое может моделировать процесс человеческого восприятия и назвал его «перцептроном» [4].

Советский Союз не отставал в этой гонке по созданию интеллектуальных машин. Среди советских учёных искусственный интеллект был главной областью научной деятельности Д.А.Поспелова. Здесь научные интересы Д. А. Поспелова связаны с моделированием поведения человека, формализацией рассуждений, общими проблемами моделирования жизненных процессов в естественных и искусственных системах. В частности, им был впервые в мире разработан подход к принятию решений, опирающийся на семиотические модели, который послужил теоретической основой ситуационного управления большими системами [4].

А что же философы? Неужели остались в стороне от всей этой канители с искусственным интеллектом? Философия сразу же критически осмыслила все эти новые теоретические разработки. Созданная на злобу дня философия искусственного интеллекта задаётся вопросами о «мышлении машин», причем вопросы эти отражают интересы различных исследователей искусственного интеллекта, философов, исследователей познавательной (когнитивной) деятельности. Ответы на эти вопросы зависят от того, что понимается под понятиями «интеллект» или «сознание», и какие именно «машины» являются предметом обсуждения.

Ну и как можно было бы предположить наиболее горячие споры в философии искусственного интеллекта вызывает вопрос возможности мышления творения человеческих рук. Вопрос «Может ли машина мыслить?», который подтолкнул исследователей к созданию науки о

моделировании человеческого разума, был поставлен еще А.Тьюрингом в 1950 г. Две основных и наиболее популярных точек зрения на этот вопрос носят названия гипотез сильного и слабого искусственного интеллекта.

Термин «сильный искусственный интеллект» ввел Джон Сёрл, его же словами подход и характеризуется: «Более того, такая программа будет не просто моделью разума; она в буквальном смысле слова сама и будет разумом, в том же смысле, в котором человеческий разум – это разум» [5].

В противовес этому подходу, сторонники гипотезы слабого искусственного интеллекта предпочитают рассматривать программы лишь как инструмент, позволяющий решать те или иные задачи, которые не требуют полного спектра человеческих познавательных способностей.

В своем мысленном эксперименте под названием «Китайская комната», Джон Сёрл показывает, что даже прохождение знаменитого теста Тьюринга может не являться достаточным критерием наличия у машины подлинного процесса мышления. Отсюда следует, что мышление есть процесс обработки находящейся в памяти информации: анализ, синтез и самопрограммирование. Точно такую же позицию занял и американский ученый Р.Пенроуз, который в своей книге «Новый ум короля» аргументирует невозможность получения процесса мышления на основе формальных систем.

Итак, теория сильного искусственного интеллекта предполагает, что компьютеры могут приобрести способность мыслить и осознавать себя как отдельную личность (в частности, понимать собственные мысли), хотя и не обязательно, что их мыслительный процесс будет подобен человеческому. Теория слабого искусственного интеллекта отвергает такую возможность.

В философской литературе, касательно рассуждений о понятии «искусственного интеллекта» часто упоминается дружественный искусственный интеллект. Он определяется как гипотетический тип искусственного интеллекта, не оказывающий негативного влияния на человека. Многие концепции дружественного искусственного интеллекта включают утверждение о том, что он не только не будет приносить вреда человечеству, но и будет оказывать материально-информационную поддержку людям, вплоть до полного обеспечения желаний и потребностей каждого отдельно взятого человека. «Дружественный» в данном контексте является техническим термином, который не означает, что искусственный интеллект обязательно будет испытывать к человечеству чувства привязанности и близости.

Чтобы окончательно определиться с возможностями искусственного интеллекта, в философии необходимо дать определение непосредственно

понятию «интеллект». Существуют разные точки зрения на этот вопрос. Аналитический подход предполагает анализ высшей нервной деятельности человека до низшего, неделимого уровня (функция высшей нервной деятельности, элементарная реакция на внешние раздражители (стимулы), раздражение синапсов совокупности связанных функцией нейронов) и последующее воспроизведение этих функций. Существует определение интеллекта как способности решать интеллектуальные задачи. Здесь под интеллектуальной задачей понимается та задача, у которой не существует известного алгоритма решения. То есть задача, для которой нужно создать алгоритм с нуля. Например, доказательство недоказанной теоремы, научное открытие, художественная деятельность и так далее. Некоторые специалисты за интеллект принимают способность рационального, мотивированного выбора, в условиях недостатка информации. То есть интеллектуальной просто считается та программа деятельности, которая сможет выбрать из определённого множества альтернатив, например, куда идти в случае «налево пойдёшь ...», «направо пойдёшь ...», «прямо пойдёшь ...».

Ну и, наконец, закономерным шагом в развитии философских размышлений об искусственном интеллекте стало внедрение концептов этого понятия в специальную философскую науку – эпистемологию – науку о достоверном знании в рамках философии. Философы, занимающиеся данной проблематикой, решают вопросы, схожие с теми, которые решаются инженерами искусственного интеллекта о том, как лучше представлять и использовать знания и информацию.

Самая сложная проблематика в рамках философского исследования проблем искусственного интеллекта – этическая. Современными философами предлагается ряд этических вопросов, перед которыми предстоит остановиться разработчикам машин с алгоритмом свободного действия. К ним относятся: если в будущем машины смогут рассуждать, осознавать себя и иметь чувства, то что тогда делает человека человеком, а машину – машиной?; если в будущем машины смогут осознавать себя и иметь чувства, возможно ли будет их эксплуатировать или придется наделять их правами?; если в будущем машины смогут рассуждать, то как сложатся отношения людей и машин? Данный вопрос был не раз рассмотрен в произведениях искусства на примере противостояния людей и машин. К примеру, в романе «Дюна» Фрэнка Герберта, в романе Дэна Симмонса «Гиперион», в кинофильме «Терминатор» и т.д.; будет ли человек, которому в результате многочисленных медицинских имплантаций заменили 99 процентов тела, считаться машиной?

Все же необходимо отметить, что человек существо не просто сознательное, но также и бессознательное; не просто рациональное, но также и иррациональное. Становление человеческого «Я» происходит в течение длительного времени с формированием в нем личности путем опыта и познания. Следует сказать, что лишь человек способен «экзистировать», рефлексировать, отделять себя от окружающей среды. Не обладая биологической специализированностью [6, с.3], он воспринимает действительное во всем его многообразии и противоречивости. Сам человек существо «сугубо диалектичное», но кроме того, что он, по сравнению с другими биологическими видами, являясь «недостаточным существом», также является «существом духовным», ибо он обладает не просто какой-то совокупностью морально-нравственных ценностей и установок, мнений и суждений, но в нем живет такой неоднозначный феномен, как вера. Лишь человек стремится познать истину своего бытия и может рассматривать это через призму трансцендентного и не всегда объяснимого путем разумного осмысления. Потому мы можем задаться вопросом: а сможет ли искусственный интеллект по ходу своего совершенствования когда-либо постигнуть духовность, подобно человеку? Сможет ли он не просто осознать себя как личность и вместе с тем экзистировать подобно человеку?

Здесь следует отметить и другую сторону вопроса. Человек как творец искусственной среды – культуры, является творцом и важнейшего средства коммуникации – языка, через который он стремится к изложению своих мыслей, чувств и переживаний. Однако же он также способствует отчуждению человека, ибо бедность языка не позволяет выразить весь объем действительных переживаний индивидов. В современном интернет-пространстве, где активно применяются возможности искусственного интеллекта, информационные системы осуществляют сложный анализ и обработку введенных пользователем языковых данных. Они пытаются подобрать наиболее оптимальные варианты ответов на запрос, установить корреляцию между запрашиваемой и имеющейся информацией, однако же механика, даже совершенствующаяся, не может все еще выйти за пределы механики, ибо для нее также язык есть отчуждение.

Таким образом, философская проблематика в изучении искусственного интеллекта не только не имеет строго очерченного круга исследований, но и постоянно расширяется, вместе с расширением знаний об искусственном интеллекте. Исторически сложившиеся знания людей об искусственных механизмах, способных в отдельной деятельности заменить человека, а главное самому стать творцом новой жизни превратились в умения создавать

нечто новое, доселе небывалое – одушевленную неживую материю. В дальнейшем все эти знания были облачены в доспехи научных теорий, результатом которых и стало создание теории об «искусственном интеллекте». Что же нам ждать от полномасштабного освоения киберпространства? Атаки «клонов», терминаторов или все-таки дружественных отношений между мыслящими машинами и человечеством, как бы это парадоксально ни звучало. Ответить на этот вопрос, видимо, предстоит уже в недалеком будущем.

ИСПОЛЬЗОВАННАЯ ЛИТЕРАТУРА (REFERENCES)

1. Орешников И.М., Шкерина Т.И. Философские размышления о проблеме искусственного интеллекта // «Философия образования», № 4, 2017 г. – С. 5-11
2. Михайлов А.С. Изобретения С.Н.Корсакова // <http://www.ict.nsc.ru/jspui/bitstream/ICT/913/1/kosakov.pdf> - дата обращения 20.04.2025 г.
3. Падуа С. Невероятные приключения Лавлейс и Бэббиджа. (Почти) правдивая история первого компьютера. – М.: Манн, Иванов и Фербер, 2018 г. – 302 с.
4. Толковый словарь по искусственному интеллекту / Авторы-составители А.Н. Аверкин, М.Г. Гаазе-Рапопорт, Д.А. Поспелов. М.: Радио и связь, 1992. - 256 с.
5. Искусственный интеллект: различные взгляды на проблему // В мире науки. (Scientific American. Издание на русском языке). 1990. № 3 / <http://www.raai.org/library/books/sirl/ai.htm> - дата обращения 20.04.2025 г.
6. Андрюшина Л.В. Человек как культурное животное в немецкой антропологии начала XX века // Ученые записки Орловского государственного университета. Серия: Гуманитарные и социальные науки. 2014. №5.
7. Лекторский В.А. Возможны ли науки о человеке? // Вопросы философии. 2015. № 5. С. 3-15.
8. Гелен А. О систематике антропологии. – Москва: Прогресс, 1988г.