УДК 004.8

Философские проблемы создания и использования искусственного интеллекта

Колесников Сергей Викторович
Приамурский государственный университет имени Шолом-Алейхема
Магистрант

Аннотация
Целью данной статьи рассмотреть философские проблемы создания и применения искусственного интеллекта. В результате исследования были выявлены ряд философских проблем, препятствующих созданию искусственного интеллекта на современном этапе.
Ключевые слова: искусственный интеллект, кибернетика, язык программирования, сознание, нейронная сеть, робототехника, киберэпоха.

Philosophical problems of creating and using artificial intelligence

Kolesnikov Sergey Viktorovich
Sholom Aleichem Priamursky State University
Master's student

Abstract
The purpose of this article is to consider the philosophical problems of the creation and application of artificial intelligence. As a result of the research, a number of philosophical problems have been identified that hinder the creation of artificial intelligence at the present stage.
Keywords: artificial intelligence, cybernetics, programming language, consciousness, neural network, robotics, cyberepoha.

Введение

Концепция Искусственного интеллекта отводит нас к более ранним временам, чем время появления первых компьютеров – еще к мифам Древней Греции. Гефест, греческий бог ремесленников и кузнецов, создал автоматы, работающие на него. Пигмалион вырезал из слоновой кости статую женщины и влюбился в нее. Афродита вдохнула в статую жизнь в качестве дара Пигмалиону. Женился Пигмалион уже на живой женщине. Эволюция человеческой цивилизации все же, развивается по пути замены естественного - искусственным.

Проблема «искусственного интеллекта», тесно связана с философией и в 21 веке, стала по-настоящему, острой и актуальной. Повсеместное внедрение искусственного интеллекта в жизнь, экономику, бизнес, разработка компьютерной виртуальной реальности в значительной степени

определяют развитие человеческой цивилизации и заставляют задуматься о понятии homosapiens.

Не угрожает ли развитие искусственного интеллекта самому существованию человека как вида. Некоторые исследователи предполагают наступление постантропологической или постчеловеческой эпохи [1].

Может ли машина иметь разум, психические состояния, сознание в том же смысле, что человек? Человеческий интеллект и машинный интеллект – одно и тоже?

Эти и другие вопросы будоражат умы и интересы исследователей, философов и ученых. Ответы на подобные вопросы зависят от понятия определений «интеллект» и «сознание».

Обзор исследований

На сегодняшний день было сделано много исследований и публикаций, рассматривающих философские проблемы связанных с искусственным интеллектом:

Д.Г. Доброродний в своей статье «Философские проблемы разработки Искусственного интеллекта», указывал на то что, наибольшее внимание направлено на достижение технологической сингулярности, создание сильного Искусственного интеллекта и сверхразума, а также на последствия этих достижений [2].

В своей работе «На пути к сильному искусственному интеллекту: социально – философские проблемы» В.Э. Вайцехович, И.Н. Вольнов и Г.Г. Малинецкий рассматривают возможные пути превращения современного слабого искусственного интеллекта в сильный, сравнимый с человеком по творческим способностям [3].

А.Г. Горбачева, Л.В. Мельчукова, Е.С. Евстратенко рассматривали мысленные эксперименты Тьюринга и Серла, с учетом стремительного развития информационных технологий, введены такие понятия как «контекстный ИИ, «искусственной интеллект определенного уровня» и 2 абсолютный ИИ». Рассмотрены примеры, что машина сможет пройти тесты Тьринга [4].

В статье Д.А. Вихарева, Е.А. Кныш «Проблема решения искусственным интеллектом философских вопросов», авторы рассматривают проблемы научно – технического прогресса в контексте революции искусственного интеллекта и киборгизации человека [5].

Целью исследования является изучение философских понятий «интеллект» и «сознание» и выявление причин и проблем, связанных с созданием и использованием искусственного интеллекта в современной действительности.

Интеллект, писал Жан Пиаже, – это высшая форма духовного приспособления к среде, путем организации стабильных, пространственно-временных логических структур [1].

Машинных или искусственный интеллект это, прежде всего система имитирующая решение человеком тех или иных задач в процессе жизнедеятельности. Но мышление, разум, интеллект, творчество, рефлексы и

другие виды психической активности — это продукты человеческой деятельности. Специалисты в области кибернетики, однако, находят возможным моделирование сознания, как индивидуального, так и общественного, при помощи компьютерных технологий. Создание систем способных к самообучению, образованию условных рефлексов и «умозаключениям», основанным на выявлении аналогий.

Проблема особенно проявляется в связи с тем, что нет критериев понимания, оценки того что мы получаем в качестве результата в области разработки искусственного интеллекта – механизм с набором алгоритмов, без понимания процесса, или же машину с возможностью возникновения протопсихических качеств, с задатками психики и интеллекта. Невзирая на термин «искусственный интеллект» в научном сообществе принято считать, что именно наличие сознания, а не интеллекта будет достаточным основанием признания машины – разумной [1].

Философские проблемы развития искусственного интеллекта

Философия искусственного интеллекта затрагивает огромное количество фундаментальных проблем, связанных с его созданием. В чем сущность разума? Принципы его работы? Да и вообще, есть ли возможность создания искусственного интеллекта? Конечно, есть большие успехи в создании алгоритмов и программного обеспечения, способных решать множество интеллектуальных задач эффективнее человека. Но найти ответы на эти и другие вопросы до сих пор не удалось никому.

Вопрос возможности существования или создания искусственного интеллекта определялся мировоззрением: дуалистическая традиция — это невыразимость мышления через телесное, а материалистическая традиция считала мышление производным от телесного. Декарт, исходя из дуалистической позиции, считал мышление атрибутом только человека (даже животных он описывал как «автоматы»), тогда как материалисты теоретически оставляли возможность мышления не только у людей.

Так что же считать критерием наличия разума? Простое сознание, как правило, воспринимает в качестве критерия разумности - поведение. Мы считаем кого - то или что-то разумным или неразумным, оценивая его поведение. Но есть ли связь между сущностью разума и его проявлением? Разумное поведение? Как определить разумно ли оно? И как определить по поведению разумно ли это существо?

Еще в 18 веке Дени Дидров своих «Философских размышления» заявлял о том, что если он найдет попугая, который ответит на любой вопрос, то его без тени сомнения надо считать разумным.

В 1936 году Алфред Айер рассматривал вопрос касательно других разумов: можно ли узнать, что другие люди имеют тот же сознательный опыт? «Единственным основанием, на котором я могу утверждать, что объект, который кажется разумным, на самом деле не разумное существо, а просто машина, является то, что он не может пройти один из эмпирических

тестов, согласно которым определяется наличие или отсутствие сознания» [1].

В 1950 году Алан Тьюринг задавался вопросом – «могут ли машины мыслить?» Он подчеркивал, что традиционный подход к этому вопросу состоит в том, чтобы сначала определить понятия «машина» и «интеллект». Но он выбрал другой путь, заменив исходный вопрос другим, «который тесно связан с исходным и формулируется относительно недвусмысленно».

По существу, он предлагал заменить вопрос - «думают ли машины?», вопросом «Могут ли машины делать то, что можем делать мы, как мыслящие создания?». «Компьютер можно считать разумным, если он способен заставить нас поверить, что мы имеем дело не с машиной, а с человеком?» Преимущество данного вопроса, по утверждению Тьюринга, является то что он проводит «четкую границу между физическими и интеллектуальными возможностями человека», для чего он предлагал эмпирический тест. Тест заключался в следующем: в разных комнатах находятся судья, машина и человек. Судья ведет переписку с машиной и человеком, не зная, кто из них собеседник. Время ответа на вопросы фиксированное. Если судья не сможет определить, кто есть кто, то машина смогла пройти тест Тьюринга и может считаться мыслящей. Причем, машина не просто будет подобием разума человека – она будет именно разумной, так как у нас не будет никакой возможности отличить ее поведение от поведения человека. Такая трактовка искусственного интеллекта как полноправного эквивалента естественного, получила название «сильного искусственного интеллекта»

Но тест Тьюринга совсем не подразумевает, что машина должна «понимать» суть слов и выражений, которыми она оперирует. Она должна лишь «правильно» имитировать «осмысленные» ответы.

В 1980 году Джорж Серл предложил мысленный эксперимент, критикующий тест Тьюринга, да и само представление о возможности существования разума без понимания. Д.Серл пытается имитировать знание китайского языка, которого не понимает. «…Предположим, что меня поместили в комнату, в которой расставлены корзинки, полные китайских иероглифов. Предположим также, что мне дали учебник на английском языке, в котором приводятся правила сочетания символов китайского языка, причем правила эти можно применять, зная лишь форму символов, понимать их значение необязательно. Например, правила могут гласить: «возьмите такой-то иероглиф из корзинки номер один и поместите его рядом с таким – то иероглифом из корзинки номер два». Представим себе, что находящиеся за дверью комнаты люди, понимающие китайский язык, передают в комнату наборы символов, и что в ответ я манипулирую символами согласно правилам и передаю обратно другие наборы символов» [6].

Так Д.Серл проходит подобие теста Тьюринга на знание китайского языка, которого не знает. В данном тесте Д.Серл выполняет исключительно механическую работу и может быть заменен машиной. Своим тестом Д.Серл показывает, что тест Тьюринга не является критерием наличия сознания, а лишь критерием способности манипулировать символами.

Позиция Серла к искусственному интеллекту сводится к следующему: разум оперирует смысловым содержанием (семантикой), тогда как компьютерная программа определяется синтаксической структурой. Следовательно, программы не являются сущностью разума и их наличие недостаточно для наличия разума. Разум не может сводиться лишь к выполнению компьютерной программы. То, что порождает разум, должно обладать, по крайней мере, причинно-следственными свойствами, эквивалентными соответствующим свойствам мозга. Серл перечеркивает проведенный Тьюрингом прямой путь к искусственному интеллекту.

Но в эксперименте Серла есть и свои минусы;

- для прохождения теста в книге должны быть ответы на все существующие вопросы,

- тест проходит система из самого Серла, книги правил и людей, которые эту книгу создали. А эти люди должны знать китайский язык. Хотя другие части системы языка не знают.

Весомым аргументом против теста Тьюринга, как критерия наличия разума, является то, что сам тест является тестом на «человекоподобие» а не на разумность. При прохождении теста машина должна вести себя как человек. Но не все же поведение человека разумно. Многие интеллектуальные задачи машина может решать эффективнее человека. Хотя есть и серьезные пробелы.

Возьмем для примера машинный перевод. Еще недавно возможности компьютера по переводу текста считались во много превосходящим возможности человека. Но на практике было совсем не то, несмотря на огромные объемы информации в памяти машины, получалось «коряво». Нужны знания во всех областях и понимание тем, о которых может идти речь в тексте.

Поэтому поиск возможности создания искусственного интеллекта упирается в первоначальный вопрос - а что такое разум?

В 1963 году Саймон и Ньюэлл на основании анализа языка высказали гипотезу, что сущность разума заключается в способности оперировать символами. Это дало толчок к созданию программ для решения интеллектуальных задач.

Программа Ньюэлла, Шоу и Саймона была сформулирована гораздо четче, чем теория искусственного интеллекта Тьюринга. Исследователи искусственного интеллекта хотели писать программы, которые вели бы себя как люди, но при этом не обязательно мысли как люди. Например, они написали программу для игры в шахматы, которая использовала силу суперкомпьютеров, чтобы оценить тысячу ходов до того, как выбрать один из них, но не пыталась подражать гроссмейстеру, человеку, который оценивает гораздо меньше альтернатив, но делает это гораздо умнее. Они сделали шаг от искусственного интеллекта к компьютерной имитации, заявив, что их программы не только решают проблемы и задачи, но и делают это именно как люди. В компьютерной имитации игры программист должен был бы попытаться написать программу, совершающие те или иные

вычисления и шаги что и гроссмейстер. Различие между искусственным интеллектом и компьютерной имитацией очень важно, поскольку чистый искусственный интеллект не относится к сфере психологии. Усилия в области разработки искусственного интеллекта делают предположения о формах когнитивных ресурсов, которыми люди должны обладать для того чтобы иметь интеллект, но уточнение того, каким образом люди действительно ведут себя разумно, требует полной имитации человеческого мышления, а не только поведения.

Является ли мышление разновидностью вычисления?

Вычислительная теория разума утверждает, что отношения между разумом и мозгом аналогичны отношениям между запущенной программой и компьютером. Если человеческий мозг является разновидностью компьютера, то и компьютеры могут быть как разумными, так и сознательными.

Томас Гоббс писал «рассуждение – это не что иное, как расчет». Т.е. наш интеллект проистекает из формы аналогичной арифметике. Это гипотеза системы символов, о которой говорилось выше, и она подразумевает, что искусственной интеллект возможен.

Может ли машина вызывать эмоции?

Если эмоции определенны только сточки зрения влияния на поведение, тогда эмоции могут рассматриваться как инструмент, который используют для максимизации полезности своих действий.

Ханс Моравец, учитывая такое определение, считает, что «роботы в целом будут очень эмоциональны, если будут хорошими людьми». Он говорит, что «роботы будут пытаться доставить вам удовольствие очевидным самоотверженным образом, потому что они будут испытывать острые ощущения от этого положительного подкрепления. Это можно интерпретировать как любовь».

Может ли машина быть самосознательной?

Самосознание, иногда использовалось писателями-фантастами как название основного человеческого свойства, которое делает индивидуума – человеком.

Тьюринг отбросил все другие свойства человека и свел вопрос к следующему: «может ли машина быть предметом собственной мысли? может ли она думать о себе?» Теоретически можно написать программу, которая будет сообщать о своих внутренних состояниях, но и самосознание предполагает немного большие возможности. Машина должна, не только, каким-то образом приписывать значение своему собственному состоянию, но и выдвигать вопросы без ответов: какова природа своего существования в данный момент; как сравнить ее с прошлым состоянием или будущим, пределы и ценность своего продукта деятельности, как воспринимать свою работу, чтобы ее оценили или сравнили с другими?

Человек может одновременно пребывать в нескольких сенсорных и ценностно-смысловых контекстах, но он не закреплен «намертво» ни с одним контекстом. Контекстуальная природа человеческого существования ни

исключает способности личности осознавать глобальное бытие в мире и свою свободу от каких бы то ни было отдельных обстоятельств [7].

Может ли машина быть доброжелательной или враждебной?

Данный вопрос представим в двух вариациях – «Враждебность» может быть определена в терминах «функции» или «поведении», в этом случае становится синонимом «опасная». Но и можно посмотреть с точки зрения намерений: может ли машина «сознательно» причинить вред? Так может ли машина иметь сознательные состояния?

Злополучный вопрос могут ли высокоинтеллектуальные и полностью автономные машины, быть опасны для человека подробно исследовался футуристами. Рассматривалось множество различных возможных сценариев, в которых машины могут представлять угрозу человечеству.

Одна из проблем в том, что, машины могут очень быстро приобрести автономность и возможно интеллект, необходимые для того чтобы быть опасными.

Вернор Винж предположил, что за несколько лет компьютеры станут в тысячи или миллионы раз умнее человека. Он назвал это «Сингулярностью». Он предполагал, что это может быть в некоторой степени, или, возможно, очень опасно для человека [8].

В 2009 году, на конференции по общему искусственному интеллекту, организованной AGI (обществом искусственного общего интеллекта) ученые и технические эксперты обсуждали потенциальное влияние роботов и компьютеров. Гипотетической возможности того, что они могут стать самодостаточными и будут принимать собственные решения. Отмечалось, что некоторые машины уже приобрели некоторые формы полу автономности, в том числе возможность находить самостоятельно источники энергии, или возможность самостоятельно выбирать цели для атаки с помощью оружия. Обсуждались вопросы о том, что некоторые компьютерные вирусы могут уклоняться от уничтожения и достигли «тараканского интеллекта». Но они отметили, что самосознание, отраженное в научной фантастике маловероятно, хотя есть потенциальные опасности [10].

Некоторые ученые и эксперты ставят под сомнение использование роботов в боевых действиях, особенно когда им дается некоторая степень автономности. Следует уделять больше внимания последствиям их способности принимать автономные решения, по мере того как военные роботы становятся все более сложными.

Дальнейшие шаги привели к развитию разработки искусственного интеллекта по аналогии с человеческим мозгом и его нервной системой – принцип нейросети. Нейросети не программируются, а обучаются. Это главное преимущество нейросетей перед традиционными алгоритмами и программами. В процессе обучения нейросеть должна быть способна распознавать сложные зависимости между входными и выходными данными, выполнять обобщение и анализ.

Специалисты по глубинному обучению нейросетей говорят: «Самое интересное из передовых достижений в том, что мы начинаем видеть, как эти системы начинают пытаться рассуждать» Современные нейрокомпьютеры способны решать многие задачи, но до человеческого разума им еще очень далеко.

Однако Томас Кун отметил, что всякая научная революция предвещается периодом хаоса, когда нормальная практика науки постепенно переходит в то, что называется «экстраординарной наукой». Рано или поздно повседневная практика нормальной науки обязательно приведет к открытию аномалий. Во многих случаях, что то, перестанет работать так, как предсказывала парадигма, в ряде наблюдений обнаружится то, что не вместить в существующую систему убеждений [9].

Заключение

Игнорирование философии искусственного интеллекта человечеством - неправильно, и ее роль в обществе возможно недооценена. Без понимания философии или концепции развития невозможно дальнейшее развитие прогресса. Человечество на современном этапе развития ищет пути развития, в том числе в сфере искусственного интеллекта, цифрового мира, киберэпохи. За последние десятилетия достигнуты достаточно большие успехи в имитации, моделировании человеческого мышления, применение его функций в умных компьютерах. Машины в гигантских размерах превосходят человека в хранении и обработке данных, решают различные задачи и выполняют множество функций. Но пока машины только помощники человека, помогая расширить его возможности. В это пока видится вектор направления развития системы человек плюс машина.

Скорее всего, сильный искусственный интеллект создан не будет, по крайней мере, не в ближайшее время, хотя и создано уже много технологий искусственного интеллекта, которые достаточно сильно влияют на человеческое общество и его развитие.

Библиографический список

1. Айер А. Язык, истина и логика. М.: Канон +; РООИ «Реабилитация», 2010. 240 с. URL: https://a.eruditor.one/file/2187911/
2. Доброродний Д.Г. Философские аспекты проблемы разработки искусственного интеллекта//Философия и социальные науки в современном мире. Материалы международной научной конференции к 30-летию факультета философии и социальных наук Белорусского государственного университета. 2019. С. 189-192. URL: https://www.elibrary.ru/item.asp?id=41107745
3. Войцехович В.Э., Вольнов И.Н., Малинецкий Г.Г. Ожидаемая эволюция ИИ: от слабого к сильному ИИ (философско-антропологические вопросы) // Проблемы онто-гносеологического обоснования математических и естественных наук. 2021. № 12. С. 6-10. URL: https://www.elibrary.ru/item.asp?id=48017658

4. Горбачева А.Г., Мельчукова Л.В., Евстратенко Е.С. Философский анализ проблемы создания искусственного интеллекта // Актуальные направления научных исследований: от теории к практике. 2015. № 4 (6). С. 230-233. URL: https://www.elibrary.ru/item.asp?id=24212739
5. Вихарев Д.А., Кныш Е.В. Проблема решения искусственным интеллектом философских вопросов // Modern Science. 2020. № 7-2. С. 242-246. URL: https://www.elibrary.ru/item.asp?id=43361253
6. Бердяев Н. А. Человек и машина (проблема социологии и метафизики техники) (извлечение) // Вестник Университета имени О. Е. Кутафина. 2022. №4 (92). URL: https://cyberleninka.ru/article/n/chelovek-i-mashina-problema-sotsiologii-i-metafiziki-tehniki-izvlechenie
7. Полуянов В.П. Человек как особый социокультурный феномен и методология интервального подхода // Вестник БГТУ имени В. Г. Шухова. 2012. №3. URL: https://cyberleninka.ru/article/n/chelovek-kak-osobyy-sotsiokulturnyy-fenomen-i-metodologiya-intervalnogo-podhoda
8. Kurzweil R. The singularity is near: When humans transcend biology. Penguin, 2005. URL: https://books.google.ru/books?hl=ru&lr=&id=9FtnppNpsT4C&oi=fnd&pg=PT22&dq=+Ray+Kurzweil+The+Singularity+Is+Near&ots=K4aiWF41zF&sig=0-GAp_W7D1QMhT6CW8P0P5fr_F8&redir_esc=y#v=onepage&q=Ray%20Kurzweil%20The%20Singularity%20Is%20Near&f=false
9. Гроф С. За пределами мозга. Рождение, смерть и трансценденция в психотерапии. – Litres, 2022. URL: https://books.google.ru/books?hl=ru&lr=&id=eIJ8CAAAQBAJ&oi=fnd&pg=PT2&dq=Гроф+С.+За+пределами+мозга+&ots=orq7o4NOwu&sig=GDc6C3pMPDz4MZFy_VqvpIOn_m4&redir_esc=y#v=onepage&q=Гроф%20С.%20За%20пределами%20мозга&f=false
10. The Second Conference on Artificial General Intelligence. URL: http://agi-conf.org/2009/workshop.php (дата обращения 04.12.2022)