

Феномен искусственного интеллекта в философской традиции

Викторов Антон Владимирович,
аспирант, РАНХиГС
E-mail: aviktorov@me.com

Философская проблема искусственного интеллекта имеет богатую историю. Начиная с модели Тьюринга, где искусственный интеллект только начинал разрабатываться до современных философских моделей Бострома, Чалмерза и прочих мыслителей, где искусственный интеллект уже не видится, как простая вычислительная машина. Подход к искусственному интеллекту сильно изменился с начала исследований в этой области. Появляются новые трактовки самого сознания, из-за чего появляются новые версии искусственного интеллекта. Многие философские аргументы нацелены на доказательство невозможности создания искусственного интеллекта, ярким примером тому являются аргументы Деннета об интенциональности и Серла об особой форме понимания, свойственной только человеческому сознанию. При этом по-прежнему ключевым остается философский вопрос – чем является этот искусственный интеллект, каковы его сущности, черты и в чем собственно разница с человеческим сознанием. Возможно ли создать такой искусственный интеллект, который заменил бы человека.

С каждым годом становится все важнее, в условиях взрывного развития искусственного интеллекта, развивать и философские традиции, и дискурс вокруг этого феномена.

Ключевые слова: Искусственный интеллект, Этика искусственного интеллекта, Суперинтеллект, Мировая безопасность, Феномен искусственного интеллекта, философский дискурс.

Появление искусственного интеллекта в философском дискурсе известно еще из истории Античной философии. Аристотель и Пифагор стремились объяснить то, что выходит за рамки классического понимания в чувственном опыте, создавая для этого такие особые системы рационального объяснения, как математические формы, облекающие материю, взять к примеру пифагорейское числовое описание мироздания, как предок современных цифровых систем. В дальнейшем развитии философской традиции это стремление к описанию неизвестных явлений продолжилось, например, в философии Лейбница, который утверждал мировую гармонию, Булевой алгебре, кодирующей любые высказывания на ложные «0» и истинные «1», разностной машине Ч. Бэббиджа и некоторых других. И уже в XX веке появились первые полноценные описания того, что мы сейчас называем искусственным интеллектом.

Расцвет дискуссий вокруг дискурса об искусственном интеллекте пришелся именно на XX век, связано это было, предположительно, с развитием научного знания и появления множества концепций философии науки. Позитивизм, как одно из направлений, появляется именно в XX веке, и оно говорит, что философия нужна для того, чтобы устранить выдуманные, ненаучные проблемы, которые не относятся к чувственной реальности, а также философия должна заниматься анализом языка и выявлением в нем нелогичных конструкций, которые не дают отсылок к реальности. Сущность позитивизма состояла в том, что он отвергал все умозрительные размышления философии, а на первый план выдвигал эмпирический опыт. Философия должна обобщать опыт наук, а не заниматься абстрактными размышлениями. Первый позитивизм ставил перед собой задачу обобщения опыта, а роль философии снижалась. Философ и основоположник позитивизма Огюст Конт считал, что «существует несколько стадий развития человечества, каждой из которых было свойственно собственное понимание и объяснение мира» [2]. Первая стадия была теологической, где объяснением было религиозное видение, вторая стадия была метафизическая, где объяснение природы происходило с помощью категории причинности, а третья стадия – позитивная, научная, именно здесь появляется рациональное осмысление природы, отсюда и берет свое название позитивизм. В дальнейшем оно переросло в направление логического позитивизма, где уделялось большое внимание философии языка и постпозитивизм, где анализировались проблемы роста научного знания и его сущности.

Все это сместило акценты и создало плодородную почву для зарождения традиций и развития философского дискурса в области новых типов знаний, в том числе таких как искусственный интеллект, это дало надежды на то, что философия более не будет связана с метафизическими построениями, мало относящимися к реальности, а напротив, будет служить обществу, оказывая пользу для всего человечества в лице развития новых цифровых технологий. Развитие искусственного интеллекта привело к появлению первых цифровых машин и сильно изменило представления людей о цифровизации. Появились надежды, что в скором будущем машинный интеллект станет наравне с человеческим или даже его превзойдет. Но, как мы видим, эти ожидания

не сбылись и предполагаемого прорыва на сегодняшний день не произошло.

Безусловно, искусственный интеллект с момента его разработки сильно изменился, появились колоссальные изменения, однако говорить о том, что он стоит наравне с человеческим весьма преждевременно. Совершенствуются алгоритмы поиска, машины могут решать более сложные задачи, чем те, которые стояли перед ними изначально. Но вся проблема состоит не в том, что искусственному интеллекту не хватает мощности, чтобы превзойти человека. Как раз в мощности искусственный интеллект очень даже превосходит человека. Он способен обрабатывать такие массивы данных, которые бы не были доступны человеку ни на каком этапе развития. Все дело в том, что до сих пор в философии искусственного интеллекта нет четкого определения, чем является этот искусственный интеллект, каковы его сущности, черты и в чем собственно ключевая разница с человеческим сознанием. Эта проблема намного сложнее, чем кажется на первый взгляд, поскольку именно определение искусственного интеллекта задает его рамки (включая рамки философского дискурса) и дает понять в сравнении чем является сознание. Если бы мы получили непротиворечивое и устраивающее всех определение искусственного интеллекта, то дискуссии вокруг него бы прекратились в тот же момент.

Традиционные вопросы, которые задаются при попытке определения сущности искусственного интеллекта сводятся к тому, чтобы дать отличия искусственного интеллекта от человеческого сознания, и звучат они следующим образом. Возможно ли создать такой искусственный интеллект, который будет похож на человеческое сознание? В какой мере человеческое сознание воспроизводимо? Где находятся границы интеллекта и сознания? Эти и многие другие вопросы задаются в связи с попыткой разработки непротиворечивой теории искусственного интеллекта. Однако и по сей день мы не имеем четкого ответа ни на один из представленных вопросов.

Те определения искусственного интеллекта, которые мы имеем на данный момент, страдают одним большим недостатком – они очень общие и не конкретные. Швейцарский психолог Жан Пиаже считает, что «интеллект есть то, что обладает интегративным функционалом всех процессов психики, что в свою очередь и обеспечивает нахождение и адаптацию человека в мире». Есть множество других определений: в итоге, как замечает В. Ладов, «человеку никак не удается найти уникальное место, он оказывается подобным либо системе искусственного интеллекта (при выборе парадигмы «интеллект как исчисление понятий»), либо обычному биологическому организму (при выборе парадигмы «интеллект как восприятие»)» [3].

Важным поворотом в сфере изучения искусственного интеллекта считается событие, которое произошло летом 1956 года в университете города Ганновер, в колледже Дартмут [4]. Это событие – семинар, который был посвящен, в том числе, проблемам искусственного интеллекта и нейронным сетям, где присутствовали философы сознания. Сами результаты семинара не дали особых откровений и революционных открытий, однако семинар позволил зародиться самому философскому дискурсу об искусственном интеллекте. Именно там появилось название для еще только зарождающейся науки в области сознания – искусственному интеллекту, которое было предложено американским информатиком и разработчиком основ функционального программирования Джоном Маккарти. В дальнейшем он будет удостоен премии А. Тьюрин-

га за огромный вклад в области исследований искусственного интеллекта.

В вопросе об определении того, что является искусственным интеллектом нет единого мнения, однако имеется три доминирующих трактовки. Первая из них употребляет этот термин в значении научной сферы, целью которой является воссоздание процессов познания и человеческого мышления для повышения эффективности обработки данных техникой. Вторая трактовка определяет искусственный интеллект, как набор различных механизмов, устройств, которые являются интеллектуальными. И третья трактовка определяет искусственный интеллект, как набор представлений о том, что такое разум, что есть познание применительно к человеческому сознанию, что дает возможность постановки проблемы воссоздания интеллекта.

В современном мировом сообществе кругозор использования искусственного интеллекта охватывает множество областей, к которым относится и распознавание речи, и теория математических доказательств, и возможность диагностики болезней, а также более обширные области, вроде теории восприятия. Искусственный интеллект также применяется для реализации проектов любого рода, сложности и направления. Актуальность данной области знаний таким образом не вызывает никаких сомнений.

Философ и кибернетик Н. Нильсон давал собственное определение области искусственного интеллекта: «в самом общем смысле – это решение «интеллектуальных» задач с помощью автоматических методов, в первую очередь, с помощью вычислительных машин. Но какую деятельность следует считать интеллектуальной, а какую нет? Это не вполне ясно» [5].

Для того, чтобы дать определение понятию искусственный интеллект предлагается использовать системный подход, который рассматривает любой феномен, как систему, состоящую из множества элементов, каждый из которых играет важную роль. Интеллект, в общем виде, есть то, что объединяется в систему функций познания субъекта. При всем при этом высокий порог интеллекта считается тот, что образует единство из таких функций, как память, чувствительные способности и другие. В этом смысле, интеллект позволяет решить не только задачи, которые имеют заданный алгоритм, но и такие, которые не имеют такого алгоритма.

Самое важное уточнение относительно понятия искусственного интеллекта заключается в том, что сам искусственный интеллект, а точнее попытки по его созданию сводятся к задаче имитации человеческой способности к мышлению.

Так, первая полноценная попытка создания искусственного интеллекта (имитации человеческого сознания) была произведена в середине XX века великим математиком Аланом Тьюрингом в статье «Вычислительные машины и разум», изданной журналом «Mind» более 70 лет назад. Самый первый вопрос, которым задается Тьюринг является вопрос о возможности мыслить для искусственной машины. Обычный ответ на этот вопрос начинался с определения понятий машины и интеллекта. Однако Тьюринг решает поступить иначе. Вместо того, чтобы спрашивать возможна ли мысль у машины, он спрашивает о том, может ли машина поступать так, как поступаем мы. По мысли Тьюринга, он сохраняет в своей формулировке строгую границу между физическими и интеллектуальными способностями индивида.

Для демонстрации своих подходов А. Тьюринг предложил тест. Классическая интерпретация теста заключалась в том, что человек общается с компьютером и с человеком. На основании полученных ответов на заданные

вопросы, человек должен в итоге определить с кем он все-таки разговаривает – компьютером или человеком, при этом не видя самого человека, и не зная находится ли он на том конце. Программа должна создаваться с таким уклоном, чтобы человек не смог понять с кем он разговаривает. Таким образом, предполагалось доказать возможность создания умной машины – искусственного интеллекта, который способен воссоздать человеческое мышление. Тест Тьюринга был раскритикован учеными и философами, но несомненно особенным образом повлиял на философский дискурс в области искусственного интеллекта. Отсюда также следует актуальность данной проблемы, поскольку она до сих пор не является решенной. Тест Тьюринга, во-первых, затронул проблему разграничения сильного и слабого искусственного интеллекта. Сильный интеллект – это и есть человеческий разум, сознание, мышление, а слабый – лишь имитация каких-либо простых действий, например, действий в соответствии с четко заложенным алгоритмом. Данное разграничение будет подробнее рассмотрено далее.

Чуть ранее теста Тьюринга, в 1936 году, философ Альфред Айер обратился к проблеме существования других сознаний, сознаний людей, и задался вопросом: как мы узнаем, что другие люди имеют свой собственный опыт? И только один ответ, по мысли Айера, был истинным – это то, что «машина не может пройти ни один эмпирический тест, который выявляет существование ли сознание у его испытуемого или нет» [1]. Эта попытка описания очень близка тому, что высказывал Тьюринг, но тем не менее достоверно неизвестно, знал ли Тьюринг об этой мысли Айера. Но не будем обвинять никого из философов в заимствовании идей, поскольку есть одно весьма распространенное мнение, которое, как нам кажется, отражает всю ситуацию. Это мнение гласит, что «вся философия – это лишь заметки на полях Платона». Отчасти это мнение истинно, поскольку так или иначе всю новую философскую традицию всегда можно герменевтически вывести из предыдущей, а поэтому не остается смысла в критике Тьюринга по этому аспекту. И сегодня тест Тьюринга до сих пор не потерял своей актуальности, но все же проходит более успешно в связи с возрастающей мощностью вычислительных систем, которые научились быстро обрабатывать информацию.

В настоящее время все еще остается актуальным вопрос об имитации человеческого интеллекта путем искусственного. Еще в конце XX века программа «Deep Blue» одержала победу в шахматном поединке над Г. Каспаровым, что говорит о высокой мощности обработки данных, но опять же не говорит об имитации сознания человека. В настоящее время мы видим, как искусственный интеллект справляется с написанием различных письменных работ, ярким примером недавнего является дипломная работа, написанная искусственным интеллектом ChatGPT.

Наиболее успешные попытки развития теорий искусственного интеллекта принадлежат традиции аналитической философии. Философ Дэвид Чалмерз вводил разделение для искусственного интеллекта, чтобы показать границы его различных версий. Это разграничение говорит о существовании двух видов искусственного интеллекта: сильного и слабого. Слабый искусственный интеллект нацелен на воспроизводство простых функций машины, не посягая на человеческое сознание, таковы, например, роботы-обслуживающий персонал, переводчики, шахматные программы и прочее. Сильная же версия искусственного интеллекта пытается найти возможности воспроизведения человеческого интеллекта, человеческого сознания.

Также этим разграничением пользовался и американский философ Джон Серл, который является создателем знаменитого мысленного эксперимента «Китайской комнаты», в которой он опровергает попытку построения сильного искусственного интеллекта похожего на человеческий. Джон Серл строит свой эксперимент на критике теста Тьюринга. Сущность эксперимента Серла состоит в следующем. Существует человек, который находится в комнате с карточками, на которых изображены иероглифы, при этом он не знает ни одного китайского иероглифа. Но у него есть инструкции по взаимодействию с ними. Эти инструкции подобны инструкциям компьютера. Также существует человек, который находится за пределами комнаты, который знает китайский язык и передает вопросы в комнату, на которые ждет ответа. Получается, будто бы человек, который знает все инструкции, знает язык, поскольку верно отвечает на вопросы, а поэтому понимает его. Так, тест Тьюринга проходит успешно, но понимания не появляется. Следовательно, такая проверка интеллекта, сознания не является легитимной. Для Серла важно было показать, что существует разница семантики и синтаксиса. В этом эксперименте, в комнате, человек обладает лишь синтаксисом, тогда как сознание подразумевает смысл, то есть семантику. Человек воспринимает нечто, выделяет некоторые смыслы, а не выводит какую-то информацию просто из набора правил. Таким образом, мы можем говорить о том, что машины могут выполнять часть операций, на которое способно человеческое сознание, но на мой взгляд, это не даёт нам право говорить о том, что машины действительно могут мыслить.

Аргументы Серла против теста Тьюринга, а также сам тест Тьюринга вызвал многочисленные дискуссии вокруг проблемы возможности создания искусственного интеллекта. Так, развитие этой проблемы имело место в философии Дэвида Чалмерза. Ключевым аргументом Чалмерза состоит в том, что сознание должно обладать способностью субъективного восприятия. Поэтому даже очень мощные интеллектуальные машины не обладают им. То же самое можно сказать и о животных, поскольку хотя они и могут обладать сознанием, но их интеллект явно ниже уровня человеческого интеллекта.

В предположениях Чалмерза также присутствует мысль о том, что без физического тела трудно представить сознание, хотя он и не отрицает, что в будущем появится какая-либо система, которая будет говорить об обратном.

Дискурс об искусственном интеллекте также развивается философом сознания Дэвидом Деннетом. Деннет отрицает позицию квалиа, согласно которой у каждого человека есть особые переживания. Функционализм, как одно из направлений философии сознания, представляет собой такую теорию, в рамках которой «то, что делает нечто ментальным состоянием определенного типа, зависит не от своего внутреннего устройства, а от того, как оно функционирует, какую роль играет в системе, частью которой является» [6]. Если Чалмерза интересовал вопрос что такое сознание, то Деннет интересуется вопросом то как оно работает. Наше сознание работает по аналогии функции машины, с заложенными программами внутри него, по мысли Деннета. При этом философ не отрицает ментальные состояния вообще, но сводит их к паттернам поведения человека. В этой связи он говорит об искусственном интеллекте как о том, что понимает ничуть не хуже человека, их интенциональные направленности равны. Таким образом, мир будет наполнен смыслом как для человека, так и для искусственного интеллекта, причем ни один из них не будет выражать больше или меньше понимания.

Так, все изложенные проблемы могут решаться при важной оговорке – если интеллект человека не является совокупностью биофизической природы. Это значит, что мы отрицаем редукцию человеческого сознания к структурам мозга, к деятельности мозга. Этой редукцией занимаются философы бихевиористы и натуралисты, которые утверждают, что все действия человека – это результат несвободной деятельности нейронной активности в мозге каждого конкретного человека. Однако такая редукция есть ни что иное как попытка исключить свободные действия человека и свести всю его активность, в том числе творческую, познавательную и прочие, к работе нейронных связей, к материальным частям мозга, которые отвечают за активность человека. Но такая редукция похожа на то, как если бы мы сравнили сюжет какой-либо компьютерной игры с работой движка самой игры.

Мы же исходим из того, что человеческое сознание не есть нечто только функциональное, что может быть сведено к отдельным материальным связям. Сознание человека – это сложное явление, которое обладает многими функциями, не сводясь к какой-либо одной, не имея редукции к деятельности в мозге, к материальным связям, а напротив, имеющее особое эмерджентное свойство субъективного восприятия и, кроме того, имеющее в своем распоряжении органы чувств, психику и прочее. Из этого становится очевидным тот факт, что сознание невозможно описать в каких-либо внешних терминах, и как вывод тот уровень искусственного интеллекта, который мы имеем на данный момент, не совпадает с тем уровнем, который имеется в сознании человека. Однако какие результаты нас ждут в дальнейшем, пока доподлинно неизвестно.

По определению, человек – это биосоциальное существо, которое при этом не сводится к своим биологическим и социальным свойствам, а действует на более высоком уровне, на уровне культуры. Человек постоянно воссоздает символическую коммуникацию, которая случается не только между самими индивидами, но и между другими системами, в том числе, искусственным интеллектом. Причем биологическая часть человека иногда не просто не является тем, к чему сводится человек, но и более того, что иногда противоречит биологической природе. Это ярко заметно на примерах морального поведения человека, где биологические факторы были бы предпочтительнее, нежели то, как человек поступает в сфере нравственного выбора. Биология, безусловно, остается основой человеческого тела, физической активности, поскольку она задает базовые процессы в организме, позволяя существовать человеку на физическом уровне, но при этом сейчас она все менее влияет на человека. Именно индетерминированность человека позволяет появиться возможности человека к созданию государств, культур, соблюдению традиций, а также к развитию коммуникации и разветвленной системе компьютерных технологий, которые порождают новые сферы искусственного интеллекта. Вопрос о сознании-тела, философская традиция которого идет еще со времен дуализма Декарта и даже ранее, заставляет постоянно переосмысливать возможность о создании такого интеллекта, который будет максимально похож на человеческий. Поскольку сознание развивается не по предзаданным законам, а человек в целом не сводится к совокупности своих частей, будь то физические или иные, то вопрос о создании искусственного интеллекта остается открытым. Для этой проблемы еще не очерчены все возможные грани и ограничения, но уже сейчас мы видим большие прорывы в этой области. Уже сейчас мы видим

то, насколько сильно слабый искусственный интеллект продвинулся в выполнении повседневных, и не только, задач. Но на сегодняшний день, к сожалению, нет полноценных решений создания сильного искусственного интеллекта, а поэтому нам только предстоит познать эту область.

Литература

1. Айер А. Язык, истина и логика. М: «Канон+» РООИ «Реабилитация», 2010. – 240 с.
2. Гимадиева Л.И. Философия позитивизма // Скиф. 2021. № 12 (64). С. 374. URL: <https://cyberleninka.ru/article/n/filosofiya-pozitivizma>
3. Ладов В.А. Философские проблемы искусственного интеллекта / Томск: Томск. гос. ун-т., 2005. URL: http://fsf.tsu.ru/wp_test/wp-content/uploads/tutorials/AI211.pdf
4. Люггер Д. Искусственный интеллект. Стратегии и методы решения сложных проблем / Д. Люггер. – 4-е изд. – М.: Вильямс, 2003. – 864 с.
5. Нильсон Н. Искусственный интеллект. Методы поиска решений / Н. Нильсон. – М.: Мир, 1973. – 273 с.
6. Сирл Дж. Разум, мозг и программы / Дж. Сирл // Глаз разума / Д. Хофштадтер, Д. Деннетт. – Самара: Бахрах-М, 2003. – С. 314–331.

THE PHENOMENON OF ARTIFICIAL INTELLIGENCE IN THE PHILOSOPHICAL TRADITION

Viktorov A.V.
RANEPA

The philosophical problem of artificial intelligence has a long history. Starting with the Turing model, where artificial intelligence development was just beginning to the modern philosophical models of Bostrom, Chalmers and other philosophers, where artificial intelligence no longer appears to be a simple computing machine. The approach to artificial intelligence has changed dramatically since the beginning of research in this field. There are new interpretations of consciousness itself, and new versions of artificial intelligence are emerging. Many philosophical arguments are aimed at proving the impossibility of artificial intelligence, as exemplified by Dennett's arguments about intensiveness and Searle's special form of understanding, unique to human consciousness. At the same time, the philosophical question remains the key – what is this artificial intelligence, what are its essence, features and what is the difference with the human consciousness. Whether it is possible to create an artificial intelligence that would replace the human. Every year becomes more and more important, in the explosive development of artificial intelligence, to develop both philosophical traditions and discourse around this phenomenon.

Keywords: Artificial Intelligence, Artificial Intelligence Ethics, Super Intelligence, World Security, Artificial Intelligence Phenomenon, Philosophical Discourse.

References

1. Ayer A. Language, truth and logic. M: "Kanon +" ROOI "Rehabilitation", 2010. – 240 p.
2. Gimadieva L.I. Philosophy of positivism // Skif. 2021. No. 12 (64). P. 374. URL: <https://cyberleninka.ru/article/n/filosofiya-pozitivizma>
3. Ladov V.A. Philosophical problems of artificial intelligence / Tomsk: Tomsk. state Univ., 2005. URL: http://fsf.tsu.ru/wp_test/wp-content/uploads/tutorials/AI211.pdf
4. Lugger D. Artificial intelligence. Strategies and methods for solving complex problems / D. Lugger. – 4th ed. – M.: Williams, 2003. – 864 p.
5. Nilson N. Artificial intelligence. Solution search methods / N. Nilson. – M.: Mir, 1973. – 273 p.
6. Searle J. Mind, brain and programs / J. Searle // Mind's eye / D. Hofstadter, D. Dennett. – Samara: Bahrah-M, 2003. – S. 314–331.