

машины оказывается вполне достаточно для того, чтобы вынести решение о том, способна ли она мыслить; для прохождения теста Тьюринга вопрос о наличии приватного опыта несущественен.

В конечном итоге проведенный историко-философский анализ оснований области исследования ИИ и философии сознания позволяет сделать следующий вывод. При принятии теста Тьюринга как достаточного критерия наличия мышления у машины вопрос «что пытаются создать разработчики ИИ?» получает закономерный ответ: эффективно имитирующую человеческое языковое поведение систему, чья внутренняя организация не является существенной. В этом случае у разработчиков ИИ нет потребности в дальнейшем обращении к понятийному аппарату философии сознания для ответа на этот вопрос, так как, определив мышление как эффективную имитацию, они получают исчерпывающее описание машины, которую стремятся создать.

При отказе от теста Тьюринга как от такого достаточного критерия поиск ответа на вопрос «что пытаются создать разработчики ИИ?» требует дополнительного обращения к философии сознания. Целью подобного обращения становится попытка сформулировать дополнительные содержательные дескрипции для определения того, какая машина может считаться мыслящей. Ядром этого набора дескрипций являются такие понятия, как сознание (consciousness), интенциональность, приватный опыт и т. д., которые являются понятийным аппаратом философии сознания. Без конкретной реализации этих характеристик ментального (mind) в машине сторонники подобного подхода не видят возможности создания ИИ.

МЕСТО ТЕОРЕМ ГЕДЕЛЯ О НЕПОЛНОТЕ В ДИСКУССИИ ОБ АЛГОРИТМИЗАЦИИ МЫШЛЕНИЯ

И. И. Красов

*магистрант 2 курса направления «Интеллектуальные
системы в гуманитарной сфере» Департамент философии
Уральского гуманитарного института Уральского
федерального университета имени первого Президента России
Б. Н. Ельцина, г. Екатеринбург*

Введение

Проблема алгоритмизации мышления исторически пересекается с программой создания систем искусственного интеллекта. В данной статье предпринята попытка раскрыть спецификацию проблемы алгоритмизации мышления через разворачивающийся спор об ее возможности, который разделил сообщество на две противоборствующие линии: менталистов и механицистов. В основе статьи лежит анализ идей Ю. Л. Ершова и В. В. Целищева,

представленных ими в работе «Алгоритмы и вычислимость в человеческом познании»⁶¹.

В первой части нашей статьи мы раскроем истоки и проблематику алгоритмизации мышления, и покажем, что ложится в основу при ответе на вопрос: возможна ли алгоритмизация мышления?

Вторая часть статьи посвящена раскрытию фона дискуссии между ментализмом и механицизмом. Мы познакомимся с первыми представителями ментализма: Дж. Лукасом и Р. Пенроузом.

Следующую, третью часть мы посвятим изложению места теорем Геделя о неполноте в рамках вышеуказанного спора о возможности алгоритмизации мышления. Мы выясним как интерпретация этих теорем проявляет себя в выяснении соотношения между мышлением человека и деятельности компьютера.

Далее, в четвертой части мы выясним в чем кроется проблема движения от математических фактов к философским выводам. И как эта проблема, в границах разбираемой нами дискуссии менталистов и механицистов, показывает герменевтический характер дискуссии, и что этот характер значит.

В пятой части мы остановимся на раскрытии позиции Геделя в дискуссии. Мы покажем, как идеи Геделя соотносятся с критикой механизма. Для этого мы обратимся к сформулированной Геделем дилемме и проведем ее анализ.

Выяснив, что Гедель выступает против механицистов, можно задать вопрос: являются ли идеи Геделя менталистскими? Последняя, шестая часть нашей статьи показывает, какие проблемы не дают нам идеи Геделя полностью соотнести с идеями Лукаса и Пенроуза.

1. Программа алгоритмизации мышления: истоки и установка проблематики

Когда мы обращаемся к исследованиям в области алгоритмизации мышления, то мы чаще всего сталкиваемся с вопросом «может ли машина мыслить?». Постановкой такого вопроса мы обязаны А. Тьюрингу.

Ершов и Целищев признают, что идеи выдвинутые Тьюрингом помогли ограничить этот вопрос в психологическом плане. Это ограничение связано с тем, что Тьюринг сформулировал тест, следуя которому можно определить обладает ли машина мышлением. Однако для полного понимания проблематики алгоритмизации мышления авторы «Алгоритмы и вычислимость в человеческом познании» постулируют необходимость

⁶¹ Ершов Ю. Л. Алгоритмы и вычислимость в человеческом познании / отв. ред. К. Ф. Самохвалов, А. С. Морозов. Новосибирск: Изд-во СО РАН, 2012.

найти «более формальный аспект проблемы, который бы имел ярко выраженный философский интерес»⁶².

В дальнейшем к поиску решения вопроса о возможности мышления машины присоединяется ряд когнитивных исследований, в основе которых лежат попытки компьютерного моделирования работы мозга с опорой на результаты достигнутые в физиологии, математике, компьютерных исследованиях и т. д. По сути, перед нами встает программа создания систем искусственного интеллекта. Имеет смысл обратить внимание на работу «Об интеллекте»⁶³, проделанную в этом направлении Дж. Хокинзом вместе С. Блейкли. В этой книге Хокинс формулирует теорию работы мозга на основе модели «память-предсказание», отталкиваясь он тезиса невролога В. Маунткастла, который можно сформулировать так: все части неокортекса, не зависимо от своих специализаций, работают на основе одного алгоритма. По мнению Хокинза, программа создания искусственного интеллекта (в сильном его понимании) не увенчалась успехом не из-за проблем технического характера, а из-за принципиальной проблемы.

Придерживаясь схожего взгляда, Ершов и Целищев видят суть проведенного ими исследования в не практической попытке проведения алгоритмизации мышления, а в установлении принципиальной возможности подобной алгоритмизации. Необходимо обратиться к достаточно значимому результату и при этом заслуживающему тщательного философского осмысления. Данным требованиям соответствует теоремы Геделя о неполноте, «которые, с одной стороны, представляют собой один из самых известных и значительных результатов в математической логике, а с другой – предмет множества самых разнообразных философских спекуляций, которые никак не хотелось бы повторять»⁶⁴.

В конечном итоге, фокус внимания авторов останавливается, как собственно и нашего, на рассмотрении спора двух направлений, дающих на вопрос о возможности алгоритмизации мышления противоположные ответы. Главная проблема этой дискуссии заключается в разных интерпретациях теорем Геделя о неполноте. Дискуссия отличается большим разбросом в методах и стилистике аргументации, которая, в свою очередь, выделяется недостаточной завершенностью.

⁶² Ершов Ю. Л. Алгоритмы и вычислимость в человеческом познании / отв. ред. К. Ф. Самохвалов, А. С. Морозов. Новосибирск: Изд-во СО РАН, 2012. С. 268.

⁶³ Хокинс Дж., Блейкли С. Об интеллекте / пер. с англ. М.: ООО «И. Д. Вильямс», 2007.

⁶⁴ Ершов Ю. Л. Алгоритмы и вычислимость в человеческом познании / отв. ред. К. Ф. Самохвалов, А. С. Морозов. Новосибирск: Изд-во СО РАН, 2012. С. 268.

2. Фон спора «ментализма» и «механизма»

Мы остановимся на двух персоналиях, для того, чтобы очертить фон указанного выше спора. Первым выступает Р. Пенроуз и его две работы «Новый ум короля» и «Тени разума», где им используются теоремы Геделя о неполноте для защиты тезиса о неалгоритмичности мышления человека. Вторым и по совместительству катализатором самого спора стал Дж. Лукас, а именно его полемическая статья «Умы, механизмы и Гедель», постулирующая превосходство человеческого разума над машиной. Вместе Пенроуз и Лукас вошли в малочисленную группу исследователей, которые в споре о возможности алгоритмизации мышления принадлежат к направлению «ментализма». Странники же противоположного направления, которые в первую очередь придерживались редуционного подхода в отношении ментальной деятельности, выступали под флагом «механизма». Этот спор, как указывают авторы «Алгоритмы и вычислимость в человеческом познании», маловероятно можно было бы назвать серьезным, даже несмотря на сложность поставленной проблемы, в виду малочисленности одной из сторон, если бы не одно обстоятельство. Этим обстоятельством, которое внесло существенный вклад в разгар дискуссии, является свидетельство Хао Вана о том, что Гедель являлся противником «механизма». Таков фон проблемы, которая выразилась в споре между менталистами, решительно выступающими против возможности алгоритмизации человеческого мышления, и механицистами, утверждающими, что формализовать человеческое мышление возможно.

3. Место теорем Геделя о неполноте в споре о возможности алгоритмизации мышления

Теоремы Геделя о неполноте имеют такое важное значение для дискуссии между менталистами и механицистами потому, что они представляются как строгий результат относительно фундаментальных ограничений человеческого мышления. Теоремы Геделя, сформулированные для формальных систем математики, могут быть интерпретированы таким образом, что будут представлять ответы о возможностях мышления компьютера: «Поскольку теоремы Геделя формулируются для формальных систем математики, в частности для формальной арифметики, а формальная система эквивалентна идеализированной вычислительной машине, то теоремы Геделя могут быть интерпретированы как говорящие нечто и о возможностях компьютерного «мышления»»⁶⁵. Так сложилось, что интерпретации строящиеся вокруг теорем Геделя приводили к разграничению между математиками и философами. Однако в виду интерпретации, приведенной нами выше, возникает

⁶⁵ Там же . С. 271.

проблематика, которая интересует оба сообщества, математическое и философское, а именно отношение «компьютерного» мышления и мышления человека.

Само сопоставление компьютера и мозга является довольно проблемным. Так, выдвигая, модель, по которой мозг уподобляется компьютеру, мы узнаем, что достоверных и точных знаний о работе компьютера в нашем распоряжении значительно больше, чем о нашем мозге. О подобном пишет и Хокинс, когда рассуждает о провалах попыток строительства систем искусственного интеллекта: нам необходимо разобраться в работе нашего мозга, прежде чем проводить работу по созданию ИИ, в связи с этим, Хокинс и выстраивает свою теорию работы мозга, а в частности неокортекса, которому им отводится ведущая роль.

Не смотря на различные проблемы, указанные выше, проведение параллелей между мозгом и компьютером допустимо, поэтому наши знания о компьютере в некоторой степени могут быть перенесены на наши представления о деятельности мозга, и более того, на устройство человеческого мышления в целом. Остается вопрос только в том, какие из полученных знаний о компьютере нам стоит приложить? В связи с этим и становятся интересными теоремы Геделя о неполноте, в границах которых выражается фундаментальное ограничение о возможности полностью формализовать интуитивное знание. И тогда нам следует задаться вопросом о том, можем ли мы результаты математической логики перенести на человеческое мышление? Как показала история, можно привести сложное хитросплетение аргументов, как в сторону положительного ответа на поставленный вопрос, так и в пользу отрицательного, что и вылилось в спор между менталистами и механицистами.

Относительно расплывчатую тематику о сопоставление компьютера и когнитивных возможностей человека можно сузить до более точных положений, уточняющих проблематику:

Во-первых, наша проблематика ограничивается теоремами Геделя о неполноте, а, в частности, соответствующими их интерпретациями, связанными с сопоставлением человеческого мышления и мышления компьютера.

Во-вторых, под мышлением подразумевается ограниченность «математическим мышлением», т. е. способностью рассуждать с «математической определенностью»⁶⁶.

В-третьих, ведущим вопросом признается вопрос о возможности алгоритмизации мышления, с учетом всех возможных отношений между человеком и машиной: постулирования превосходства человека, и алгоритмической природой его мышления, или признание

⁶⁶ Там же. С. 273.

эквивалентности машины человеку, или же принятия превосходства машины над человеком.

4. Герменевтический характер дискуссии

Проблема соотношения между машинным интеллектом и человеческим мышлением пролегал в русле того, что мы движемся от математических результатов к философским заключениям. Для нас принципиально важно не встать на путь, на котором мы будем оправдывать сомнительные философские выводы точно определенными математическими фактами. Историческим примером такого толка, как указывают авторы «Алгоритмы и вычислимость в человеческом познании», являются интерпретации теорем Геделя о неполноте, поэтому мы вслед за авторами должны сделать отступление, которое поможет нам понять спор между менталистами и механицистами более точно.

В чем же проблема, когда мы пытаемся проводить философские интерпретации математических фактов? Проблема возникает тогда, когда мы постулируем некоторые философские следствия из математических фактов. А допустимо ли постулировать такие следствия? В подобных случаях структура проводимой аргументации, может быть довольно сложной и запутанной, и, что примечательно, такая ситуация происходит из-за того, что в основе этой структуры лежат неявные философские посылки. Во время строгого анализа основной задачей исследователей и должна быть экспликация таких неявных посылок, что на практике оказывается очень трудным.

Однако, если просто избегать философских посылок, в том числе и сомнительных, то всякая аргументация от математики к философии прекращается. Если же использовать крайне сомнительные посылки, то интерес к аргументации практически утрачивается. Довольно часто нам приходится пользоваться философскими допущениями, о сути которых нам неизвестно, пока мы не заходим в тупик, придя к неправдоподобным выводам, или впад в противоречие. В таком случае мы должны отказаться от наших посылок. Такие решения Бенацераф называет «герменевтическими». Таким образом, наши посылки не могут быть даны нам в результате априорного выбора, а наша структура аргументации не может быть подвержена строгой дедуктивной трактовке. В случае проводимого исследования спора менталистов и механицистов о возможности алгоритмизации мышления проявляется герменевтическая, или, как предпочитают писать Ершов и Целищев, «диалектическая» суть споров подобного рода, которые отталкиваются от математических фактов, и где нет конечного аргумента «за».

Какое же философское допущение принимается во время интерпретаций теорем Геделя о неполноте? В случае нашего

исследования, философское заключение заключается в убеждении, «что понятие арифметического утверждения, не основанное на формальном доказательстве, слишком бедно для того, чтобы нести с собой полноценную концепцию истины»⁶⁷.

В основу формальных моделей, ложатся наши представления, которые носят интуитивный характер, о значении символов и выражений в той же математике. Такие представления должны были бы отражаться в наших формальных моделях, однако случается так, что мы из формальных моделей извлекаем ограничения на понимание оснований этих моделей, то есть исходных символов. Исходя из вышеуказанного, теоремы Геделя приводят к следующим заключениям некоторых исследователей: из факта о неполноте формальных моделей полагается, что формализованные теории не могут схватить интуитивное содержание математического мышления. На данную трактовку можно вполне оправдано дать возражение, что даже при условии достаточного схватывания наших исходных значений формальной моделью, мы маловероятно можем ограничения формализма переносить на исходную математическую практику.

5. Позиция Геделя в дискуссии. Дилемма Геделя и ее анализ

Какую же позицию принимал сам создатель теорем о неполноте, Гедель? Как мы выяснили из свидетельства Хао Вана, Гедель выступал против «механизма». Содержание философский исканий Геделя приоткрылось в 1995 году, когда вышел третий том посмертных публикаций Геделя, среди которых находилась лекция из целой серии таких лекций посвященных Гиббсу. Из этой лекции следует, что Гедель полагает доказанными два результата, говорящих нам о соотношения человеческого ума и машины:

«1) человеческий ум не способен к формулировке (или механизации) всех математических интуиций; если они сформулированы, появляются новые, например, о непротиворечивости;

2) либо человеческий ум превосходит все машины (более точно, может решить больше теоретико-числовых проблем), либо существуют такие теоретико-числовые проблемы, которые неразрешимы для человеческого ума»⁶⁸.

Второе же положение и породило многочисленные споры, так как имеет важные следствия для решения вопроса о соотношения человека и машины. Как мы видим положение построено в виде дилеммы, где принятие одной и отвержение другой из частей дилеммы указывает на философскую позицию, которой придерживается исследователь. Исходя из различных комментариев, в том числе

⁶⁷ Там же. С. 276.

⁶⁸ Там же. С. 279.

и Хао Вана, становится ясным то, что Гедель отказался от первой части диллемы, что фактически роднит его с позицией менталистов.

Теоремы Геделя о неполноте связаны с работой, проделанной Тьюрингом. Именно появление понятия машины Тьюринга и связанные с ними концепции лежат в основе спора о соотношении машинного «интеллекта» и ума человека. Тьюринг полагал эквивалентность между мозгом и машиной, что и критиковал Гедель. Такое соотношение могло быть верным только, по мнению Геделя, если добавить следующие посылки:

1) Ум отдельно от материи не существует.

2) Функционирование мозга аналогично функционированию машины.

Гедель считал, что первая посылка будет опровергнута. Однако мы можем задаться вопросом: имеют ли отношения теоремы Геделя о неполноте к этим положениям, или какое-либо отношение к возможностям человеческого интеллекта? Э. Нагель и Дж. Ньюмен дали крайне утвердительный ответ на эти вопросы, когда проводили критику позиций Тьюринга. Они, исходя из теорем Геделя, полагали, что человеческие способности не формализованы до конца и принципиально не могут быть формализованы полностью.

Позиция Геделя на озаглавленный выше вопрос была не столь прямолинейна, как у Нагеля и Ньюмена. Гедель пришел к другим выводам, которые имеют высокую важность для решения вопроса соотношения человека и компьютера. Для научного сообщества эти выводы стали доступны в виде дилеммы, которую сам Гедель сформулировал следующим образом:

«Либо математика неполна (незавершаема) в том смысле, что очевидные аксиомы никогда не могут быть вмещены в конечные правила, то есть, что человеческий ум (даже в сфере чистой математики) бесконечно превосходит возможности любой конечной машины, либо существуют абсолютно неразрешимые диофантовые проблемы определенного типа (не исключается, что истинны оба члена этой дизъюнкции, так что, строго говоря, существуют три альтернативы)»⁶⁹.

Дилемму можно выразить и другим способом. Первая часть указывает на превосходство ума над машиной за счет того, что вся математика не может быть вмещена в конечные правила формальной системы. Вторая часть повествует нам о том, что математика делится на две части: субъективную и объективную, где субъективная отвечает познаваемым истинам, а объективная ответственна за платонический мир истин математики.

⁶⁹ Там же. С. 281.

Здесь нам необходимо остановиться на важном разграничении между умом и мозгом. Это разграничение можно выразить таким вопросом: что мы имеем в виду, когда говорим, что человеком доказываются теоремы с математической определенностью — человеческий мозг или его ум? Проблема заключается не только в том, какой термин мы предпочтем. Если считать, что способность доказательства является результатом отдельной части мозга, которая является аналогом компьютера, то, исходя из первой части дилеммы, ум превосходит мозг. Такой взгляд кажется довольно странным: ум человека превосходит мозг, место где это самый ум условно «находится». Полагаем, что для Геделя с данным заявлением не возникло бы проблем, так как из указанного нами выше им полагалось, что будет опровергнуто, что ум не способен существовать от материи, в нашем случае мозга. Следуя же второй части дилеммы, выходит так, что ум человека, приравненный к машине, не может познать ряд математических истин, которые становятся фактически неразрешимыми предложениями.

Как мы указывали выше, позиция Геделя в отношении постулирования превосходства ума над машиной отличалась от взглядов Нагеля и Ньюмена. Если ум способен решать все математические проблемы, то он должен решать вопрос о своей непротиворечивости. Менталисты решают эту проблему постулируя заранее, что математическое мышление было бы невозможно, если математик был бы внутренне противоречив.

Гедель же показывает разрешение этого вопроса следующим образом: если мы полагаем, что получение математического знания происходит по некоторому алгоритму, т. е. считаем, что способность человека к доказательству математических истин носит характер механического процесса, то у нас нет никаких гарантий, что получаемые результаты при такой деятельности будут непротиворечивыми. Иными словами, Гедель подводит нас к тому, что ум, трактуемый как машина, не может доказать свою непротиворечивость. При этом заявление ума о собственной непротиворечивости есть математическое предложение, и тогда, следуя второй теореме Геделя о не полноте, доказать внутри системы свою собственную непротиворечивость ум не может. Это значит, что ум, приравниваемый к машине, не имеет возможности понять принципы своего собственного функционирования.

6. Связь идей Геделя с идеями Лукаса и Пенроуза

Вопрос о том, в какой степени идеи Геделя, которые были высказаны им в Гиббсовской лекции, связаны с идеями Лукаса и Пенроуза остается открытым и этому есть ряд причин. Наиболее очевидной причиной является расхождение в общих философских позициях этих мыслителей. Так Лукаса, чаще всего, относят

к «идеалистам», так как он указал целью своей аргументации ниспровержение материализма. За Пенроузом признают материалиста, так как он считает, что перестроенная физика может объяснить ментальные состояния, а Геделя определяют к дуалистам.

Так же Геделем в указанной выше лекции вообще не употребляется понятие машины, когда он обращается к фундаментальным вопросам математического познания. У Лукаса же вопрос машины занимает важное место в структуре его аргументации.

И, наконец, в самой аргументации менталистов есть различия технического характера. Лукас применяет первую теорему Геделя о неполноте в неформальной аргументации, а вторую теорему Геделя приводит в популярной формулировке: «Если система S непротиворечива, это не может быть доказано в S ». Х. Патнэм указывает на то, что Лукас допускает ошибку, так как путает два различных утверждения, которые одновременно можно было назвать: «утверждение, что S непротиворечива». По мнению Патнэма, Лукас путает обычное высказывание о том, что методы, которые используют математики, непротиворечивы, с математическим утверждением, которое возникает при применении теорем Геделя к гипотетической формализации данных методов. Интересно, что Патнэм указывает на допущение схожей ошибки Нагелем и Ньюменом, т. е. неправильное применение теорем Геделя.

Пенроуз же в своей аргументации прибегает не к самой теореме Геделя, а к теореме Тьюринга, которая является ее следствием. Он применяет ее ко всей мировой активности всех математиков и утверждает, что эту активность не может схватить никакой алгоритм, никакой жесткий свод правил, вписанный в машину Тьюринга. Пенроуз, таким образом, полагает, что машина Тьюринга не может симулировать мозг человека.

Заключение

В данной статье мы познакомились с одним из основных аспектов проблематики алгоритмизации мышления: вопросом принципиальной возможности этой алгоритмизации. Нами был реконструированы основы дискуссии между двумя направлениями, дающими принципиально альтернативные ответы на поставленный вопрос: ментализмом и механизмом. Мы продемонстрировали, что ведущую роль, в рамках этого спора, играют различные интерпретации теорем Геделя о неполноте: мы выяснили, что интерпретации этих теорем дают возможность различных ответов на вопрос о соотношении человеческой и машинного «мышления».

Нами было показано, что характер дискуссии носил герменевтический характер, т. е. в аргументация, которая будучи отталкивающейся от математических фактов, так не обнаруживается «убийственных», неоспоримых аргументов.

Мы выяснили, что Гедель занимает позицию, родственную позиции менталистов, однако вопрос о связи идей, высказанных Геделем, с идеями Лукаса и Пенроуза так и остается открытым в виду различных философских воззрений мыслителей, использование терминов и технической реализации аргументации. При раскрытии позиции Геделя, а именно его взгляда на превосходство ума над компьютером, мы провели анализ, сформулированной им дилеммы. Мы выяснили, что если считать ум эквивалентным машине, то ум сталкивается с рядом проблем, одна из которых заключается в том, что в подобном случае человеческий ум не может понять принципы своего функционирования.