ИЗДАТЕЛЬСТВО грамота

Манускрипт • Manuscript

ISSN 2949-3242 (online)
ISSN 2618-9690 (print)

2026. Том 19. Выпуск 3 | 2026. Volume 19. Issue 3
Материалы журнала доступны на сайте (articles and issues available at): manuscript-journal.ru

RU

Историко-философские предпосылки проблемы искусственного интеллекта: от мифов к теории вычислимости

Мальцев В. А.

Аннотация. Цель исследования – реконструировать становление проблемы «разум и машина» от древних представлений до теории вычислимости, проследив основные этапы формализации рассуждения и её внутренние пределы. В работе рассмотрена линия формализации рассуждения, в которой Аристотель, Р. Луллий, Т. Гоббс, Г. В. Лейбниц, Дж. Буль, Г. Фреге, Д. Гильберт, А. Чёрч и А. Тьюринг последовательно превращают его в механически выполнимую процедуру. Одновременно выделена встречная линия, в которой Г. В. Лейбниц, К. Гёдель, А. Тарский и А. Тьюринг обнаруживают пределы формализации средствами самого формализма. Научная новизна исследования состоит в том, что разрозненные историко-философские эпизоды сведены в единый нарратив, который показывает развитие проблемы и выявляет в работах прошлого идеи и связи, продуктивные для современного анализа искусственного интеллекта как феномена и как участника научного познания. В результате исследования установлено, что оценка системы по внешне разумоподобному поведению восходит к мифу и получает методологическое оформление у А. Тьюринга. Операциональная эффективность процедуры не равнозначна субъектной разумности, а синтаксическая корректность не порождает семантики, поскольку уровень интерпретации лежит вне процедуры. Пределы формального метода заданы его устройством, а не дефицитом вычислительных ресурсов, следовательно, они не снимаются наращиванием масштаба систем. Вклад моделей ИИ в познание следует оценивать скорее по приросту проверяемости результата, нежели по правдоподобию ответа. Саму систему корректнее описывать как средство познавательной деятельности, а не как её субъект. Выдвинутые положения применимы в дальнейших исследованиях этой области.

EN

Historical and philosophical foundations of the artificial intelligence problem: from myths to computability theory

V. A. Maltsev

Abstract. The study aims to reconstruct the emergence and evolution of the “mind and machine” problem from ancient perceptions to computability theory, tracing the key stages in the formalization of reasoning along with its intrinsic limits. The article examines the lineage of the formalization of reasoning, wherein Aristotle, Ramon Llull, Thomas Hobbes, G. W. Leibniz, George Boole, Gottlob Frege, David Hilbert, Alonzo Church, and Alan Turing sequentially transformed reasoning into a mechanically executable procedure. Concurrently, a counter-lineage is identified in which Leibniz, Kurt Gödel, Alfred Tarski, and Turing uncovered the limits of formalization using the tools of formalism itself. The scientific novelty of the study lies in synthesizing disparate historical and philosophical episodes into a cohesive narrative that elucidates the development of the problem and uncovers ideas and interconnections in historical scholarship that remain productive for the contemporary analysis of artificial intelligence both as a phenomenon and as a participant in scientific inquiry. The study establishes that evaluating a system based on externally mind-like behavior traces back to myth and receives its methodological consolidation in Turing’s work. Operational procedural efficacy is not equivalent to subjective rationality, and syntactic correctness does not generate semantics, as the interpretive layer lies outside the formal procedure itself. The limits of the formal method are determined by its internal architecture rather than by computational resource constraints; consequently, they cannot be eliminated merely by scaling up systems. The contribution of AI models to scientific

Научная статья (original research article) | https://doi.org/10.30853/mns20260234

© 2026 Авторы. © 2026 ООО Издательство «Грамота» (© 2026 The Authors. © 2026 Gramota Publishing, LLC). Открытый доступ предоставляется на условиях лицензии CC BY 4.0 (open access article under the CC BY 4.0 license): https://creativecommons.org/licenses/by/4.0/

inquiry should be evaluated by the increase in the verifiability of results rather than by the plausibility of responses. The system itself is more accurately characterized as an instrument of cognitive activity rather than as its epistemic subject. The proposed tenets offer a framework for further inquiry in this domain.

Введение

Актуальность исследования обусловлена тем, что системы искусственного интеллекта (далее – ИИ) участвуют сегодня в научной работе, образовании и экспертном суждении, создавая внешне обоснованные утверждения, которые входят в состав знания прежде, чем установлен их познавательный статус. Таким системам приписывают предикаты понимания, рассуждения и намерения, тогда как условия применения этих предикатов остаются неустановленными. Распространено и допущение, по которому наращивание вычислительных ресурсов снимает философские ограничения. Между тем различие между выполнением формальной процедуры и мышлением разрабатывалось задолго до появления вычислительных машин, и обсуждение современных систем во многом опирается на предшествующие достижения философии, логики и математики. Отсюда необходимость обращения к историко-концептуальным предпосылкам проблемы «разум и машина», выраженной в наиболее общем виде вопросом о том, может ли машина мыслить и действовать разумно, подобно человеку (Георгиу, 2022, с. 32-35).

Для достижения вышеуказанной цели исследования необходимо решить следующие задачи:

– провести историко-философскую реконструкцию проблемы «разум и машина», начиная с её исходной мифологической формы;
– выявить основные этапы формализации рассуждения и установить, что достигается на каждом из них;
– обозначить пределы формализации, выделив неформализуемый остаток на каждом этапе;
– определить, чем полученные выводы полезны для современных обсуждений ИИ.

Методологическую основу исследования составляет диалектический метод, позволяющий рассматривать проблему «разум и машина» в её историческом развитии, взаимосвязи противоположно направленных процессов и качественном преобразовании проблемы. Применяются общенаучные методы анализа и синтеза, абстрагирования, обобщения и сравнения, а также принцип единства исторического и логического. Отдельно следует отметить проблемно-ориентированную историко-философскую реконструкцию, задающую основание отбора и сопоставления рассматриваемых позиций. Современные системы ИИ непосредственно не исследуются; применимость полученных выводов к нейросетевым архитектурам требует отдельной проверки.

Материал исследования составили «Первая аналитика» Аристотеля (1989), «Рассуждение о методе» и «Размышления о первой философии» Р. Декарта (1989; 1994), «Левиафан» Т. Гоббса (Hobbes, 1651), «Монадология» Г. В. Лейбница (1982), «Математический анализ логики» и «Исследование законов мышления» Дж. Буля (Boolle, 1847; 1854), «Исчисление понятий» и «Основные законы арифметики» Г. Фреге (1893; 1967), работа К. Гёделя (Gödel, 1962) о формально неразрешимых предложениях, «Понятие истины в языках дедуктивных наук» А. Тарского (1999), статьи А. Тьюринга (1937; 1950) о вычислимых числах и о вычислительных машинах и интеллекте. Учение Р. Луллия, формалистская программа Д. Гильберта и тезис А. Чёрча рассматриваются по исследовательской литературе, мифологический материал об искусственных существах – по работе А. Майор (Mayor, 2018).

Теоретическую базу исследования составляют работы по истории логики и силлогистике (Бочаров, 1984; Черноскутов, 2021), включая комментарий Р. Смита к «Первой аналитике» (Aristotle, 1989), исследования работ Р. Луллия (Koetsier, 2016; Crossley, 2011; Sales, 2011), литература о проекте универсальной характеристики (Milkov, 2006; Журавлева, 2012) и об аргументе «мельницы» (Lodge, 2014; Rozemond, 2014), работы о программе Д. Гильберта и теоремах о неполноте (Zach, 2007; Михайлова, 2015; Raatikainen, 2003), а также исследование истории тезиса А. Чёрча (Целищев, 2009). Дополнительно учитывались концепции, раскрывающие соотношение искусственной агентности и разумной субъектности (Шевцов, 2012; Георгиу, 2022).

Практическая значимость исследования состоит в том, что полученные в работе положения о границах формального метода, о различении операциональной эффективности и субъектности, а также о роли ИИ в познавательной деятельности могут быть применены при анализе эпистемического статуса современных систем ИИ. Результаты также могут учитываться при разработке нейросетевых моделей и при оценке заявляемых для них когнитивных возможностей.

Обсуждение и результаты

Устойчивые мифологические сюжеты о самодвижущихся артефактах, искусственно созданных существах, а также о границе между живым и изготовленным обнаруживаются еще в древней культуре. При этом важно подчеркнуть, что на данном этапе речь ещё не идёт о полноценных философских теориях – перед нами первичные формы осмысления проблемы.

Так, А. Майор (Mayor, 2018, p. 1-2) в исследовании «Боги и роботы: мифы, машины и древние мечты о технологиях» рассматривает мифы об искусственных существах – о Талосе, Пандоре, созданиях Дедала, Прометея и Гефеста – как ранние размышления о возможности имитации и даже возвышения над природой. Их объединяет не просто мотив оживления материи, а пограничный статус между живым и искусственным.

Особое значение имеет фигура Талоса – бронзовой статуи стражника Крита, искусственного самодвижущегося существа с устойчивой охранительной функцией: он распознаёт приближение чужих и производит эффект автономного поведения. Однако миф не наделяет Талоса речью, рефлексией и целеполаганием, его описание исчерпывается наблюдаемым поведением и телесными признаками, а вопрос о внутренних состояниях не ставится (Mayor, 2018, p. 7-19). При этом обнаруживается онтологическая двойственность. С одной стороны, Талос принадлежит к сфере технически изготовленного и сохраняет статус функционального агента, с другой – его жизнь зависит от кроваво-красной жилы, повреждение которой приводит к вытеканию ихора, утрате силы и смерти. Данный атрибут принадлежит не самому Талосу, а способу его описания: не располагая языком для искусственного как такового, миф характеризует изготовленное предикатами живого тела. Та же операция – перенос признаков одного онтологического уровня на другой – воспроизводится и в современном анализе ИИ: искусственная система оценивается через признаки человеческой разумности, но внешнее подобие разумного поведения ещё не тождественно сознанию, пониманию и субъектности.

Анализ мифа позволяет обнаружить недостаточность оценки по внешнему поведению, однако иного основания сам миф не даёт. Для того чтобы судить о том, означает ли выполнение процедуры мышление, требуется прежде установить, что в рассуждении вообще поддаётся формализации. Поэтому в дальнейшем необходимо установить возможность представления рассуждения как организованной по правилам и воспроизводимой процедуры.

Логическое учение Аристотеля, позднее объединённое в корпус «Органон», задаёт систему анализа высказываний и умозаключений. Как отмечает В. А. Бочаров, аристотелевское высказывание предполагает связь имени и глагола, причём трактовка последнего соответствует современному пониманию предиката; в силлогистике основные типы высказываний различаются по утверждению и отрицанию, а также по общности и частности (Бочаров, 1984, с. 41-44). В «Первой аналитике» силлогизм определяется как умозаключение, в котором из положенных посылок с необходимостью следует нечто отличное от них, причём без нужды в дополнительном термине (Aristotle, 1989, p. 2). Правильность вывода тем самым не зависит от его содержания и устанавливается по расположению терминов – по одной из трёх фигур, различающихся положением среднего термина. При этом из самого определения ещё не следует, что силлогизм обязательно должен иметь ровно две посылки или сразу сводиться к одной из фигур: такая редукция выступает результатом дальнейшего доказательства Аристотеля (Aristotle, 1989, p. 109-110). Значение трудов Аристотеля для проблемы «разум и машина» состоит в том, что формальная корректность вывода становится проверяемой независимо от психологического акта, в котором она совершается. Неформализованным остаётся приведение рассуждения к фигуре – задача, которую Аристотель рассматривает отдельно и решает средствами эвристики, приходя к выводу, что рассуждающий должен сам выделить посылки, определить их количество и восполнить пропущенное (Aristotle, 1989, p. 51, 161).

Следующий шаг в линии формализации рассуждения связан с Р. Луллием. *Ars lulliana* (лат.; англ. Llull's Art; рус. «луллианское искусство», или «метод Луллия») – философско-теологическая и логико-комбинаторная система, в которой базовые понятия, значимые для межрелигиозной полемики, переводятся в знаковую форму и через фигуры, таблицы и правила комбинации используются для рационального обоснования истины (Koetsier, 2016, p. 55, 61-67). Наглядно-механический характер этой системы проявляется, в частности, в четвёртой фигуре. Как отмечает Дж. Н. Кроссли (Crossley, 2011, p. 52), она включала три диска, независимо вращающихся вокруг центральной оси, и их вращение давало различные комбинации. Т. Салес (Sales, 2011, p. 29) усматривает здесь идею конечного набора правил и исходных истин, из которых порождается множество производных. Р. Луллий соединял базовые понятия линиями и предписывал следовать им при выведении следствий, так что рассуждение предстаёт движением по заданным связям (Sales, 2011, p. 30). Тем самым мыслительная операция получает знаково-комбинаторную организацию, что делает возможным порождение новых выводов.

Механика фигур даёт, однако, лишь соединения знаков. Их истинность связывалась не с процедурностью, а с тем, что исходные понятия мыслились как выражение божественных атрибутов, из которых истины следуют с необходимостью (Koetsier, 2016, p. 61-67). Поэтому правила Р. Луллия имеют преимущественно синтаксический характер, в то время как семантический вывод, как подчёркивает Дж. Н. Кроссли (Crossley, 2011, p. 46, 52-53), остаётся на стороне интерпретатора.

Следующий этап связан с новоевропейским механицизмом, в рамках которого механизм перестаёт пониматься только как внешний артефакт и становится общей объяснительной моделью природных и телесных процессов. У Р. Декарта (1989) этот подход выражается прежде всего в трактовке телесности животных. В пятой части «Рассуждения о методе» Р. Декарт (1989, с. 281-282) прямо сопоставляет тело животного с машиной, указывая на то, что человеческое искусство производит множество автоматов и самодвижущихся инструментов немногими деталями, тогда как в теле животного число частей несравненно больше, и потому такое тело правомерно рассматривать как машину, сделанную руками божьими. Действия, совершающиеся без участия мышления, объясняются устройством органов и материей, из которой тело составлено, а не признанием за ним разумной души (Декарт, 1989, с. 277). К. П. Шевцов (2012, с. 214-215) обращает внимание на разобранный Р. Декартом пример дворцовых фонтанов: фигуры в них кажутся самодвижущимися и осмысленно реагирующими, тогда как их движение производится скрытым механизмом – водой, трубами и рычагами. Видимость осмысленного действия, следовательно, ничего не добавляет к устройству, которое его производит. Отметим, что у Р. Декарта (1994, с. 62-63) вещь только мыслящая и непротяжённая противопоставлена вещи исключительно протяжённой и не мыслящей, и автомат относится ко второй из них, не претендуя на статус мыслящего Я. Из этого следует онтологическая неполнота автомата.

Р. Декарт (1989) проводит демаркацию между человеком и машиной, сходной с человеческим телом и подражающей его действиям, по двум признакам. Первый относится к языку. Машина откликалась бы на внешние воздействия заранее предусмотренными репликами, однако не соединяла бы слова по-новому, отвечая на то, что при ней сказано, – умение, доступное, по замечанию философа, и наименее развитому человеку. Дополнительно Р. Декарт (1989) отделяет производство знаков от осмысленной речи на примере птиц, отмечая, что они произносят слова не хуже людей, однако не говорят, поскольку говорить – значит показывать, что произносимое помыслено. Второй признак относится к области применения. Машина исполняла бы отдельные действия не хуже человека и даже лучше, но в других непременно оказалась бы несостоятельной, ибо действует не сознательно, а сообразно устройству своих органов, каждый из которых пригоден для строго определённой операции. Разум же есть универсальное орудие, служащее при самых разных обстоятельствах (Декарт, 1989, с. 283-284). Превосходство в отдельном действии, следовательно, не свидетельствует о разумности. Значима именно способность действовать в разнообразии контекстов, нежели точность в отдельной операции.

Вопрос о всеобщем языке составляет у Р. Декарта (1989) отдельный предмет рассмотрения. Как показывает Н. Милков (Milkov, 2006), в переписке с М. Мерсенном обсуждается проект идеографии, отражающей человеческие мысли и связанной с mathesis universalis (лат. «всеобщая математика»; проект общего метода познания по образцу математического рассуждения, основанного на порядке и мере); такой язык помогал бы суждению, представляя предметы столь ясно, что ошибка стала бы почти невозможной. Однако осуществление проекта требует, чтобы все мысли человеческого ума могли быть расположены в порядке, подобном естественному порядку чисел, – условие, ввиду которого Р. Декарт признаёт замысел чрезмерно трудным и утопическим (Milkov, 2006, p. 1-2).

У Т. Гоббса (Hobbes, 1651) проблема разума и машины получает развитие в анализе суждения и рассуждения. В пятой главе «Левиафана» он прямо определяет рассуждение как счёт. Рассуждая, человек складывает и вычитает следствия общих имён, используемых для обозначения и сообщения мыслей. Эта модель распространяется не только на числа, но и на геометрические объекты, политические решения, правовые отношения и логические следствия слов (Hobbes, 1651, p. 30). Дополнительно Т. Гоббс считает, что причиной абсурдных выводов выступает отсутствие порядка, то есть когда рассуждение начинается не с определений и устойчивых значений имён (Hobbes, 1651, p. 33).

Далее закономерно обратиться к философии Г. В. Лейбница (1982), поскольку именно у него линия формализации рассуждения получает одно из наиболее развёрнутых выражений в новоевропейской мысли. Проект универсального языка, сохранявший у Р. Декарта характер труднореализуемой возможности, становится здесь предметом последовательной философско-логической разработки.

Универсальная характеристика (Characteristica universalis) – проект символического исчисления, всеобщий метод получения достоверного знания и обоснования истинности суждений. Он двусоставен: характеристика есть одновременно и способ исчисления понятий, и всеобщий язык, обладающий собственной грамматикой (Журавлева, 2012, с. 69). Образцом точности выступает математика, поскольку числа однозначны и пригодны для проверки, однако характеристика не исчерпывается ею, так как арифметика и алгебра лишь лучшие из существующих её примеров, но не её предел. Универсальная характеристика призвана свести противоположные аргументы «к числовым терминам» так, чтобы «даже основания могли быть взвешены» (Milkov, 2006, p. 3). Рассуждение определяется тем самым как процедура. Спор переводится из области неопределённой естественной речи в область точного знака и правила, а истина устанавливается корректностью знаковой операции. Указанное понимание рассуждения сопоставимо с тем, на котором строится символическая модель ИИ, где интеллектуальная деятельность описывается через репрезентацию знания и процедуры вывода.

Значение проекта не ограничивается строгими науками. Н. Милков (Milkov, 2006, p. 3) обращает внимание на то, что наибольшую ценность Г. В. Лейбниц усматривал именно там, где прямое применение строгой процедуры невозможно и приходится действовать через предположение, оценку степеней и выбор альтернатив, в частности в медицине, военном искусстве и политике. Это позволяет рассматривать универсальную характеристику не только как проект формального доказательства, но и как проект рационализации практического суждения.

Ограничения проекта обнаруживаются, однако, в самой философии Г. В. Лейбница. Достоверное знание достижимо не во всякой области, поскольку сами истины бывают двух родов: истины разума и истины факта. Первые сводимы к тождественным высказываниям и потому в наибольшей мере поддаются исчислению; вторые связаны с бесконечным анализом условий и причин, полностью доступным только Богу (Журавлева, 2012, с. 72, 75-76). Их познание остаётся поэтому вероятностным, а формальная процедура не устраняет неопределённость, только упорядочивает её, оставляя за субъектом суждение и оценку степени обоснованности. Соответственно, и предел проекта задаётся строением своего предмета. Анализ условий восходит к бесконечности, и наращивание вычислительной мощности его не завершает.

Уязвимость обнаруживается и на уровне собственных предпосылок. Как отмечает А. С. Журавлёва (2012, с. 69), характеристика укоренена в метафизическом контексте монадологии и потому не сводима к логико-математическому построению. Исчисление держится на двух допущениях: что анализ понятий доходит до простых и что порядок знаков отвечает порядку вещей. Первое относится к познанию, второе к бытию, и обосновать их исчисление не может. Правильно выполненная операция даёт поэтому истинный результат лишь при верности допущений, взятых извне. Следовательно, данный проект нельзя представить как нейтральную технику, поскольку характеристика не работает на одном лишь устройстве знака, а опирается и на метафизику системы.

Кроме того, Г. В. Лейбниц (1982) сам обозначает предел механического объяснения психического. Он предлагает представить машину, устроенную так, чтобы мыслить, чувствовать и воспринимать, увеличенной настолько, чтобы в неё можно было войти, как в мельницу. При осмотре внутреннего устройства обнаружились бы лишь части, толкающие друг друга, но не нашлось бы ничего, что объясняло бы перцепцию (Лейбниц, 1982, с. 415-416). Следовательно, восприятие следует искать не в составной машине, а в простой субстанции. П. Лодж (Lodge, 2014, p. 79-82) также подчёркивает, что § 17 «Монадологии» направлен против объяснения ментального через одну только механическую организацию частей.

М. Роземонд (Rozemond, 2014, p. 7-9) предупреждает против прочтения «мельницы» как простой ошибки или аргумента от незнания. Перцепция не может быть модификацией материи, поскольку она активна, а материя пассивна, и сам Г. В. Лейбниц, понимая риск возражения, не сводил свою позицию к тому, что сознание внутри машины попросту не обнаруживается (Rozemond, 2014, p. 3, 12-13). Механическое взаимодействие частей объясняет движение и связь элементов, но не внутреннее единство восприятия как активного состояния. Тем самым представимость операций мышления в знаково-процедурной форме сама по себе не свидетельствует о наличии сознания, понимания и субъективного опыта.

То, что у Г. В. Лейбница оставалось замыслом всеобщего исчисления, в программе формализации Буля получает строго алгебраический характер и становится действующей системой операций над классами. Задачу своего исследования Дж. Буль (Boole, 1854, p. 1) определяет как выявление законов операций ума, посредством которых совершается рассуждение, и выражение этих законов на символическом языке исчисления. При сохранении правил соединения символов одна и та же формальная операция получает при этом разные интерпретации в различных областях знания (Boole, 1847, p. 3-4). Иными словами, содержание суждения временно отступает на второй план: предметом анализа становится его формальная структура, которая обрабатывается как операция над символами и затем вновь получает содержательную интерпретацию.

Как отмечает Ю. Ю. Черноскутов (2021, с. 5, 10-14), последовательная интерпретация терминов как классов в соединении с методами символической алгебры позволяет Дж. Булю построить более гибкую теорию силлогизма, выходящую за пределы ограничений аристотелевской традиции. Если у Аристотеля правильность вывода задаётся структурой силлогизма и отношениями между терминами, то теперь эти отношения получают форму алгебраических выражений и подлежат исчислению. Логика перестаёт быть классификацией правильных форм вывода и становится процедурой символического преобразования, что готовит понимание интеллектуальной деятельности как формальной операции над символами. Ретроспективно это подтверждается тем, что алгебра Дж. Буля стала важной основой современной символической логики и проектирования цифровых вычислительных схем.

Логический проект Г. Фреге (Frege, 1967) состоит в переходе от алгебраизации логики к построению специального формального языка, предназначенного для строгого представления структуры доказательства. В предисловии к «Исчислению понятий» Г. Фреге (1967) ставит задачу выработать максимально строгий метод доказательства, исключающий психологизм из анализа мысли. Каждый шаг вывода должен быть явно зафиксирован, подчинён общему правилу и свободен от скрытых интуитивных переходов (Frege, 1967, p. 5-8). «Исчисление понятий» выступает средством проверки рассуждения.

Сопоставляя свой проект с достижениями Дж. Буля (Boole, 1854), Г. Фреге (Frege, 1979, p. 10) указывает, что работал с тем же материалом, но «другим способом». Различие состоит в математической модели: Дж. Буль (Boole, 1854) обобщает алгебру и рассматривает логические отношения по аналогии с операциями над классами, тогда как Г. Фреге (Frege, 1967, p. 21-24) вводит в логику аппарат функции и аргумента. Поэтому его метод позволяет анализировать не только отношения между объёмами понятий, но и внутреннюю структуру суждения: предикатную форму, квантификацию, условные зависимости и последовательность доказательства (Frege, 1967, p. 12-17, 24-29). Именно это делает его проект более пригодным для формализации тех структур рассуждения, которые не сводятся к включению и пересечению классов.

Если у Дж. Буля (Boole, 1854) законы мышления сохраняют связь с анализом операций ума, то уже у Г. Фреге (Frege, 1893, S. XV-XVIII) логика отделяется от индивидуальных состояний субъекта – описывается не фактический процесс мышления человека, а правила, обеспечивающие корректность вывода. Тем самым формализация получает нормативный смысл в том, что доказательство должно быть проверяемой процедурой, независимой от интуиции рассуждающего, а механическая исполнимость такой процедуры оказывается следствием строгости, а не её целью.

В начале XX в. вопрос о формализации рассуждения получает новое развитие в формалистской программе Д. Гильберта. Как показывает Р. Зах (Zach, 2007, p. 411), указанная программа опиралась на идеи аксиоматического метода и финитной теории доказательств. Их связь отвечала задаче обосновать математику её собственными средствами. Теории, оперирующие бесконечным, должны были получить оправдание в рассуждении о конечных, обозримых конфигурациях знаков, надёжность которых не ставится под сомнение.

Н. В. Михайлова (2015) подчёркивает, что методологическое значение формализма Д. Гильберта состоит не в стремлении «переписать» математику символами, а в достижении предельной строгости, при которой само доказательство становится объектом исследования. Возникает метаматематика, где доказательства, формулы и правила вывода подлежат содержательному анализу (Михайлова, 2015, с. 534-541). Доказательство предстаёт при этом конечной синтаксической конфигурацией, поддающейся строго определённым преобразованиям, – объектом, о котором рассуждают с той же точностью, что о числах. Этим создаётся условие,

при котором рассуждение может рассматриваться как механически выполнимое. Ранее, когда доказательство служило способом получить результат, перечень допустимых шагов не был задан заранее и корректность каждого перехода зависела от суждения рассуждающего; теперь же оно допускает проверку на выполнимость без обращения к содержанию. Обоснование системы, однако, требует финитных средств, надёжность которых не подлежит сомнению, их Д. Гильберт характеризует через обозримость и интуитивную данность. Строгого определения финитному он не дал, и точный объём этого понятия остаётся спорным (Raatikainen, 2003, p. 160). Неформализованное основание, устранённое внутри системы, возвращается тем самым в её обоснование.

Развитие математической логики показало, однако, что полная формализация не устраняет проблем обоснования. К. Гёдель (Gödel, 1962, p. 57, 62, 65-66) рассматривает формальные системы, описанные выше, и устанавливает в первой теореме о неполноте, что если такая система достаточно выразительна для представления арифметики, её аксиомы и правила рекурсивно задаваемы и она ω-непротиворечива, то в ней существуют арифметические утверждения, не доказуемые и не опровержимые её собственными средствами. Вторая теорема усиливает этот вывод, поскольку достаточно сильная непротиворечивая система не может доказать собственную непротиворечивость своими средствами (Gödel, 1962, p. 70-71). Соответственно, относительность доказуемости неустранима, она определена принятой системой аксиом и правил, а усиление системы порождает лишь новое неразрешимое в ней утверждение. Превосходство истины над выводимостью теоремами при этом не доказывается – они устанавливают неразрешимость, тогда как истинность неразрешимого утверждения усматривается при содержательном прочтении арифметики.

Из этого не следует, будто формальный метод утратил своё значение. Различать нужно два утверждения. Первое состоит в том, что программа Д. Гильберта в первоначальном виде и полном объёме неосуществима. П. Раатикайнен (Raatikainen, 2003, p. 169) пишет, что она окончательно показана нереализуемой теоремами К. Гёделя (Gödel, 1962) и связанными результатами, причём уже первая теорема опровергает её, если требование консервативности признать для программы существенным (Raatikainen, 2003, p. 164). Второе утверждение принадлежит Р. Заху (Zach, 2007, p. 411), соединяющему тот же вывод с указанием на множество частных успехов программы и на порождённые ею успехи в логической теории и метатеории.

Следовательно, формализация как метод остаётся состоятельной и продуктивной. Несостоятельным оказывается притязание замкнуть её на себя, обосновав формальную систему её собственными средствами. Пределы лежат в области обоснования, а не в области применения, и установлены они средствами той метаматематики, которую программа Д. Гильберта и создала. Вопрос об истине, понимании и основаниях суждения формальной процедурой при этом не решается.

Проблема границ формализации также поднимается в семантической концепции истины А. Тарского (1999, с. 36-39), где он различает объектный язык, выражения которого являются предметом анализа, и метаязык, средствами которого задаются условия истинности этих выражений. Это различие связано с требованием материальной адекватности определения истины. Предложение считается истинным тогда и только тогда, когда имеет место утверждаемый им факт (Тарский, 1999, с. 22-24). При этом А. Тарский (1999, с. 33-36) отмечает, что язык, в котором определяется истинность, не должен смешиваться с языком, истинность выражений которого определяется; в противном случае возникает риск семантических парадоксов. Следовательно, формальное построение выражения, его преобразование по правилам и установление его истинностного статуса относятся к разным уровням анализа и для признания вывода истинным недостаточно его внутренней процедурной корректности, поскольку требуется внешний уровень семантической интерпретации.

Заключительный в рамках настоящей работы этап развития линии формализации рассуждения связан с работами А. Чёрча и А. Тьюринга, где формализуется понятие механически выполнимой процедуры. Необходимость такого шага обнаружилась там, где доказыванию подлежало отсутствие метода. Утверждение, что общего способа решения нет, относится ко всем возможным процедурам, перебрать которые поочерёдно нельзя. Отрицательное доказательство требует заранее установить, что считается допустимой процедурой, тогда как программа Д. Гильберта такого класса не задавала.

А. Чёрч подошёл со стороны формализма, отождествив эффективную вычислимость с λ-определимостью, позднее – с общерекурсивными функциями. Это отождествление он предлагал именно как определение вычислимости, после чего вывел из него неразрешимость отдельных задач (Целищев, 2009, с. 4-5). Основание такого шага оставалось внешним, поскольку обширность класса таких функций свидетельствовала лишь об объёме, оставляя без ответа вопрос, почему формализм охватывает всякое вычисление. Отсюда недоверие К. Гёделя к предложенному определению, не устранённое ни сменой формализма, ни доказанной вскоре эквивалентностью экспликаций (Целищев, 2009, с. 5). Поэтому следующим шагом становится анализ того, чем является само вычисление.

Требуемый ответ дал А. Тьюринг (Turing, 1937). Он также решал вопрос о том, что следует считать эффективно вычислимым, исходя при этом из анализа самого процесса вычисления. Разбирая деятельность вычисляющего человека, он удерживает лишь то, от чего зависит результат, – конечный набор различаемых знаков, ограниченность записи, обозримой в один момент, конечное число состояний и разложимость вычисления на элементарные операции над обозреваемым участком (Turing, 1937, p. 249-252). Такая деятельность, отвлечённая от прочих свойств вычислителя, получает представление в виде абстрактной вычислительной машины. Антропологическое основание конструкции отметил Л. Витгенштейн, заметив, что «машины Тьюринга – это просто люди, которые вычисляют» (цит. по: Целищев, 2009, с. 4). Этим и решается проблема отрицательного

доказательства, требующая заранее очертить класс всех допустимых процедур. Класс этот определяется условиями, при которых следование правилу вообще возможно, – обозримостью знаковой конфигурации и локальностью операций над ней. К. Гёдель характеризовал достигнутое как точное определение понятия механической вычислимости, эпистемологического по происхождению, и связывал с этой работой возможность строго определить общее понятие формальной системы (Целищев, 2009, с. 4-6). Образцом процедуры служит здесь человек, взятый со стороны механически организованного действия. Для построения теории А. Тьюринг удерживает в нём лишь механически воспроизводимое, отвлекаясь от понимания, целеполагания и усмотрения смысла. Соответствие работы машины человеческому вычислению воспроизводит ту сторону человеческой деятельности, которая была отобрана заранее, и независимого основания судить о разумности машины оно не даёт. Успех модели подтверждает точность отбора, оставляя вопрос о разумности открытым.

Получает строгое определение и понятие механической процедуры, недостающее программе Д. Гильберта, который добивался независимости вывода от индивидуального усмотрения, и цена достигнутого состоит в том, что образцом процедуры послужил именно вычисляющий человек, из вывода прежде устранённый. Место общего понятия процедуры занимает у А. Тьюринга (Turing, 1937, p. 248) отождествление выражений «существует общий процесс определения» и «существует машина, которая определит», оправданность которого он прямо связывает с оправданностью самого определения вычислимого. Определение опирается на конечные операции над обозримыми конфигурациями знаков и потому проверяемо само, причём класс процедур задан условиями выполнимости помимо средств какого-либо отдельного исчисления. Соответствие этих условий всему, что естественно считать вычислимым, сохраняет статус тезиса, доказательства не имеющего, поскольку свои доводы А. Тьюринг (Turing, 1937, p. 249) сам называет обращениями к интуиции и математически неудовлетворительными.

На этом основании доказуемо и отсутствие процедуры. А. Тьюринг (Turing, 1937, p. 246-248) устанавливает, что нет машины, которая по описанию произвольной машины определяла бы, печатает ли та бесконечно много знаков первого рода, равно как и машины, определяющей, напечатает ли произвольная машина заданный знак. Общего способа установить для произвольной формулы логики предикатов, доказуема ли она, поэтому не существует, поскольку он позволял бы решить и вопрос о печати заданного знака, уже показанный неразрешимым (Turing, 1937, p. 259, 262). Отсутствие общей процедуры не означает вместе с тем наличия внемашинного средства решения той же задачи. Предел установлен для машин, между тем отсутствие подобного предела у человеческого интеллекта, по замечанию А. Тьюринга (Turing, 1950, p. 444-445), его возражатели предполагают без доказательства. Установлен он внутри метода его же средствами, как у К. Гёделя (1962) и А. Тарского (1999), и относится к устройству метода, оставаясь безразличным к состоянию техники и к наращиванию мощности. Полученные результаты показывают, что для некоторых свойств поведения произвольных вычислительных систем универсальной разрешающей процедуры не существует. Это задаёт принципиальные ограничения для полной автоматизации их верификации.

А. Тьюринг (Turing, 1937) отказывается определять термины «машина» и «мыслить», заменяя вопрос о мышлении машин вопросом об исходе имитационной игры, где роль участника играет цифровая вычислительная машина, оцениваемая по неотличимости её ответов от человеческих. Исходную формулировку он признаёт настолько бессмысленной, что обсуждать её не стоит (Turing, 1950, p. 433-434, 436, 442). Сущностное определение мышления выносится за скобки, основанием остаётся наблюдаемое поведение.

Так, механически выполнимая процедура получила точное определение, вследствие чего линия формализации пришла к завершению в виде строго определённого класса. Тем же средством установлен и предел самого метода, обнаруженный на стороне отрицательных результатов. Меняется вследствие этого и формулировка проблемы. Прежде спрашивали о мере, в которой рассуждение представимо как формальная процедура, теперь под вопросом оказывается статус системы, эту процедуру реализующей, и для его оценки требуется иной философский аппарат – функционалистское понимание психического, его феноменологическая критика, различение синтаксической корректности и семантической осмысленности.

Заключение

Реконструкция позволяет сформулировать следующие выводы:

1. Первые формы проблемы обнаруживаются в мифах об искусственных существах, где самодвижущийся артефакт получает предикаты живого и человеческого. Философски такая оценка недостаточна, поскольку переносит признаки одного онтологического уровня на другой. К середине XX в. поведенческое основание получает методологическое оформление, ведь А. Тьюринг судит о машине по неотличимости её ответов от человеческих. Та же логика воспроизводится в современном анализе ИИ, где человеческие характеристики приписываются системам на основании внешне разумоподобного поведения.

2. До появления вычислительных машин проблема развивалась как последовательное усиление формальной и операциональной стороны рассуждения: у Аристотеля вывод получает логическую структуру, у Р. Луллия – знаково-комбинаторную процедуру, у Т. Гоббса определяется как счёт над именами, у Г. В. Лейбница – как проект универсального исчисления, у Дж. Буля выражается алгебраически, у Г. Фреге переходит в специальный логический язык, у Д. Гильберта само доказательство становится конечной процедурой и предметом метаматематики. А. Чёрч и А. Тьюринг определяют само понятие формальной процедуры, и операционализируемое перестаёт быть интуитивным представлением, становясь строго заданным классом.

3. Механицизм Нового времени превращает машину в модель живого. Граница между автоматом и разумом у Р. Декарта при этом сохраняется, поскольку автомат достигает безупречности в отдельной операции, разум сохраняет применимость в любых обстоятельствах. Универсальная машина А. Тьюринга этой границы не отменяет, оставаясь универсальной внутри класса механически выполнимых процедур. Отсюда рабочий критерий, по которому частная эффективность системы ещё не свидетельствует о познавательной субъектности.

4. Сильнейшие проекты операционализации ограничены изнутри. Р. Декарт обнаруживает это первым, всеобщая идеография предполагает мысли, упорядоченные подобно числам, – условие, невыполнимость которого и делает проект утопическим. Г. В. Лейбниц аргументом «мельницы» показывает невыводимость перцепции из взаимодействия частей, а сам его проект опирается на метафизику монадологии, что лишает его нейтральности. К. Гёдель устанавливает неполноту достаточно сильных формальных систем, А. Тарский – необходимость различения объектного языка и метаязыка. Меняется не значимость формального метода, а понимание его границ и условий использования. По отношению к ИИ данные результаты дают основание предполагать, что ограничения, обусловленные устройством формального метода, не снимаются одним лишь наращиванием масштаба систем; применимость данного вывода к нейросетевым архитектурам требует отдельной проверки.

5. Всю реконструкцию сопровождает различение синтаксиса и семантики. У Р. Луллия правила комбинируют знаки, тогда как смысл вывода остаётся за интерпретатором. Р. Декарт замечает, что произносить слова ещё не значит говорить, если произносимое не помыслено. У Дж. Буля независимость операции от содержания становится достоинством метода, у А. Тарского – требованием метаязыка. Уровень интерпретации лежит вне процедуры, поэтому безупречная форма машинного ответа о понимании не свидетельствует.

6. Историческим достижением линии формализации оказалась проверяемость вывода. У Г. Фреге логика описывает правила признания вывода корректным, минуя фактический ход мысли, Д. Гильберт делает проверяемым само доказательство, А. Чёрч и А. Тьюринг определяют класс механически проверяемого. Автоматическая исполнимость выступает следствием достигнутой строгости. Значит, и вклад ИИ в познание уместно измерять приростом проверяемости результата, нежели правдоподобием ответа.

7. Изначально формальное средство замышлялось как опора суждения, а не как его замена. Приведение рассуждения к фигуре остаётся у Аристотеля делом рассуждающего, надёжность финитных средств удостоверяется у Д. Гильберта интуитивной данностью, а наибольшую ценность исчисления Г. В. Лейбниц видел там, где строгая процедура неприменима и истины факта познаются лишь вероятностно. Уместнее поэтому описывать ИИ в качестве средства познавательной деятельности, чем в качестве её субъекта.

Перспективы дальнейших исследований состоят в разработке полученных положений. Предстоит выработать не заимствованный у человеческой разумности аппарат описания искусственных систем, проверить применимость выводов линии формализации к нейросетевым архитектурам, операционализировать критерий переноса решений между контекстами и разграничить техническую невозможность и концептуальную недостаточность. Отдельного рассмотрения требуют внешний уровень интерпретации, процедуры верификации машинных ответов и методология применения ИИ в познавательной деятельности.

Материалы исследования | Research materials

1. Декарт Р. Рассуждение о методе, чтобы верно направлять свой разум и отыскивать истину в науках // Декарт Р. Сочинения: в 2 т. / сост., ред., вступ. ст. В. В. Соколова. М.: Мысль, 1989. Т. 1.
2. Декарт Р. Размышления о первой философии // Декарт Р. Сочинения: в 2 т. / сост., ред. и примеч. В. В. Соколова; пер. с лат. С. Я. Шейнман-Топштейн. М.: Мысль, 1994. Т. 2.
3. Лейбниц Г. В. Монадология / пер. с фр. Б. П. Боброва // Лейбниц Г. В. Сочинения: в 4 т. / ред., сост., авт. вступ. ст. и примеч. В. В. Соколова. М.: Мысль, 1982. Т. 1.
4. Тарский А. Понятие истины в языках дедуктивных наук / пер. с польск. В. Л. Васюкова // Философия и логика Львовско-Варшавской школы / сост. В. А. Смирнов, В. Л. Васюков; отв. ред. Е. Н. Шульга. М.: Российская политическая энциклопедия (РОССПЭН), 1999.
5. Aristotle. Prior Analytics / transl. with introd., notes, and comment. by R. Smith. Indianapolis – Cambridge: Hackett Publishing Company, 1989.
6. Boole G. The Mathematical Analysis of Logic: Being an Essay Towards a Calculus of Deductive Reasoning. Cambridge: Macmillan, Barclay, & Macmillan, 1847.
7. Boole G. An Investigation of the Laws of Thought, on Which Are Founded the Mathematical Theories of Logic and Probabilities. L. – Cambridge: Walton and Maberly: Macmillan and Co., 1854.
8. Frege G. Grundgesetze der Arithmetik: Begriffsschriftlich abgeleitet. Jena: Verlag von Hermann Pohle, 1893. Bd. 1.
9. Frege G. Begriffsschrift, a Formula Language, Modeled upon That of Arithmetic, for Pure Thought / transl. by S. Bauer-Mengelberg // From Frege to Gödel: A Source Book in Mathematical Logic, 1879-1931 / ed. by J. van Heijenoort. Cambridge: Harvard University Press, 1967.
10. Frege G. Boole's Logical Calculus and the Concept-Script // Frege G. Posthumous Writings / ed. by H. Hermes, F. Kambartel, F. Kaulbach; transl. by P. Long, R. White. Oxford: Basil Blackwell, 1979.
11. Gödel K. On Formally Undecidable Propositions of Principia Mathematica and Related Systems / transl. by B. Meltzer; introd. by R. B. Braithwaite. Edinburgh – L.: Oliver and Boyd, 1962.

12. Hobbes T. Leviathan, or The Matter, Forme and Power of a Commonwealth Ecclesiasticall and Civil. L., 1651.
13. Turing A. M. On Computable Numbers, with an Application to the Entscheidungsproblem // Proceedings of the London Mathematical Society. 1937. Ser. 2. Vol. 42. No. 1. https://doi.org/10.1112/plms/s2-42.1.230
14. Turing A. M.Computing Machinery and Intelligence // Mind. 1950. Vol. 59. No. 236. https://doi.org/10.1093/mind/LIX.236.433

Источники | References

1. Бочаров В. А. Аристотель и традиционная логика: анализ силлогистических теорий. М.: Изд-во Московского университета, 1984.
2. Георгиу Т. С. Решение проблемы «сознание – тело» и искусственный интеллект // Вестник Тверского государственного университета. Серия: Философия. 2022. № 4 (62). https://doi.org/10.26456/vtphilos/2022.4.032
3. Журавлева А. С. Универсальная характеристика в метафизическом проекте Г. В. Лейбница // Вестник Пермского университета. Философия. Психология. Социология. 2012. Вып. 2 (10).
4. Михайлова Н. В. Программа формализма Гильберта как работающее философское направление обоснования математики // Российский гуманитарный журнал. 2015. Т. 4. № 6. https://doi.org/10.15643/libartrus-2015.6.12
5. Целищев В. В. История тезиса Черча // Гуманитарные науки в Сибири. Серия: Философия и право. 2009. № 1.
6. Черноскутов Ю. Ю. О силлогистике Дж. Буля // Дискурс. 2021. Т. 7. № 2. https://doi.org/10.32603/2412-8562-2021-7-2-5-15
7. Шевцов К. П. Память телесных автоматов в механической вселенной Декарта // Исторические, философские, политические и юридические науки, культурология и искусствоведение. Вопросы теории и практики. 2012. № 7-2 (21).
8. Crossley J. N. Ramon Llull's Contributions to Computer Science // Ramon Llull: From the Ars Magna to Artificial Intelligence / ed. by A. Fidora, C. Sierra. Barcelona: Artificial Intelligence Research Institute, IIIA-CSIC, 2011.
9. Koetsier T. The Art of Ramon Llull (1232-1350): From Theology to Mathematics // Studies in Logic, Grammar and Rhetoric. 2016. Vol. 44. No. 1. https://doi.org/10.1515/slgr-2016-0004
10. Lodge P. Leibniz's Mill Argument Against Mechanical Materialism Revisited // Ergo. 2014. Vol. 1. No. 3. https://doi.org/10.3998/ergo.12405314.0001.003
11. Mayor A. Gods and Robots: Myths, Machines, and Ancient Dreams of Technology. Princeton – Oxford: Princeton University Press, 2018. https://doi.org/10.2307/j.ctvc779xn
12. Milkov N. A New Interpretation of Leibniz's Project for Characteristica Universalis // Hans Poser, Einheit in der Vielheit, Proceedings of the 8th International Leibniz-Congress, 2006.
13. Raatikainen P. Hilbert's Program Revisited // Synthese. 2003. Vol. 137. https://doi.org/10.1023/A:1026287119565
14. Rozemond M. Mills Can't Think: Leibniz's Approach to the Mind-Body Problem // Res Philosophica. 2014. Vol. 91. No. 1. https://doi.org/10.11612/resphil.2014.91.1.1
15. Sales T. Llull as Computer Scientist, or Why Llull Was One of Us // Ramon Llull: From the Ars Magna to Artificial Intelligence / ed. by A. Fidora, C. Sierra. Barcelona: Artificial Intelligence Research Institute, IIIA-CSIC, 2011.
16. Zach R. Hilbert's Program Then and Now // Philosophy of Logic / ed. by D. Jacquette. Amsterdam: Elsevier, 2007. https://doi.org/10.1016/B978-044451541-4/50014-2

Информация об авторах | Author information

RU **Мальцев Владислав Антонович**[1]
[1] Вятский государственный университет, г. Киров

EN **Vladislav Antonovich Maltsev**[1]
[1] Vyatka State University, Kirov

[1] *vladmaltsev71@yandex.ru*

Информация о статье | About this article

Дата поступления рукописи (received): 01.09.2026; опубликовано online (published online): 22.09.2026.

Ключевые слова (keywords): философия искусственного интеллекта; машина Тьюринга; формализация рассуждения; теоремы Гёделя о неполноте; субъект познания; philosophy of artificial intelligence; Turing machine; formalization of reasoning; Gödel's incompleteness theorems; epistemic subject.