

УДК 165,8 + 004.81
DOI: 10.15372/PS20240505
EDN: NQYHVX

Н. И. Кожокару

ФИЛОСОФСКИЕ КОНЦЕПЦИИ И. КАНТА КАК ПРЕДПОСЫЛКИ СОЗДАНИЯ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА*

В статье даны основные определения искусственного интеллекта, приведены основные положения опередившей свое время теории познания И. Канта, отмечено их соответствие теоретическому обоснованию создания искусственного интеллекта как решение первой части триединой задачи по созданию антропоморфного искусственного интеллекта. На основании изучения научных работ исследователей, непосредственно занимающихся созданием систем искусственного интеллекта, было установлено, что не только теория познания Канта, но и его трансцендентальная логика, а также априорная психология являются теоретической основой для создания суперсовременных систем искусственного интеллекта. Сочетание когнитологии Канта, трансцендентальной логики Канта и априорной психологии Канта современные исследователи, работающие в области создания искусственного интеллекта, называют принципами Канта. По их мнению, без соблюдения указанных принципов создание сильного антропоморфного искусственного интеллекта практически невозможно. Созданию искусственного интеллекта препятствуют серьезные трудности в области естественных наук, в частности нейрофизиологии, весьма далекой от объяснения принципов работы человеческого мозга. В этой связи понимание механизма работы созданных систем искусственного интеллекта также оставляет желать лучшего – порой сами разработчики не способны определить, каким путем были получены конечные выводы в ходе запрограммированной обработки исходных данных. В связи с указанными проблемами автором проведена параллель между взрослением ребенка и возрастанием объема знаний человечества, который, возможно, пока не соответствует решению задачи по созданию сильного искусственного интеллекта.

Ключевые слова: искусственный интеллект; теория познания Канта; трансцендентальная логика; априорная психология; объяснимый искусственный интеллект

* Статья подготовлена в рамках государственного задания ГАУТН «Цифровизация и формирование современного информационного общества: когнитивные, экономические, политические и правовые аспекты». Регистрационный номер НИОКТР123022000042-0. Код темы FZNF-2023-0004. Регистрационный номер темы 1022040800826-5-5.2.1;6.3.1;5.9.1.

N. I. Kozhokaru**KANT'S PHILOSOPHICAL CONCEPTS AS PREREQUISITES FOR THE CREATION OF ARTIFICIAL INTELLIGENCE**

The article gives the basic definitions of artificial intelligence, cites the main provisions of Kant's theory of cognition, which was ahead of its time, and notes their correspondence to the theoretical justification for the creation of artificial intelligence as a solution to the first part of the triune task of creating anthropomorphic artificial intelligence. Based on the study of scientific works of researchers directly involved in the creation of artificial intelligence systems, it was found that not only Kant's theory of cognition, but also his transcendental logic and a priori psychology are the theoretical basis for the creation of ultramodern artificial intelligence systems. The combination of Kant's cognitive science, Kant's transcendental logic and Kant's a priori psychology is called Kant's principles by modern researchers working in the field of artificial intelligence creation. In their opinion, it is practically impossible to create a strong anthropomorphic artificial intelligence without observing these principles. The creation of artificial intelligence is hindered by serious difficulties in the field of natural sciences, in particular neurophysiology, which is very far from explaining the principles of the human brain. In this regard, the understanding of the mechanism of operation of the created artificial intelligence systems also leaves much to be desired – sometimes the developers themselves are unable to determine how the final conclusions were obtained in the course of programmed processing of initial data. In connection with these problems, the author drew a parallel between the growing up of a child and the increasing amount of knowledge of mankind, which, perhaps, does not yet correspond to the solution of the task of creating a strong artificial intelligence.

Keywords: artificial intelligence, Kant's theory of cognition, transcendental logic, a priori psychology, explainable artificial intelligence.

В Encyclopaedia Britannica «способность компьютера выполнять задачи, которые могут выполнять разумные существа» определяется как «искусственный интеллект» (ИИ). Помимо приведенного выше точного, но общего определения, наиболее информативны, на наш взгляд, следующие:

- научные исследования, направленные на моделирование процессов познания и мышления;
- различные устройства, механизмы, программы, которые в соответствии с определенными критериями можно было бы назвать интеллектуальными;
- комплекс понятий о знаниях, человеке и о разуме, которые позволяют актуализировать моделирование интеллекта [11; 29; 31; 32].

Впервые термин «искусственный интеллект» использовал американский ученый информатик Д. Маккарти в 1956 году [33].

За 175 лет до этого, в 1781 году, был опубликован основной философский труд Канта «Критика чистого разума» [1]. Похоже, никто с тех пор не подошел настолько близко к пониманию того, как происходят мыслительные процессы в голове homo sapiens. В этой связи обращение ученых, стремящихся создать компьютер, способный

выполнять «задачи, которые могут выполнять разумные существа», к научному наследию Канта представляется совершенно естественным, как с точки зрения формальной, так и трансцендентальной, кантовской, логики.

По мнению ряда современных философов [2; 19; 36], а также ученых различных специальностей, работающих в области создания искусственного интеллекта, Кант создал теоретическую основу для современных исследований в области информационных технологий, в частности создания искусственного интеллекта [3, 15, 16, 17].

Речь, безусловно, идет об антропоморфном варианте ИИ. Для Канта, совершившего коперниканский переворот в философии, философия – это «наука об отношении всякого знания к существенным целям человеческого разума». В антропологическом повороте, автором которого считается Сократ, а последователями – Протагор, Р. Декарт, Э. Гуссерль, М. Шелер, важнейшая роль, по мнению М. Хайдеггера и М. Фуко, также принадлежит Канту [18; 42]. Он, сформулировав в «Критике чистого разума» три основных вопроса философии: «Что я могу знать? Что я должен делать? На что я смею надеяться?», в «Логике» задал четвертый: «Что есть человек?» «На первый вопрос, – говорит Кант, – отвечает метафизика, на второй – мораль, на третий – религия, и на четвертый – антропология. Но в сущности все это можно было бы свести к антропологии, ибо три первых вопроса относятся к последнему» [1; 27; 28].

Как нам представляется, создание сильного ИИ антропоморфного характера предполагает решение триединой задачи.

Первая часть – когнитологическое обоснование возможности создания ИИ, понимание, что происходит в процессе познания. Эта часть триединой задачи отвечает на вопрос «Что происходит?»

Вторая часть – понимание, как происходит процесс познания в головном мозге человека, ответ на вопрос «Как происходит?»

И третья часть – уточнение, как этот процесс скопировать, смоделировать, то есть создать ИИ: «Как сделать, повторить, смоделировать?»

Согласно данным изданной после первой научной конференции по ИИ, состоявшейся в 1956 году в Дартмуте, практически однозначно мнение о том, что наибольший вклад в решение первой части триединой задачи был внесен более 250 лет назад Кантом. Так, в 2022 году в Великобритании была издана книга «Кант и искусственный интеллект», содержащая десятки статей как философов, так и специалистов в области ИИ [36].

Д. Маккарти, наиболее известный специалист, родоначальник

исследований в области ИИ и автор термина «искусственный интеллект», к основополагающим принципам построения ИИ относит внутреннюю согласованность и всеобщность, а также априорность математики, что полностью совпадает с воззрениями когнитологии Канта. Несмотря на некоторые философские разногласия, Маккарти, автор и сторонник трансцендентального реализма, в отличие от трансцендентального идеализма Канта признает уникальность вклада последнего в теоретическое обоснование ИИ [33].

С.Л. Катречко так и назвал свою статью – «Кантовская концепция сознания (познания) как модель “искусственного интеллекта”» [2]. Какова же, по Канту, общая структура элементарного познавательного акта, приводящая к получению синтетического знания?

Первый этап – первоначальный синтез схватывания. О нем Кант пишет: «синтез схватывания, имеющий эмпирический характер, необходимо должен сообразоваться с синтезом апперцепции, который имеет интеллектуальный характер и содержится в категории совершенно *a priori*» [1; 26].

Второй – синтез апперцепции. По Канту, «синтетическое единство апперцепции есть высший пункт, с которым следует связывать все применение рассудка, даже всю логику и, вслед за ней, трансцендентальную философию, более того, эта способность и есть сам рассудок» [26].

Третий – фигурный (пространственный) синтез.

Четвертый – схематический (временной) синтез.

Пятый – образно-понятийный синтез.

Шестой – заключительный пропозициональный понятийно-понятийный синтез [1].

По второй части триединой задачи дело как за философами, объясняющими общие закономерности мышления человека, так и за специалистами-нейрофизиологами, биофизиками и биохимиками. В этом направлении сделан гигантский шаг вперед. Специалисты в области теории нейронных сетей У. Маккалок и В. Питтс сумели доказать, что любая функция естественной нервной системы, которая логически может быть описана конечным числом слов, осуществима через во многом эквивалентную ЭВМ формальную нервную сеть; данные функции можно моделировать. Но возможно ли строго логично описать работу мозга? С точки зрения формальной, аристотелевской логики – нет. Тут, по словам современных ученых, работающих в области создания сильного искусственного интеллекта, на помощь приходит трансцендентальная логика Канта: «Мы ... устанавливаем идею науки о чистом, происходящем из рассудка и разума, знании, посредством

которого предметы мыслятся вполне а priori. Такая наука, определяющая происхождение, объем и объективное значение подобных знаний, должна называться трансцендентальной логикой» [1; 35; 43].

Однако, что касается успехов нейрофизиологов, биофизиков, биохимиков и других ученых, изучающих механизмы работы человеческого мозга на клеточном уровне, оснований для оптимизма немного. Для примера приведем результаты работы ученых исследовательской лаборатории Института Солка, которые в рамках научного сотрудничества с аналогичными группами коллег по всему миру изучили более полумиллиона клеток головного мозга с целью их типирования и классификации. Статья по результатам данного исследования, опубликованная в журнале «Science» менее года назад, является первой научной работой, в которой методы типирования и классификации клеток головного мозга, до сих пор примененные на мышах, были использованы при изучении человеческого мозга. По словам профессора Д. Экера, директора лаборатории геномного анализа Солка и исследователя Медицинского института Г. Хьюза, «опубликованные данные представляют собой первые тесты того, могут ли эти подходы работать на образцах человеческого мозга, и мы были воодушевлены тем, насколько хорошо они реализуются. Это действительно начало новой эры в науке о мозге, когда мы сможем лучше понять, как мозг работает, развивается, стареет и подвергается влиянию болезней». Ясно, что до полного понимания работы мозга, его анатомических особенностей и физиологических механизмов мышления человека еще очень далеко. В отличие от объясняющей общие закономерности человеческого мышления теории познания и трансцендентальной логики Канта, истинность и практическая значимость которых была неоднократно подтверждена результатами естественнонаучных исследований, в том числе в области создания первых систем искусственного интеллекта [41].

Для краткого анализа состояния третьей части триединой задачи опишем систему формирования антропоморфного ИИ. Ее «мозг» представляет активную ассоциативную однородную структуру, многомерную рецепторно-эффекторную нейроподобную растущую сеть, состоящую из множества нейроноподобных элементов, связанных синоптическими связями. Несмотря на то, что создание нейроноподобного элемента – упрощенной модели биологического нейрона – это огромное достижение, которое позволяет решать массу важных для человечества задач, сильный ИИ пока не создан. Современные исследователи в различных странах мира упорно работают над решением третьей части триединой задачи. В результате создано множество систем, применяющих различные алгоритмы искусственного интеллекта,

которые зачастую представляют некое подобие «черного ящика». Во многих случаях не только пользователи результатов применения искусственного интеллекта, но и сами разработчики конкретной системы не способны определить, каким путем были получены конечные выводы в ходе запрограммированной обработки исходных данных. В этой связи появился особый и распространенный не только в научной среде термин «объяснимый искусственный интеллект (Explainable AI, XAI)» [34].

Проблемой XAI занимаются ведущие научные и образовательные учреждения мира, в частности Университет Карнеги Меллон (США), Эдинбургский университет (Великобритания), Технологический университет Мичигана (США), Корнелльский университет (США), Университет Дьюк (США), Борнмутский университет (Великобритания), Лондонский университет (Великобритания), Норвежский университет естественных и технических наук и другие. И это неспроста. Сомнения в безопасности и потребность в XAI со стороны государства и общества изложена в ряде официальных документов, таких как статья 22 Общего регламента по защите данных (ЕС)¹, которая дает каждому гражданину Европейского экономического сообщества право узнать, каким образом было принято решение, которое его касается, в частности использовании той или иной автоматизированной системы, в частности искусственного интеллекта. Закон об ответственности за работу алгоритмов (США) обязывает организации и компании обеспечить полное получение сведений о риске для конфиденциальности или безопасности гражданина, которые могут быть созданы системой искусственного интеллекта, а также о рисках принятия ошибочных, несправедливых, предвзятых или дискриминационных решений. При этом уже установлены как изначально интерпретируемые системы ИИ (линейная и логистическая регрессия, деревья решений и другие), так и системы по типу «черного ящика» (перцептрон, сверточная и рекуррентная нейронные сети, сеть долгосрочной кратковременной памяти и прочие). Так или иначе, удачные или не очень попытки создания различных типов искусственного интеллекта сводятся к реализации когнитивной архитектуры Канта, не вызывающей опасений или сомнений в своей истинности [14; 34].

Р. Эванс, соискатель ученой степени доктора философии в области вычислительной техники Имперского колледжа Лондона, в своей диссертации излагает принципы создания системы индуктивно-логического программирования Apperception Engine, которая решает целый

¹ URL: <https://eur-lex.europa.eu/eli/reg/2016/679/oj>

ряд проблем, преследовавших нейронные сети с момента их создания, путем перехода к гибридной нейросимволической архитектуре, сочетающей в себе низкоуровневое восприятие и высокоуровневую апперцепцию [15]. По словам Эванса, «Apperception Engine в его нынешнем виде является результатом некоторых фундаментальных решений, которые были приняты на ранних этапах проекта, а также ответов на некоторые основные вопросы о том, как интерпретировать и реализовывать Канта». Приведем те вопросы, ответы на которые искал и в той или иной мере нашел Эванс, чтобы констатировать, что без философских концепций Канта создание суперсовременной и наиболее перспективной системы ИИ было бы невозможно:

1. Когда Кант говорит, что каждая последовательность детерминаций должна быть подкреплена каузальным правилом, означает ли это, что (1) должно существовать каузальное правило, в которое верит агент? Или, что гораздо слабее, (2) агент должен просто верить в существование каузального правила?

2. Когда Кант говорит, что суждения – это правила, имеет ли он в виду (1) эксплицитные правила, сформированные из дискретных символов? Или он может иметь в виду, что некоторые суждения – это просто (2) имплицитные правила (например, процедура, которая подразумевается в весах нейронной сети)?

Каков же ответ на первый вопрос Эванса? Во Второй аналогии опыта Кант пишет: «Все изменения происходят по закону связи причины и действия... Если, следовательно, мы ощущаем, что нечто происходит, то мы всегда предполагаем, что этому предшествует нечто другое, за чем оно следует в соответствии с неким правилом».

Кантовское понятие причинности в изложенной выше формулировке в настоящее время является предметом дискуссий ученых, работающих в области создания искусственного интеллекта. Не подвергая сомнению суть утверждения, они расходятся во мнениях по поводу его толкования: означает ли оно, что (1) всякий раз, когда происходит преемственность, существует правило, которое, по мнению агента, гарантирует преемственность? Или же оно означает, что (2) всякий раз агент верит, что существует некое правило, которое определяет правопреемство, даже если он не знает, что это за правило [1; 4; 5].

Сторонники второй, более слабой интерпретации (со слов сторонников первой, сильной), считают, что для восприятия преемственности вовсе не обязательно сформировать каузальное суждение: достаточно признать, что необходимо его сформировать. Например, для Б. Лонгюесс восприятие последовательности означает готовность искать причинное правило, это не значит, что он должен найти его.

Утверждение, что «все, что происходит, предполагает нечто другое, из чего оно следует по какому-то правилу» не означает, что мы познаем это правило, но что мы так устроены, что ищем его, ибо только его предпосылка позволяет нам распознать постоянное, которому мы приписываем изменяющиеся свойства [30].

Другие ученые, в том числе М. Фридман, придерживаются первой, более сильной интерпретации [19]. Приведем мнение Эванса, которое стало ключевым для создания Apperception Engine: «Если мы примем первую, более сильную интерпретацию, то любая реализация теории Канта будет представлять собой систему, которая может быть использована для предсказания будущих состояний, ретроспекции прошлых состояний и вменения недостающих данных. Эта способность заполнять пробелы в сенсорном потоке доступна только потому, что возможно фактически конструировать правила для объяснения последовательности явлений. Если бы мы реализовали вторую, более слабую интерпретацию, то можно было бы просто поверить в существование некоего правила – не было бы необходимости искать это правило, было бы достаточно знать, что оно где-то существует. В этом случае невозможно предвидеть будущее или реконструировать прошлое» [15].

Достаточно дискуссионным является и ответ на второй вопрос Эванса «по Канту». Для удобства читателя повторим вопрос: «Когда Кант говорит, что суждения – это правила, имеет ли он в виду (1) эксплицитные правила, сформированные из дискретных символов? Или он может иметь в виду, что некоторые суждения – это просто (2) имплицитные правила (например, процедура, которая подразумевается в весах нейронной сети)?»

Первая интерпретация, предполагающая, что суждения являются явными правилами, использующими дискретные символы в языке мышления, является формой того, что Брэндом называет регуляризмом [6].

Вторая интерпретация допускает правила, которые являются универсальными (они применимы ко всем объектам определенного типа), необходимыми (они применимы во всех ситуациях), но неявными, то есть имплицитными: правило может быть не выражено в кратком предложении на естественном или формальном языке. В качестве конкретного примера второй интерпретации можно привести нейрологическую машину [12]. Это нейронная сеть, которая имитирует прямое построение цепочки деепричастных оборотов, но не представляя клаузулы в явном виде. «Правила» нейронной логической машины неявно

выражены в весах (большой тензор значений FLOPS²) нейронной сети и не могут быть преобразованы в краткие человекочитаемые правила. Тем не менее эти правила универсальны и необходимы, они применимы ко всем объектам во всех ситуациях.

Большинство авторов считают, что правила Канта – это явные правила, состоящие из дискретных символов [5; 7; 20]. Эванс находит практическую причину для того, чтобы предпочесть первую интерпретацию в терминах эксплицитных правил. Отчасти привлекательность механизма апперцепции созданной им системы заключается в том, что теории, найденные ею, могут быть прочитаны, поняты и верифицированы. Например, теория, полученная Apperception Engine из трассы Сокобана (компьютерная игра с сильным ветвлением дерева решений, которая сопоставима с шахматами и большой глубиной, созданная Хиroyоки Имбаяси в 1981 году), не просто верна, а доказательно верна. Если необходимо понять, как и о чем думает машина, следует предпочесть эксплицитные правила (объяснимый искусственный интеллект, ХАИ, о котором было написано ранее)». Другая, возможно более фундаментальная, причина предпочтения явных правил состоит в том, что они позволяют нам проверить, удовлетворяются ли условия единства Канта. Например, для того чтобы проверить каждая ли последовательность подкрепляется каузальным суждением, необходимо иметь возможность проверить созданные правила. Непонятно, как система, оперирующая имплицитными правилами, может определить, были ли условия единства Канта действительно удовлетворены.

При создании Apperception Engine в качестве архитектурного проекта для системы машинного обучения была использована априорная психология Канта. Apperception Engine обеспечивает унифицированную реализацию различных способностей, описанных Кантом: воображение реализовано как набор недетерминированных правил выбора, способность суждения реализована в виде нейронной сети, а способность синтеза – в виде неконтролируемой системы синтеза программ. Эти подсистемы в высшей степени недетерминированы: воображение свободно синтезировать интуицию любым способом, способность суждения – сопоставлять интуицию с концептами любым удобным способом, а способность суждения вольна строить любые правила. Это справедливо до тех пор, пока совокупный продукт этих трех способностей удовлетворяет различным условиям единства [16; 17; 40].

² FLOPS (floating-point operations per second) – внесистемная единица, используемая для измерения производительности компьютеров, показывающая, сколько операций с плавающей запятой в секунду выполняет данная вычислительная система.

Следует привести мнения исследователей, работающих в области создания ИИ о том, что конкретные проектные решения, принятые в Apperception Engine Эванса, представляют собой лишь один из способов ответа на поставленные выше вопросы «по Канту». Существует множество других возможных архитектур. Одним из вариантов, например, представить правила в неявном виде [12] и использовать одну нейронную сеть для совместного обучения сопоставления интуиции с концепциями и изучения весов неявных правил. Другим вариантом может быть использование гибридной архитектуры, в которой нейронная сеть, обученная методом градиентного спуска, сопоставляет интуиции в понятия, а другая символическая система объединяет понятия в правила. Суть в том, что существует много способов реализовать когнитивную архитектуру Канта и работа в этом направлении продолжается [7; 8; 13; 21; 22; 23; 37].

Сочетание когнитологии Канта, трансцендентальной логики Канта и априорной психологии Канта современные исследователи, работающие в области создания искусственного интеллекта, называют принципами Канта. Без соблюдения указанных принципов создание сильного антропоморфного ИИ практически невозможно [25; 36; 38; 39; 40].

Что же дальше? Долго ли нам ждать появления сильного искусственного интеллекта, для возможности создания которого так много сделал Кант? Возможно ли это в принципе?

Человечество с момента возникновения homo sapiens обращено в неизведанный мир с нулевым запасом научных знаний. Практически *tabula rasa*. Полная аналогия с только что родившимся ребенком. Естественно, ребенок растет, многому учится и в 12–13 лет способен решать гораздо более сложную задачу, чем в 8. В каком же возрасте находится человечество, неспособное (пока?) решить задачу создания сильного ИИ? Следует сказать еще об одном варианте. Неспособный решить задачу ребенок может быть просто глуп (по Канту: «отсутствие способности суждения есть, собственно, то, что называют глупостью, и против этого недостатка нет лекарства») [1, с. 178]. Однако думать такое о человечестве, частью которого является великий Кант, с нашей точки зрения, неправильно.

Литература

1. *Кант И.* Критика чистого разума. М.: АСТ, 2022. С. 110.
2. *Катречко С.Л.* Кантовская концепция сознания (познания) как модель «искусственного интеллекта» // Новое в искусственном интеллекте. Методологические и теоретические вопросы / Под ред. Д.И. Дубровского, В.А. Лекторского. М.: ИИнтелЛЛ, 2005. С. 137–140.

3. *Allais L.* Kant, non-conceptual content and the representation of space. *Journal of the History of Philosophy*, 47(3):383–413, 2009.
4. *Achourioti T., van Lambalgen M.* A formalization of Kant's transcendental logic. *The Review of Symbolic Logic*, 4(2):254–289, 2011.
5. *Achourioti T., van Lambalgen M.* Kant's logic revisited. *IfCoLog Journal of Logics and Their Applications*, 4:845–865, 2017.
6. *Brandom R.* *Making It Explicit*. Harvard University Press, 1994.
7. *Brandom R.* *Between Saying and Doing*. Oxford University Press, 2008.
8. *Brandom R.* How analytic philosophy has failed cognitive science. *Towards an Analytic Pragmatism (TAP)*, 2009. P. 121–133.
9. *Clark A.* Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, 36(3):181–204, 2013.
10. *Cropper A., Muggleton S.* Learning efficient logic programs. *Machine Learning*, 2018. P. 1–21.
11. *Dennett D.* Artificial intelligence as philosophy and as psychology. *Brainstorms*, 1978. P. 109–26.
12. *Dong H., Jiayuan M.* Neural logic machines. arXiv preprint arXiv:1904.11694, 2019.
13. *Dyckhoff R., Negri S.* Geometrisation of first-order logic. *Bulletin of Symbolic Logic*, 21(2):123–163, 2015.
14. Document 32016R0679 Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation)
15. *Evans R., Sergot M.* Formalizing Kant's rules. *Journal of Philosophical Logic*, pages 1–68, 2019.
16. *Evans R.* Kant on constituted mental activity. *APA on Philosophy and Computers*, 2017.
17. *Evans R.* A Kantian cognitive architecture. In *On the Cognitive, Ethical, and Scientific Dimensions of Artificial Intelligence*, pages 233–262. Springer, 2019.
18. *Fodor J.* *The Language of Thought*. Harvard University Press, 1975.
19. *Friedman M.* *Kant and the Exact Sciences*. Harvard University Press, 1992.
20. *Gomes A.* Kant on perception: Naive realism, non-conceptualism, and the B-deduction. *The Philosophical Quarterly*, 64(254):1–19, 2013.
21. *Gerstenberg T., Tenenbaum J.* Intuitive theories. *Oxford Handbook of Causal Reasoning*, pages 515–548, 2017.
22. *Hamrick J.* Analogues of mental simulation and imagination in deep learning. *Current Opinion in Behavioral Sciences*, 29:8–16, 2019.
23. *Hernandez-Orallo J.* Beyond the Turing test. *Journal of Logic, Language and Information*, 9(4):447–466, 2000.
24. *Hernandez-Orallo J., Martinez-Plumed F.* Computer models solving intelligence test problems: Progress and implications. *Artificial Intelligence*, 230:74–107, 2016.
25. *James W., Burkhardt F.* *The Principles of Psychology*, volume 1. Macmillan London, 1890.
26. *Kant I.* What is enlightenment? In *Practical Philosophy*, pages 11–22. Cambridge University Press, 1784.
27. *Kant I.* *Critique of the Power of Judgment*. Cambridge University Press, 1790.
28. *Kant I.* The metaphysics of morals. In *Practical Philosophy*, pages 353–604. Cambridge University Press, 1797.
29. *Kim H.* 4 Tracing the Origins of Artificial Intelligence: A Kantian Response to McCarthy's Call for Philosophical Help // *Kant and Artificial Intelligence*. 2022. P. 129–143.
30. *Longuenesse B.* *Kant and the Capacity to Judge*. Princeton University Press, 1998.

31. *Lake B., Ullman T.* Building machines that learn and think like people. *Behavioral and Brain Sciences*, 40, 2017.
32. *McCarthy, J.* "Programs with Common Sense" at the Wayback Machine (archived October 4, 2013). In *Proceedings of the Teddington Conference on the Mechanisation of Thought Processes*, 756–91. London: Her Majesty's Stationery Office, 1959.
33. *Mihaly H.* Explainable AI: A Brief History of the Concept. *ERCIM News* (134): 9–10, 2023.
34. *Pinosio R.* The Logic of Kant's Temporal Continuum. PhD thesis, University of Amsterdam, 2017.
35. *Schönecker D.* Kant and Artificial Intelligence Published by De Gruyter, 2022.
36. *Sellars W.* The role of imagination in Kant's theory of experience. In *In the Space of Reasons*, pages 454–466. Harvard University Press, 1978.
37. *Stephenson A.* Kant's Theory of Experience. PhD thesis, University of Oxford, 2013.
38. *Stephenson A.* Kant on the object-dependence of intuition and hallucination. *The Philosophical Quarterly*, 65(260):486–508, 2015.
39. *Stephenson A.* Imagination and inner intuition. *Kant and the Philosophy of Mind*, 2017.
40. *Swanson L.* The predictive processing paradigm has roots in Kant. *Frontiers in Systems Neuroscience*, 10:79, 2016.
41. *Tian W., Zhou J.* Single-cell DNA methylation and 3D genome architecture in the human brain. *Science*. 2023.
42. *Waxman W.* Kant's Anatomy of the Intelligent Mind. Oxford University Press, 2014.
43. *Wolff R.* Kant's Theory of Mental Activity. Harvard University Press, 1963.

References

1. *Kant, I.* (2022). *Kritika chistogo razuma [The Critique of Pure Reason]*. M.: AST, P. 110. (In Russ.)
2. *Katrchko, S.L.* (2005). Kantovskaya koncepciya soznaniya (poznaniya) kak model «iskusstvennogo intellekta» [Kant's concept of consciousness (cognition) as a model of "artificial intelligence"] // *Novoe v iskusstvennom intellekte. Metodologicheskie i teoreticheskie voprosy [New in Artificial Intelligence. Methodological and theoretical issues]* M.: IntelLL. P. 137–140. (In Russ.)
3. *Allais, L.* (2009). Kant, non-conceptual content and the representation of space. *Journal of the History of Philosophy*, 47(3):383–413.
4. *Achourioti, T., van Lambalgen, M.* (2011). A formalization of Kant's transcendental logic. *The Review of Symbolic Logic*, 4(2):254–289.
5. *Achourioti, T., van Lambalgen, M.* (2017). Kant's logic revisited. *IfCoLog Journal of Logics and Their Applications*, 4:845–865.
6. *Brandom, R.* (1994). *Making It Explicit*. Harvard University Press.
7. *Brandom, R.* (2008). *Between Saying and Doing*. Oxford University Press.
8. *Brandom, R.* (2009). How analytic philosophy has failed cognitive science. *Towards an Analytic Pragmatism (TAP)*. P. 121–133.
9. *Clark, A.* (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, 36(3):181–204.
10. *Cropper, A., Muggleton, S.* (2018). Learning efficient logic programs. *Machine Learning*. P. 1–21.
11. *Dennett, D.* (1978). Artificial intelligence as philosophy and as psychology. *Brainstorms*. P. 109–26.
12. *Dong, H., Jiayuan, M.* (2019). Neural logic machines. arXiv preprint arXiv:1904.11694.

13. *Dyckhoff, R., Negri, S.* (2015). Geometrisation of first-order logic. *Bulletin of Symbolic Logic*, 21(2):123–163.
14. Document 32016R0679 Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation).
15. *Evans, R., Sergot M.* (2019). Formalizing Kant's rules. *Journal of Philosophical Logic*, pages 1–68.
16. *Evans, R.* (2017). Kant on constituted mental activity. *APA on Philosophy and Computers*.
17. *Evans, R.* (2019). A Kantian cognitive architecture. In *On the Cognitive, Ethical, and Scientific Dimensions of Artificial Intelligence*, pages 233–262. Springer.
18. *Fodor, J.* (1975). *The Language of Thought*. Harvard University Press.
19. *Friedman, M.* (1992). *Kant and the Exact Sciences*. Harvard University Press.
20. *Gomes, A.* (2013). Kant on perception: Naive realism, non-conceptualism, and the B-deduction. *The Philosophical Quarterly*, 64(254):1–19.
21. *Gerstenberg, T., Tenenbaum, J.* (2017) Intuitive theories. *Oxford Handbook of Causal Reasoning*, pages 515–548.
22. *Hamrick, J.* (2019). Analogues of mental simulation and imagination in deep learning. *Current Opinion in Behavioral Sciences*, 29:8–16.
23. *Hernandez-Orallo, J.* (2000). Beyond the Turing test. *Journal of Logic, Language and Information*, 9(4):447–466.
24. *Hernandez-Orallo, J., Martinez-Plumed, F.* (2016). Computer models solving intelligence test problems: Progress and implications. *Artificial Intelligence*, 230:74–107.
25. *James, W., Burkhardt, F.* (1890). *The Principles of Psychology*, volume 1. Macmillan London.
26. *Kant, I.* (1784). What is enlightenment? In *Practical Philosophy*, pages 11–22. Cambridge University Press.
27. *Kant, I.* (1790). *Critique of the Power of Judgment*. Cambridge University Press.
28. *Kant, I.* (1797). The metaphysics of morals. In *Practical Philosophy*, pages 353–604. Cambridge University Press.
29. *Kim, H.* (2022). 4 Tracing the Origins of Artificial Intelligence: A Kantian Response to McCarthy's Call for Philosophical Help // *Kant and Artificial Intelligence*. C. 129–143.
30. *Longuenesse, B.* (1998). *Kant and the Capacity to Judge*. Princeton University Press.
31. *Lake, B., Ullman, T.* (2017). Building machines that learn and think like people. *Behavioral and Brain Sciences*, 40.
32. *McCarthy, J.* (1959). "Programs with Common Sense" at the Wayback Machine (archived October 4, 2013). In *Proceedings of the Teddington Conference on the Mechanisation of Thought Processes*, 756–91. London: Her Majesty's Stationery Office.
33. *Mihaly, H.* (2023). Explainable AI: A Brief History of the Concept. *ERCIM News* (134): 9–10.
34. *Pinosio, R.* (2017). *The Logic of Kant's Temporal Continuum*. PhD thesis, University of Amsterdam.
35. *Schönecker, D.* (2022). *Kant and Artificial Intelligence* Published by De Gruyter.
36. *Sellars, W.* (1978). The role of imagination in Kant's theory of experience. In *In the Space of Reasons*, pages 454–466. Harvard University Press.
37. *Stephenson, A.* (2013). *Kant's Theory of Experience*. PhD thesis, University of Oxford.

38. *Stephenson, A.* (2015). Kant on the object-dependence of intuition and hallucination. *The Philosophical Quarterly*, 65(260):486–508.

39. *Stephenson, A.* (2017). Imagination and inner intuition. *Kant and the Philosophy of Mind*.

40. *Swanson, L.* (2016). The predictive processing paradigm has roots in Kant. *Frontiers in Systems Neuroscience*, 10:79.

41. *Tian, W., Zhou, J.* (2023). Single-cell DNA methylation and 3D genome architecture in the human brain. *Science*.

42. *Waxman, W.* (2014). *Kant's Anatomy of the Intelligent Mind*. Oxford University Press.

43. *Wolff, R.* (1963). *Kant's Theory of Mental Activity*. Harvard University Press.

Сведения об авторе

Кожокару Наталья Игоревна – Государственный академический университет гуманитарных наук (119049, Москва, ул. Мароновский переулок, 26).

natashakozhokaru@gmail.com

Information about the author

Kozhokaru, Natalia Igorevna – State Academic University for the Humanities (26, Maronovskii pereulok, Moscow, 119049, Russia).

Дата поступления: 30.07.2024