



УДК 101
<https://doi.org/10.17072/2078-7898/2026-3-382-393>
EDN: CCDOJA

Поступила: 31.07.2026
Принята: 24.08.2026
Опубликована: 01.10.2026

СМЫСЛ БЕЗ СУБЪЕКТА: ТРАНСЦЕНДЕНТАЛЬНЫЕ СТРУКТУРЫ ПОЗНАНИЯ В ФИЛОСОФИИ И. КАНТА И АРХИТЕКТУРА НЕЙРОСЕТЕЙ

Торощин Егор Александрович

Пермский государственный национальный исследовательский университет (Пермь)

В статье осуществляется систематическое сопоставление гносеологической архитектуры трансцендентальной философии И. Канта с принципами построения и функционирования современных нейронных сетей. В отличие от существующих работ, затрагивающих эту параллель фрагментарно, статья проводит последовательное, поэлементное сопоставление всех четырех уровней кантовской теории познания — априорных форм чувственности, категорий рассудка, схематизма и учения о трех синтезах (схватывания, воспроизведения и узнавания) — с конкретными архитектурными и алгоритмическими решениями нейросетей: входными репрезентациями данных, операциями агрегации и функциями активации, механизмами регуляризации, обеспечивающими обобщение, а также этапами обработки информации от низкоуровневых признаков до итоговой классификации. Применяется метод структурной, а не онтологической аналогии, что позволяет избежать метафорического отождествления мышления и вычисления. Показано, что обнаруживаемый структурный изоморфизм не случаен: он отражает фундаментальное свойство любой познавательной системы, требующей активного конструирования модели мира на основе поступающих данных. Вместе с тем обосновывается принципиальный предел данной аналогии: трансцендентальное единство апперцепции — рефлексивное самосознание, делающее возможным подлинный смысл, — не имеет и по существу не может иметь аналога в распределенных архитектурах современного ИИ. Этот вывод соотносится с аргументом Дж. Сёрла о «Китайской комнате» и с дискуссией о предиктивной обработке в философии сознания. Отсюда следует центральный тезис статьи: нейросеть воспроизводит механизм смыслопорождения, но не является носителем смысла. Выводы значимы для дискуссий о перспективах сильного ИИ (AGI) и его этических измерениях.

Ключевые слова: трансцендентальная философия, И. Кант, искусственный интеллект, нейронные сети, гносеология, сильный ИИ.

Для цитирования:

Торощин Е.А. Смысл без субъекта: трансцендентальные структуры познания в философии И. Канта и архитектура нейросетей // Вестник Пермского университета. Философия. Психология. Социология. 2026. Вып. 3. С. 382–393.
<https://doi.org/10.17072/2078-7898/2026-3-382-393>. EDN: CCDOJA

MEANING WITHOUT A SUBJECT: TRANSCENDENTAL STRUCTURES OF COGNITION IN I. KANT'S PHILOSOPHY AND THE ARCHITECTURE OF NEURAL NETWORKS

Yegor A. Toroshchin

Perm State University (Perm)

The article undertakes a systematic comparison between the epistemological architecture of Kant's transcendental philosophy and the principles of design and operation of contemporary neural networks. Unlike existing studies, which address this parallel only fragmentarily, the article offers a comprehensive, element-by-element comparison of all four levels of Kant's theory of cognition — the a priori forms of sensibility, the categories of understanding, schematism, and the doctrine of the three syntheses (apprehension, reproduction, and recognition) — with specific architectural and algorithmic solutions found in neural networks: input data representations, aggregation operations and activation functions, regularization mechanisms that ensure generalization, and the stages of information processing from low-level features to final classification. The method applied is one of structural rather than ontological analogy, which avoids a metaphorical identification of thinking with computation. It is shown that the structural isomorphism thus revealed is not accidental: it reflects a fundamental property of any cognitive system, namely the need to actively construct a model of the world on the basis of incoming data. At the same time, the article establishes a principled limit to this analogy: the transcendental unity of apperception — the reflexive self-consciousness that makes genuine meaning possible — has, and in principle can have, no analogue in the distributed architectures of contemporary AI. This conclusion is related to J. Searle's Chinese Room argument and to the debate over predictive processing in the philosophy of mind. From this follows the central thesis of the article: a neural network reproduces the mechanism of meaning-generation but is not a bearer of meaning. The conclusions bear on discussions on the prospects for artificial general intelligence (AGI) and its ethical dimensions.

Keywords: transcendental philosophy; Immanuel Kant; artificial intelligence; neural networks; epistemology; AGI.

To cite:

Toroshchin Ye.A. [Meaning without a subject: transcendental structures of cognition in I. Kant's philosophy and the architecture of neural networks]. *Vestnik Permskogo universiteta. Filosofia. Psihologia. Sociologia* [Perm University Herald. Philosophy. Psychology. Sociology], 2026, issue 3, pp. 382–393 (in Russian), <https://doi.org/10.17072/2078-7898/2026-3-382-393>. EDN: CCDOJA

Введение

Сегодня можно говорить о настоящем ренессансе кантовских идей в области исследований искусственного интеллекта (ИИ). Этот ренессанс обусловлен не внешними обстоятельствами, не модой или стремлением гуманитарных дисциплин к легитимизации через связь с актуальными технологиями, — он определяется внутренней логикой развития самой науки об ИИ. По мере того как нейросетевые системы достигают

все более высоких результатов в задачах восприятия, классификации и генерации [LeCun Y. et al., 2015], исследователи с нарастающей остротой ставят вопрос: что именно происходит внутри этих систем? Какова природа тех «знаний», которые в них кодируются? Эти вопросы неизбежно выводят за пределы инженерии — в область философии познания.

Трансцендентальная философия Канта предлагает один из наиболее разработанных ответов

на вопрос о том, как вообще возможно познание. Именно поэтому ряд исследователей все настойчивее указывает на то, что положения кантовской гносеологии описывают механизм, принципиально сходный с тем, что реализуется в нейросетевых архитектурах [Козолупенко Д.П., 2023, с. 120]. Более того, структура разума в философии Канта начинает рассматриваться как своего рода «архетип искусственного интеллекта» [Григорова Я.В., Комаров С.В., 2024, с. 464], а категории рассудка понимаются как «логические драйверы синтеза продуктивного воображения», формальный механизм которых принципиально отделен от работы живого сознания и может быть воспроизведен технически [Беттони М., 1995, с. 133].

Вместе с тем в существующей литературе эта параллель обычно проводится лишь по отдельным направлениям: исследователи вспоминают отдельные элементы трансцендентальной философии Канта при анализе нейросетей и ИИ, но не сводят эти элементы в единую систему. В настоящей статье предлагается восполнить этот пробел, последовательно сопоставив все четыре уровня кантовской теории познания с конкретными архитектурными и алгоритмическими решениями, используемыми в современных нейронных сетях.

Исследование направлено на то, чтобы выявить структурное сходство между гносеологической архитектурой трансцендентальной философии Канта и принципами организации нейросетей и вместе с тем определить границу такого сходства, связанную с вопросами смысла и самосознания.

В качестве основного метода используется структурная аналогия, при которой сопоставляются функциональные характеристики двух систем — трансцендентальной гносеологии и нейросетевой архитектуры, а не их онтологический статус. Поэтому речь не идет о том, что нейросеть буквально «мыслит по Канту» или непосредственно воплощает кантовские понятия. Сопоставление строится иначе: формальные структуры, которые Кант рассматривал как условия возможности познания, сравниваются с

архитектурными решениями, к которым независимо пришли разработчики нейросетей.

Состояние проблемы в современной литературе

Интерес к связи между философией Канта и исследованиями ИИ имеет относительно давнюю историю. Одним из первых ее обозначил М. Беттони, предложивший использовать кантовскую архитектуру рассудка как основу для проектирования человекоориентированных программных систем [Беттони М., 1995]. В отечественной традиции В.Н. Брюшинкин использует кантовские категории трансцендентальной логики как потенциальную аналогию для построения модели мира в системе ИИ [Брюшинкин В.Н., 1990, 1991]. Однако в тот период данные идеи носили скорее метафорический характер и не имели систематического развития.

Новый импульс эта дискуссия получила в связи с успехами глубокого обучения в 2010-х гг. Так, Д.П. Козолупенко рассматривает трансцендентальную философию Канта как теоретическое основание для исследований в области ИИ, акцентируя внимание прежде всего на понятии трансцендентального субъекта и его отличии от любой вычислительной системы [Козолупенко Д.П., 2023]. Я.В. Григорова и С.В. Комаров анализируют проблему ИИ через призму кантовской философии в целом, вводя понятие «архетипа искусственного интеллекта» применительно к кантовской структуре разума [Григорова Я.В., Комаров С.В., 2024]. Н.И. Кожокару отмечает, что создание сильного антропоморфного ИИ невозможно без учета философских принципов Канта [Кожокару Н.И., 2024, с. 58]. А.Г. Пушкарский специально рассматривает значение философии сознания Канта для современных исследований по ИИ, приходя к выводу о принципиальной несводимости кантовского трансцендентального субъекта к любой вычислительной модели [Пушкарский А.Г., 2025].

В англоязычной традиции значимым вкладом стала работа Ю. Хуэя, в которой исследу-

ется напряжение между кантовской критической философией и современными системами ИИ в этическом и политическом измерениях [Hui Y., 2026]. Также попытки реинтерпретировать кантовскую гносеологию как вычислительную архитектуру предпринимались независимо от отечественной дискуссии. Р. Эванс прямо истолковывает «Трансцендентальную аналитику» как описание когнитивной архитектуры и реализует ее в виде работающей программы для задач индуктивного обучения [Evans R., 2019]. К. Бакнер показывает, что операция трансформационной абстракции в сверточных сетях структурно воспроизводит функцию эмпирической генерализации, которую кантианская традиция приписывала синтезирующей деятельности рассудка, и, обобщая эту линию анализа, прослеживает связь между историей эмпиризма и современным машинным обучением [Buckner C., 2018, 2024]. П. Грабен предлагает нейросетевую (GAN-based) реконструкцию кантовского учения о гармонии познавательных способностей, распространяя структурную аналогию на область эстетического суждения [Beim Graben P., 2024]. Эти работы подтверждают продуктивность кантовской оптики для анализа ИИ, но, как и рассмотренные выше отечественные исследования, затрагивают лишь отдельные звенья гносеологической архитектуры, не давая целостной картины.

В имеющейся литературе практически отсутствуют попытки провести систематическое, уровень за уровнем, сопоставление кантовской гносеологии с конкретными архитектурными решениями нейронных сетей — от входных слоев и функций активации до механизмов самовнимания и выходной классификации. Именно этот пробел настоящая статья стремится заполнить.

Априорные формы чувственности и входная архитектура нейросетей

Начнем с самого основания познавательного процесса — с того, как вообще возможна встреча познающего субъекта с миром. В «Трансцендентальной эстетике» Кант обосновывает тезис о том, что пространство и время

являются не свойствами вещей самих по себе, но априорными формами нашей чувственности — необходимыми условиями всякого созерцания. «Пространство есть не что иное, как только форма всех явлений внешних чувств, то есть субъективное условие чувственности, при котором единственно и возможны для нас внешние созерцания» [Кант И., 1964, с. 50]. Время выступает как форма внутреннего чувства, упорядочивающая последовательность наших состояний. Принципиально важно, что эти формы не выводятся из опыта, но, напротив, делают сам опыт возможным: они предшествуют любому восприятию и являются его условием.

В архитектуре нейронных сетей мы обнаруживаем структурно сходный механизм. Любая сеть работает не с «сырой» реальностью, а с данными, уже организованными определенным образом. Изображения преобразуются в тензоры пиксельных значений с сохранением пространственной структуры; тексты — в эмбединги или последовательности токенов; временные ряды — в выборки с заданным шагом дискретизации. Это преобразование происходит до начала какой-либо содержательной обработки данных. Входные слои и способы репрезентации данных выполняют функцию, строго аналогичную кантовским формам чувственности: они задают те априорные условия, при которых данные вообще могут быть восприняты и обработаны сетью.

Параллель здесь не поверхностная. Подобно тому как кантовский субъект не может созерцать вещи вне пространства и времени, нейросеть не может обрабатывать данные, не организованные в соответствии с ее входной архитектурой. И эта архитектура задается до обучения и не изменяется в его процессе — она действительно *априорна* по отношению к любому опыту данной конкретной сети. Сверточная нейросеть (CNN) [LeCun Y. et al., 1998] «видит» мир как структурированное пространство локальных признаков; рекуррентная сеть (RNN) — как упорядоченную во времени последовательность; трансформер — как множество элементов, каждый из которых соотно-

сен со всеми остальными. Тип архитектуры определяет, какой «мир» вообще доступен сети — точно так же, как кантовские формы чувственности определяют, каким образом мир может явиться субъекту.

Следует, однако, зафиксировать и существенное отличие. Кантовские формы чувственности — пространство и время — универсальны: они одни и те же для всех познающих субъектов и не варьируются от случая к случаю. Входная архитектура нейросети, напротив, определяется разработчиком и варьируется в зависимости от задачи. Это означает, что «априоризм» нейросети носит инструментальный, а не трансцендентальный характер: он задан извне, а не конститутивен для самой способности к познанию.

Категории рассудка и вычислительные операции

Чувственность дает лишь многообразие ощущений, которое само по себе еще не есть знание. Подлинное познание начинается там, где в дело вступает рассудок с его категориями — чистыми понятиями, которые вносят в опыт формы связи. «Одни лишь созерцания без понятий слепы» — этот кантовский тезис указывает на необходимость активной синтетической деятельности рассудка [Кант И., 1964, с. 106]. Кант выделяет четыре класса категорий: количества (единство, множество, цельность), качества (реальность, отрицание, ограничение), отношения (субстанция и акциденция, причина и действие, взаимодействие) и модальности (возможность, действительность, необходимость).

В нейросетях этим категориям соответствуют конкретные вычислительные операции, реализуемые в различных типах слоев: категории количества находят свой аналог в операциях агрегации — пулинге (max pooling, average pooling), суммировании, батч-нормализации [Ioffe S., Szegedy C., 2015]. Все эти операции выполняют одну и ту же функцию: преобразуют множество отдельных элементов (пикселей, значений активаций) в единое целое, сворачивая многообразие к единству. Именно это Кант описывал как функцию категорий количества. Категории качества реализуются через функции активации.

Функция ReLU (Rectified Linear Unit) [Nair V., Hinton G.E., 2010] осуществляет операцию, структурно соответствующую кантовскому «отрицанию»: она устанавливает порог, ниже которого значение полагается равным нулю — отсутствующим. Сигмоидальная функция ограничивает диапазон входных значений, переводя их в интервал от нуля до единицы, а пороговые функции тем самым позволяют различать наличие и отсутствие определенного признака. Далее особенно интересен переход к категориям отношения. Полносвязный слой можно рассматривать как функциональное соответствие связи между «субстанцией» — устойчивым узлом сети — и ее «акциденциями», т. е. входными сигналами, которые этот узел получает и учитывает при вычислении. В сверточных слоях, где фиксируются локальные зависимости между краями, текстурами и другими признаками, обнаруживается аналогия с причинностью: определенные признаки устойчиво следуют друг за другом в процессе обработки. Механизм самовнимания (self-attention) в трансформерах, впервые описанный А. Васвани с соавт. [Vaswani A. et al., 2017], дает другой вариант такой связи — взаимодействие элементов последовательности, где каждый элемент может соотноситься с остальными, а характер этих взаимных связей влияет на формирование итогового представления.

Разумеется, нейросеть не «понимает» причинность в философском смысле и не применяет категории сознательно. Но она вычислительно воспроизводит ту функцию установления связей и отношений, которую Кант приписывал категориям рассудка. Это и есть функциональный изоморфизм: не тождество природы, а тождество выполняемой роли в познавательном процессе.

Схематизм и процесс машинного обучения

Центральной и наиболее трудной проблемой «Критики чистого разума» является вопрос о том, каким образом чистые понятия рассудка, не содержащие ничего чувственного, могут применяться к данным чувственного созерцания. Ответом служит учение о схематизме. «Чистые априорные понятия кроме функции рассудка в

категориях должны а priori содержать еще формальные условия чувственности (именно внутреннего чувства), заключающие в себе общее условие, при котором единственно и можно применять категорию к какому-нибудь предмету» [Кант И., 1964, с. 177]. Таким условием выступает схема — не образ, а «представление о методе»: общий способ построения образа, применимый к бесконечному множеству конкретных случаев.

Например, схема треугольника — это не образ какого-либо конкретного треугольника, а правило его построения в пространстве, которое позволяет опознать треугольником любую конкретную фигуру с тремя углами вне зависимости от ее размера, цвета или ориентации. Схема носит процедурный характер: это знание того, *как* действовать, а не знание того, *что* есть. Здесь мы обнаруживаем параллель с механизмом работы машинного обучения. Так, обученная сеть не должна просто сохранять отдельные примеры из обучающей выборки: в таком случае речь шла бы скорее о запоминании, чем об обобщении. Результатом обучения становятся более общие способы преобразования входных данных в выходные, распределенные в матрицах весов. Алгоритм обратного распространения ошибки [Rumelhart D.E. et al., 1986] и применяемые вместе с ним средства регуляризации — L1/L2-регуляризация, дропаут [Srivastava N. et al., 2014], ранняя остановка — все они ограничивают возможность простого заучивания примеров в работе нейросети. Благодаря этому сеть должна выделять такие процедуры преобразования, которые сохраняют применимость и при работе с новыми данными. В этом отношении регуляризацию можно понимать как условие сохранения «схематичности» знания, поскольку она не позволяет свести его к совокупности отдельных, частных случаев.

Кантовская схема по своему характеру является процедурой, и в этом отношении знание, формируемое нейросетью, также имеет процедурный характер: оно выражается прежде всего в способах преобразования входных сигналов, а не в хранении самих эмпирических данных. Поэтому аналогия здесь затрагивает не только

устройство двух систем, но и сам способ понимания знания. В обоих случаях на первый план выходит его конструктивная сторона: знание связано не столько с готовым представлением объекта, сколько с определенным способом организации и преобразования материала.

Три синтеза и этапы обработки информации

В первом издании «Критики чистого разума» Кант разворачивает учение о трех синтезах, через которые многообразие ощущений превращается в единый предмет познания. Это учение описывает познание как динамический, поэтапный процесс, а не как мгновенный акт — и именно поэтому оно обнаруживает особенно богатый потенциал для сопоставления с архитектурой нейронных сетей.

Первый синтез — схватывание (апрегензия) в созерцании — есть первичное упорядочивание впечатлений в пространстве и времени. Его функция — априорное объединение многообразного, упорядочивание «по формальному условию внутреннего чувства, а именно времени» [Кант И., 1964, с. 701]. В нейросетях этому синтезу соответствует этап извлечения низкочастотных признаков. Ранние слои сверточных сетей выделяют границы, цветовые переходы, текстуры, т. е. производят первичное структурирование «сырого» ввода, не несущее еще никакого семантического содержания, но делающее дальнейшую обработку возможной.

Второй синтез — узнавание (рекогниция) в понятии. Он связывает уже синтезированное многообразие с понятием и тем самым позволяет удержать единство объекта. На этом этапе многообразие получает определенную идентичность: оно выступает уже не просто как совокупность данных, а как распознаваемый предмет. В нейросетях этому можно функционально сопоставить выходной этап классификации или генерации: нейросеть «соотносит» обработанный материал с определенным классом: например, изображением, тональностью текста или следующим токеном, либо формирует итоговый результат.

Третий синтез — воспроизведение в воображении — связан с удержанием уже полученных

представлений и их соединением с последующими. Без такого удержания восприятие распадалось бы на отдельные, не связанные между собой моменты. В нейросетевых системах функционально близкую задачу решают механизмы памяти и контекста. Так, рекуррентные сети, в частности LSTM [Hochreiter S., Schmidhuber J., 1997] и GRU [Cho K. et al., 2014], сохраняют «скрытое состояние», в котором учитывается информация о предыдущих шагах последовательности. В трансформерах эту возможность расширяет механизм самовнимания: каждый элемент последовательности может непосредственно учитывать другие элементы, устанавливая зависимости, которые не ограничиваются непосредственной временной близостью.

Обучение нейросети также имеет выраженную временную структуру, что позволяет провести дополнительную параллель с кантовским пониманием времени как априорной формы синтеза. Батчи, т. е. порции данных, одновременно поступающие на вход нейросети, можно сопоставить с синхронным схватыванием множества элементов. Эпоха, в свою очередь, представляет собой полный проход по обучающей выборке, после которого сеть вновь возвращается к тем же данным и корректирует веса с учетом результатов предыдущих итераций. Возникает повторяющийся цикл: обработка данных сменяется обновлением параметров, затем сеть снова обращается к материалу. В этом можно увидеть функциональное сходство с синтезом воспроизведения, поскольку устойчивая модель формируется за счет связывания текущей обработки с результатами уже выполненных шагов.

Граница аналогии: смысл без субъекта

Выявленное структурное сходство между философией И. Канта и работой нейросетей имеет философское значение. Однако важно обозначить и их различие. Так, в кантовской гносеологии есть одно принципиальное положение, которому в архитектуре нейросетей невозможно подобрать прямой аналог, — трансцендентальное единство апперцепции. Формула «Я мыслю» выражает требование, согласно которому все

многообразие представлений должно быть способно сопровождаться одним и тем же «Я мыслю», относящимся к одному и тому же субъекту [Кант И., 1964, с. 191]. Это единство не возникает как итог обработки информации, а выступает условием, при котором сама обработка может быть отнесена к единому субъекту. Поэтому апперцепция обеспечивает не возможность обработки эмпирических данных, а возможность осознать эту переработку как принадлежащую *мне*, т. е. субъекту.

Это единство не является результатом обработки информации — оно выступает ее трансцендентальным условием. Именно апперцепция делает возможным не просто переработку данных, но *осознание* этой переработки как *моей*, принадлежащей единому субъекту.

Распределенные архитектуры современных нейросетей не имеют ничего соответствующего этому единству. Нейросеть не «знает», что она обрабатывает данные; у нее нет точки зрения, с которой процесс обработки был бы виден как целое. Она не рефлексит над своими операциями. Это означает, что нейросеть воспроизводит *механизм* смыслопорождения, но не является *носителем смысла*. Смысл предполагает субъекта, для которого нечто значит. В отсутствие такого субъекта мы имеем дело с обработкой информации, но не с пониманием.

Данный вывод перекликается с аргументом Дж. Сёрла о «Китайской комнате» [Searle J.R., 1980]: система может оперировать символами по правилам, не понимая ни одного из них. Нейросеть, в отличие от сёрловской комнаты, не использует явных правил — она выучивает статистические паттерны. Но в точке, касающейся отсутствия понимания, аргумент сохраняет свою силу. Х. Дрейфус указывал, что человеческое понимание укоренено в телесном, ситуированном опыте, который принципиально недоступен для дискретных вычислительных систем [Dreyfus H.L., 1992]. Нейросети, в отличие от систем, критикуемых Дрейфусом, отчасти воспроизводят некоторые черты такой ситуированности — через дообучение на специфических данных, через мультимодальность. Однако про-

блема субъективности и рефлексивного самосознания остается открытой.

Смежная дискуссия разворачивается вокруг нейрокогнитивных моделей «прогнозирующего кодирования» (predictive coding), сторонники которых — А. Кларк — прямо апеллирует к кантовскому трансцендентальному идеализму, утверждая, что познающая система никогда не имеет дела с миром «как он есть», а лишь со своей собственной наилучшей гипотезой о нем [Clark A., 2016]. Д. Захави прослеживает эту связь между предиктивным кодированием, неокантианством и трансцендентальным идеализмом [Zahavi D., 2018], а Т. Шлихт, ранее уже рассматривавший развитие когнитивной науки через призму кантовского подхода к познанию [Schlicht T., 2022], в более поздней работе показывает, что само это кантианское допущение подрывает притязания прогнозирующего кодирования на реалистическое и натуралистическое объяснение сознания [Schlicht T., 2026]. Данная дискуссия, разворачивающаяся вокруг биологических нейросетей мозга, содержит, таким образом, тот же принципиальный урок, что и предпринятый в настоящей статье анализ искусственных архитектур: чем ближе структурная аналогия с кантовской гносеологией, тем острее встает вопрос о статусе субъекта, для которого обрабатываемое нечто значит.

Из этого вытекает важное следствие для обсуждения проблемы сильного ИИ (AGI). Даже если удастся решить инженерную задачу построения «алгоритмического» единства апперцепции, т. е. создать систему, которая отслеживает и интегрирует свои собственные состояния, останется вопрос о том, будет ли такая интеграция *субъективной* в кантовском смысле, или лишь ее функциональным имитатором. Юк Хуэй справедливо указывает, что сферы этического и политического — как необходимые компоненты «человеческого» интеллекта — не сводимы к алгоритмическому расчету [Hui Y., 2026, p. 98]. Это означает, что даже технически совершенный AGI останется системой иного рода, чем человеческое сознание: не потому, что он недостаточно сложен, а потому, что слож-

ность не является тем измерением, в котором располагается разрыв.

Заключение

Проведенный анализ позволяет сделать следующие выводы.

Между гносеологической архитектурой трансцендентальной философии Канта и архитектурой нейронных сетей обнаруживается глубокий структурный изоморфизм, охватывающий все четыре уровня кантовской теории познания: априорные формы чувственности находят соответствие во входной архитектуре и способах представления данных; категории рассудка — в вычислительных операциях различных типов слоев; схематизм — в процессах обучения и формирования обобщенных процедурных правил; три синтеза — в этапах обработки информации: от низкоуровневых признаков до итоговой классификации.

Сходство такого рода нельзя считать случайным. Оно связано с общей особенностью познавательных систем. Так, поступающий материал должен быть организован в некоторую *модель* мира посредством заранее заданных способов обработки — архитектурных в случае нейросетей и трансцендентальных — в кантовской философии. Кант дал этой особенности последовательное философское описание, поэтому его гносеологическая модель и сегодня может служить эвристическим инструментом для анализа нейросетевых систем.

Однако речь идет именно о функциональном, а не об онтологическом изоморфизме. Его граница проявляется в проблеме трансцендентального единства апперцепции. Нейросеть способна воспроизводить определенные механизмы смыслопорождения, но сама не становится носителем смысла: отсутствует субъект, для которого результат обработки обладал бы значением. Отсюда следует более общий вывод, важный не только для философии ИИ, но и для дискуссий о сильном ИИ (AGI) и его этических перспективах. Различие между обработкой информации и пониманием смысла носителем этой информации сохраняется как самостоятельная философская проблема и не устраняется простым увели-

чением и усложнением вычислительных ресурсов.

Список литературы

Беттони М. Кант и кризис программного обеспечения. Предложения по построению программных систем, ориентированных на человека // Кантовский сборник. 1995. Т. 1. № 19. С. 131–137. EDN: WBASGL

Брюшинкин В.Н. Кант и «искусственный интеллект»: модели мира // Кантовский сборник: Межвузовский тематический сборник научных трудов. 1990. № 1(15). С. 80–89. EDN: YUQZSD

Брюшинкин В.Н. Кант и искусственный интеллект: трансцендентальный анализ моделей мира // Кантовский сборник: Межвузовский тематический сборник научных трудов. 1991. № 1(16). С. 84–89. EDN: YUQZYS

Григорова Я.В., Комаров С.В. Анализ проблемы искусственного интеллекта через призму философии И. Канта // Вестник Пермского университета. Философия. Психология. Социология. 2024. № 4. С. 463–470. DOI: <https://doi.org/10.17072/2078-7898/2024-4-463-470> EDN: SECXZV

Кант И. Критика чистого разума // Кант И. Сочинения: в 6 т. М.: Мысль, 1964. Т. 3. 799 с.

Кожокару Н.И. Философские концепции И. Канта как предпосылки создания искусственного интеллекта // Философия науки. 2024. № 5S. С. 49–62. DOI: <https://doi.org/10.15372/PS20240505>

Козолупенко Д.П. Между сознанием и интеллектом: трансцендентальная философия И. Канта как теоретическое основание для исследований в области искусственного интеллекта // Трансцендентальный поворот в современной философии – 8: Метафизика, эпистемология, когнитивистика и искусственный интеллект: сб. тезисов. М.: РГГУ, 2023. С. 117–121. EDN: FRHPMK

Пушкарский А.Г. Значение философии сознания Канта для современных исследований по искусственному интеллекту // Вестник Российского университета дружбы народов. Серия: Философия. 2025. Т. 29, № 2. С. 473–490. DOI: <https://doi.org/10.22363/2313-2302-2025-29-2-473-490> EDN: THULXV

Beim Graben P. A neural network account to Kant's philosophical aesthetics. URL:

<https://arxiv.org/html/2404.12395> (дата обращения: 14.07.2026). DOI: <https://doi.org/10.53765/mm2024.227>

Buckner C. Empiricism without magic: transformational abstraction in deep convolutional neural networks // Synthese. 2018. Vol. 195, № 12. P. 5339–5372. DOI: <https://doi.org/10.1007/s11229-018-01949-1>

Buckner C. From Deep Learning to Rational Machines: What the History of Philosophy Can Teach Us about the Future of Artificial Intelligence. New York: Oxford University Press, 2024. 328 p. DOI: <https://doi.org/10.1093/oso/9780197653302.001.0001>

Cho K., van Merriënboer B., Gulcehre C. et al. Learning Phrase Representations using RNN Encoder–Decoder for Statistical Machine Translation // Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), October 25–29, Doha. Qatar: Doha, 2014. P. 1724–1734. DOI: <https://doi.org/10.3115/v1/d14-1179>

Clark A. Surfing Uncertainty: Prediction, Action, and the Embodied Mind. Oxford: Oxford University Press, 2016. 320 p. DOI: <https://doi.org/10.1093/acprof:oso/9780190217013.001.0001>

Dreyfus H.L. What Computers Still Can't Do: A Critique of Artificial Reason. Cambridge, MA: MIT Press, 1992. 354 p.

Evans R. A Kantian Cognitive Architecture // On the Cognitive, Ethical, and Scientific Dimensions of Artificial Intelligence. Philosophical Studies Series, vol. 134. Cham: Springer, 2019. P. 233–262. DOI: https://doi.org/10.1007/978-3-030-01800-9_13

Hochreiter S., Schmidhuber J. Long Short-Term Memory // Neural Computation. 1997. Vol. 9, № 8. P. 1735–1780. DOI: <https://doi.org/10.1162/neco.1997.9.8.1735>

Hui Y. Kant Machine: Critical Philosophy After AI. Political Theory and Contemporary Philosophy. London: Bloomsbury Academic, 2026. 142 p. DOI: <https://doi.org/10.5040/9781350563230>

Ioffe S., Szegedy C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift // Proceedings of the 32nd International Conference on Machine Learning. 2015. Vol. 37. P. 448–456.

LeCun Y., Bottou L., Bengio Y., Haffner P. Gradient-Based Learning Applied to Document Recognition // Proceedings of the IEEE. 1998. Vol.

86, № 11. P. 2278–2324. DOI: <https://doi.org/10.1109/5.726791>

LeCun Y., Bengio Y., Hinton G. Deep learning // *Nature*. 2015. Vol. 521. P. 436–444. DOI: <https://doi.org/10.1038/nature14539>

Nair V., Hinton G.E. Rectified Linear Units Improve Restricted Boltzmann Machines // *Proceedings of the 27th International Conference on Machine Learning (ICML-10)*. 2010. P. 807–814.

Rumelhart D.E., Hinton G.E., Williams R.J. Learning Representations by Back-Propagating Errors // *Nature*. 1986. Vol. 323, № 6088. P. 533–536. DOI: <https://doi.org/10.1038/323533a0>

Schlicht T. Minds, Brains, and Deep Learning: The Development of Cognitive Science Through the Lens of Kant's Approach to Cognition // *Kant and Artificial Intelligence*. Berlin; Boston: De Gruyter, 2022. P. 1–38. DOI: <https://doi.org/10.1515/9783110706611-001>

Schlicht T. Predictive processing's flirt with transcendental idealism // *Noûs*. 2026. Vol. 60, № 1. P. 87–109. DOI: <https://doi.org/10.1111/nous.12552>

Searle J.R. Minds, Brains, and Programs // *Behavioral and Brain Sciences*. 1980. Vol. 3, № 3. P. 417–424. DOI: <https://doi.org/10.1017/s0140525x00005756>

Srivastava N., Hinton G., Krizhevsky A., Sutskever I., Salakhutdinov R. Dropout: A Simple Way to Prevent Neural Networks from Overfitting // *Journal of Machine Learning Research*. 2014. Vol. 15, № 1. P. 1929–1958.

Vaswani A., Shazeer N., Parmar N. et al. Attention Is All You Need // *Advances in Neural Information Processing Systems*. 2017. Vol. 30. P. 5999–6008. DOI: <https://doi.org/10.65215/r5bs2d54>

Zahavi D. Brain, Mind, World: Predictive Coding, Neo-Kantianism, and Transcendental Idealism // *Husserl Studies*. 2018. Vol. 34, № 1. P. 47–61. DOI: <https://doi.org/10.1007/s10743-017-9218-z>

References

Beim Graben, P. (2024). *A neural network account to Kant's philosophical aesthetics*. Available at: <https://arxiv.org/html/2404.12395> (accessed 14.07.2026). DOI: <https://doi.org/10.53765/mm2024.227>

Bettoni, M. (1995). [Kant and the crisis of software engineering. Proposals for the design of

human-oriented software systems]. *Kantovskiy sbornik* [Kant Digest]. Vol. 1, no. 19, pp. 131–137.

Bryushinkin, V.N. (1990). [Kant and "artificial intelligence". Models of the world]. *Kantovskiy sbornik* [Kant Digest]. No. 1(15), pp. 80–89.

Bryushinkin, V.N. (1991). [Kant and artificial intelligence. A transcendental analysis of models of the world]. *Kantovskiy sbornik* [Kant Digest]. No. 1(16), pp. 84–89.

Buckner, C. (2018). Empiricism without magic. Transformational abstraction in deep convolutional neural networks. *Synthese*. Vol. 195, no. 12, pp. 5339–5372. DOI: <https://doi.org/10.1007/s11229-018-01949-1>

Buckner, C. (2024). *From deep learning to rational machines. What the history of philosophy can teach us about the future of artificial intelligence*. New York: Oxford University Press, 328 p. DOI: <https://doi.org/10.1093/oso/9780197653302.001.0001>

Cho, K., van Merriënboer, B., Gulcehre, C. et al. (2014). Learning phrase representations using RNN encoder-decoder for statistical machine translation. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Qatar: Doha, pp. 1724–1734. DOI: <https://doi.org/10.3115/v1/d14-1179>

Clark, A. (2016). *Surfing uncertainty. Prediction, action, and the embodied mind*. Oxford: Oxford University Press, 320 p. DOI: <https://doi.org/10.1093/acprof:oso/9780190217013.001.0001>

Dreyfus, H.L. (1992). *What computers still can't do. A critique of artificial reason*. Cambridge: MIT Press, 354 p.

Evans, R. (2019). A Kantian cognitive architecture. On the cognitive, ethical, and scientific dimensions of artificial intelligence. *Philosophical Studies Series*. Vol. 134, pp. 233–262. DOI: https://doi.org/10.1007/978-3-030-01800-9_13

Grigorova, Ya.V. and Komarov, S.V. (2024). [Analyzing the problem of artificial intelligence through the prism of Immanuel Kant's philosophy]. *Vestnik Permskogo universiteta. Filosofiya. Psikhologiya. Sotsiologiya* [Perm University Bulletin. Philosophy. Psychology. Sociology]. Iss. 4, pp. 463–470. DOI: <https://doi.org/10.17072/2078-7898/2024-4-463-470>

Hochreiter, S. and Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*.

Vol. 9, no. 8, pp. 1735–1780. DOI: <https://doi.org/10.1162/neco.1997.9.8.1735>

Hui, Y. (2026). Kant machine. Critical philosophy after AI. Political theory and contemporary philosophy. London: Bloomsbury Academic Publ., 142 p. DOI: <https://doi.org/10.5040/9781350563230>

Ioffe, S. and Szegedy, C. (2015). Batch normalization. *Accelerating Deep Network Training by Reducing Internal Covariate Shift. Proceedings of the 32nd International Conference on Machine Learning*. Vol. 37, pp. 448–456.

Kant, I. (1964). [Critique of pure reason]. *Sochineniya v 6 t.* [Works in 6 vols]. Moscow: Mysl' Publ., vol. 3, 799 p.

Kozhokaru, N.I. (2024). [Kant's philosophical concepts as prerequisites for the creation of artificial intelligence]. *Filosofiya nauki* [Philosophy of Science]. No. 5S, pp. 49–62. DOI: <https://doi.org/10.15372/PS20240505>

Kozolupenko, D.P. (2023). [Between consciousness and intelligence. I. Kant's transcendental philosophy as a theoretical foundation for research in artificial intelligence]. *Transtsendental'nyy povорот v sovremennoy filosofii – 8: Metafizika, epistemologiya, kognitivistika i iskusstvennyy intellekt* [The Transcendental Turn in Contemporary Philosophy – Metaphysics, Epistemology, Cognitive Science and Artificial Intelligence]. Moscow: RSUH Publ., pp. 117–121.

LeCun, Y., Bengio, Y. and Hinton, G. (2015). Deep learning. *Nature*. Vol. 521, pp. 436–444. DOI: <https://doi.org/10.1038/nature14539>

LeCun, Y., Bottou, L., Bengio, Y. and Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*. Vol. 86, no. 11, pp. 2278–2324. DOI: <https://doi.org/10.1109/5.726791>

Nair, V. and Hinton, G.E. (2010). Rectified linear units improve restricted Boltzmann machines. *Proceedings of the 27th International Conference on Machine Learning (ICML-10)*. Madison: Omnipress Publ., pp. 807–814.

Pushkarskiy, A.G. (2025). [The importance of Kant's philosophy of mind for contemporary research in artificial intelligence]. *Vestnik Rossiyskogo universiteta druzhby narodov. Seriya Filosofiya* [RUDN Journal of Philosophy]. Vol. 29, no. 2, pp. 473–490. DOI: <https://doi.org/10.22363/2313-2302-2025-29-2-473-490>

Rumelhart, D.E., Hinton, G.E. and Williams, R.J. (1986). Learning representations by back-propagating errors. *Nature*. Vol. 323, no. 6088, pp. 533–536. DOI: <https://doi.org/10.1038/323533a0>

Schlicht, T. (2022). Minds, brains, and deep learning. The development of cognitive science through the lens of Kant's approach to cognition. *Kant and Artificial Intelligence*. Berlin, Boston: De Gruyter Publ., pp. 1–38. DOI: <https://doi.org/10.1515/9783110706611-001>

Schlicht, T. (2026). Predictive processing's flirt with transcendental idealism. *Noûs*. Vol. 60, no. 1, pp. 87–109. DOI: <https://doi.org/10.1111/nous.12552>

Searle, J.R. (1980). Minds, brains, and programs. *Behavioral and Brain Sciences*. Vol. 3, no. 3, pp. 417–424. DOI: <https://doi.org/10.1017/s0140525x00005756>

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I. and Salakhutdinov, R. (2014). Dropout. A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*. Vol. 15, no. 1, pp. 1929–1958.

Vaswani, A., Shazeer, N., Parmar, N. et al. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*. Vol. 30, pp. 5999–6008. DOI: <https://doi.org/10.65215/r5bs2d54>

Zahavi, D. (2018). Brain, mind, world. Predictive coding, neo-kantianism, and transcendental idealism. *Husserl Studies*. Vol. 34, no. 1, pp. 47–61. DOI: <https://doi.org/10.1007/s10743-017-9218-z>

Об авторе

Торошин Егор Александрович
ассистент кафедры философии

Пермский государственный национальный
исследовательский университет,
614068, Пермь, ул. Букирева, 15;
e-mail: egtoroshchin@gmail.com
ResearcherID: JPL-0328-2023

About the author

Yegor A. Toroshchin
Assistant Lecturer of the Department
of Philosophy

Perm State University,
15, Bukirev st., Perm, 614068, Russia;
e-mail: egtoroshchin@gmail.com
ResearcherID: JPL-0328-2023