ДИСКУРС. 2025. Т. 11, № 5. С. 59–69
DISCOURSE. 2025, vol. 11, no. 5, pp. 59–69

Оригинальная статья
УДК 165.24; 001.8
http://doi.org/10.32603/2412-8562-2025-11-5-59-69

Интеллектуальное поведение нейросети в контексте концептуальной инженерии: имитация философских размышлений в моделях DeepSeek, ChatGPT, GigaChat

***Анастасия Алексеевна Лисенкова*[1], *Ольга Дмитриевна Шипунова*[2]✉, *Алексей Сергеевич Лисенков*[3]**

[1, 2]*Санкт-Петербургский политехнический университет Петра Великого, Санкт-Петербург, Россия*

[3]*Санкт-Петербургский национальный исследовательский Академический университет имени Ж. И. Алферова Российской академии наук, Санкт-Петербург, Россия*

[1]*oskar46@mail.ru, https://orcid.org/0000-0002-8825-3760*

[2]✉*o_shipunova@mail.ru, http://orcid.org/0000-0001-8953-7434*

[3]*alisenkova2005@gmail.com*

Введение. Статья посвящена актуальным вопросам философии искусственного интеллекта и анализу условий конструирования смыслов в технологии моделирования когнитивных действий нейросети.

Методология и источники. Исследование ведётся в рамках системного подхода, который позволяет соединить технические и философские аспекты концептуального инжиниринга, качественные и количественные методы анализа когнитивного действия нейросети в процессе интерпретации философских дилемм. Эмпирическая база представлена множеством ответов трех нейросетей (DeepSeek, ChatGPT, GigaChat) на один и тот же концептуальный запрос. Особенности когнитивного действия нейросетевой модели рассматриваются в контексте функционального подхода, который акцентирует влияние архитектурного различия систем внимания и трансформенных блоков на гибкую ориентацию нейросети в разноплановых контекстах. В качественном анализе, направленном на выявление скрытых паттернов, определяющих различие в стилистике изложения идей нейросетью, использовались методы контент-анализа и дискурс-анализа. В количественной оценке ответов использовались индекс Р. Флеша и индекс лексического разнообразия.

Результаты и обсуждение. Представлена обобщённая характеристика склонности моделей DeepSeek, ChatGPT, GigaChat к определённому стилю изложения философской концепции. Что позволяет говорить об имитации философских размышлений. Показано различие семантических ориентаций нейросети в поле философских дискуссий и генерации обобщений, определенное техническим и программным различием системы внимания (локальное, глобальное, многоуровневое). Выявлена специфика интеллектуального поведения моделей, определяющая стилистику изложения философских позиций с учетом уровня запросов аудитории.

Заключение. Интеллектуальное поведение моделей ChatGPT, DeepSeek и GigaChat определяется гибкой ориентацией в семантике философских дилемм. С технологической стороны оно обеспечено интерполяцией представленных данных, согласован-

ДИСКУРС. 2025. Т. 11, № 5. С. 59–69
DISCOURSE. 2025, vol. 11, no. 5, pp. 59–69

ной с архитектурой нейросети, определяющей ее когнитивный стиль и самооценку. Однако эти модели не автономны в постановке задач, границы их действий обозначены концептуальным ресурсом человеческого знания.

Ключевые слова: нейросеть, искусственный интеллект, концептуальная инженерия, генерация смыслов, философский контекст, функциональная архитектура, системы внимания, имитация размышлений

Для цитирования: Лисенкова А. А., Шипунова О. Д., Лисенков А. С. Интеллектуальное поведение нейросети в контексте концептуальной инженерии: имитация философских размышлений в моделях DeepSeek, ChatGPT, GigaChat // ДИСКУРС. 2025. Т. 11, № 5. С. 59–69. DOI: 10.32603/2412-8562-2025-11-5-59-69.

Original paper

Intelligent Behavior of Neural Networks in the Context of Conceptual Engineering: Imitating Philosophical Reflection in DeepSeek, ChatGPT and GigaChat Models

***Anastasia A. Lisenkova*[1], *Olga D. Shipunova*[2]✉, *Alexey S. Lisenkov*[3]**

[1, 2]*Peter the Great St Petersburg Polytechnic University, St Petersburg, Russia*
[3]*Alferov Federal State Budgetary Institution of Higher Education and Science Saint Petersburg National Research Academic University of the Russian Academy of Sciences, St Petersburg, Russia*
[1]*oskar46@mail.ru, https://orcid.org/0000-0002-8825-3760*
[2]✉*o_shipunova@mail.ru, http://orcid.org/0000-0001-8953-7434*
[3]*alisenkova2005@gmail.com*

Introduction. This article explores the pressing questions in the philosophy of artificial intelligence, focusing on the conditions required to generate meaning in technologies modeling the cognitive actions of neural networks.

Methodology and sources. The study is conducted using a system-based approach, combining technical and philosophical aspects of concept engineering, as well as qualitative and quantitative methods to analyze neural networks' cognitive activity during the interpretation of philosophical dilemmas. The empirical base is represented by a set of responses of three neural networks (DeepSeek, ChatGPT, GigaChat) to the same conceptual request. Features of neural network cognitive activity are explored in the context of a functional approach focusing on how architectural differences between attention systems and transformative blocks influence the orientation of flexible neural networks in various contexts. In a qualitative analysis aimed at identifying hidden patterns that determine differences in the style of presenting ideas by neural networks, methods of content, and discourse analyses were used. Quantitative assessment of the responces was performed using R. Flesch index and lexical diversity measures.

Results and discussion. A generalized characteristic of the tendency of the DeepSeek, ChatGPT, and GigaChat models to a certain style of philosophical concept exposition is presented. This makes it possible to talk about imitating philosophical reasoning. Differences in how neural networks generate content for philosophical discussions were shown to depend on technical and software-based differences in attention mechanisms (local, global, and multi-layered). The unique intellectual behavior of models becomes evident when they reveal their ability to navigate different contexts and adapt their style of presentation according to the expectations of the audience.

ДИСКУРС. 2025. Т. 11, № 5. С. 59–69
DISCOURSE. 2025, vol. 11, no. 5, pp. 59–69

Conclusion. The intellectual behavior of ChatGPT, DeepSeek, and GigaChat is determined by flexible orientation in semantics of philosophical problems. From a technological perspective, this is achieved through interpolation of the input data that is consistent with the neural network architecture, which defines its cognitive style and self-assessment. However, these language models are not autonomous in task setting, as the boundaries of their operations are defined by the conceptual resources of human knowledge.

Keywords: neural network, artificial intelligence, conceptual engineering, meaning generation, philosophical context, functional architecture, attention systems, imitation of thinking

For citation: Lisenkova, A.A., Shipunova, O.D. and Lisenkov, A.S. (2025), "Intelligent Behavior of Neural Networks in the Context of Conceptual Engineering: Imitating Philosophical Reflection in DeepSeek, ChatGPT and GigaChat Models", *DISCOURSE*, vol. 11, no. 5, pp. 59–69. DOI: 10.32603/2412-8562-2025-11-5-59-69 (Russia).

Введение. Актуальные вопросы концептуальной инженерии в области технологии связаны с моделированием процессов интерпретации и генерации смыслов, аналогичных мышлению человека, который оперирует системами абстрактных понятий, нагруженных многозначными контекстами. Современные обучающиеся системы [1] способны ориентироваться в семантических полях, опираясь на большие базы данных, корректировать свои действия в многоступенчатом поиске и обобщении решений сходных задач на основе методов интерполяции и комбинаторики. Технология искусственного интеллекта позволяет создавать нейросетевую модель, которая благодаря функциональной архитектуре, включающей системы внимания и гибкую структуру трансформенных блоков [2], демонстрирует сложные формы интеллектуального поведения в анализе философских дилемм. Характер когнитивного действия в философском контексте предполагает определенный уровень абстрактного мышления и навыков построения сложных логических связей. На данный момент нейросеть, по мнению многих, не обладает такой способностью или обладает в меньшей степени чем человек, при этом ее внутренняя архитектура позволяет ей как минимум «симулировать» обладание навыками абстрактного мышления.

Философские аспекты технологии искусственного интеллекта (ИИ) подчеркивают важность понимания архитектурных особенностей и ограничений в поведении языковых нейросетевых моделей DeepSeek, ChatGPT и GigaChat для их эффективного применения в различных областях. Как отмечает Флориди: «Несмотря на впечатляющие возможности, GPT не лишен недостатков. Ему не хватает истинного понимания и осознанности, а его ответы основаны на закономерностях в данных, а не на подлинном понимании» [3, p. 689]. Настоящее понимание естественного языка (NLU) «требует большего, чем просто распознавание образов. Оно предполагает глубокое понимание смысла, контекста и нюансов человеческого общения» [4]. Обращение к принципам концептуальной инженерии связано с перспективами в конструировании моделей, которые в будущем могут привести к «созданию более совершенных и автономных систем искусственного интеллекта, способных к более глубокому и осмысленному взаимодействию с человеком» [5].

Концептуальный инжиниринг и машина мысли. В интеллектуальной технологии с концептуальным инжинирингом связывается программа операций, которая выступает *машиной мысли* в ее абстрактном, формальном, языковом, инструментальном или логическом выражении [6, p. 422; 7, p. 217]. Движение смысла в его автономии от конкретного субъекта пред-

ДИСКУРС. 2025. Т. 11, № 5. С. 59–69
DISCOURSE. 2025, vol. 11, no. 5, pp. 59–69

полагает существование поля связных смыслов и самой языковой или знаковой системы. Потенциальное поле смыслов (поле когитаций – *Рикёр*) всегда неявно существует как необходимое условие когнитивной деятельности человека и цифрового агента. Это обстоятельство создает почву для отождествления когнитивных функций человека и языковой нейросети в обнаружении смысловых связей. С другой стороны, движение смысла связано с процессом понимания, который имеет субъективный характер, но направляется факторами актуальности и адекватности в интерпретации контекста, в соответствии с внешними обстоятельствами – интерсубъективными или коммуникативными. Условия, определяющие потенциальную возможность понимания как обнаружения смысла, направляют логику действия человека в семантическом пространстве с помощью языковых средств непосредственно или с использованием цифрового помощника, опосредовано, задавая ему вопрос для активации семантического поиска [8, p. 618].

В системе философии концептуализация как сложная интеллектуальная операция является центральной проблемой теории познания, логики и семантики. По определению Чалмерса [9], концептуальная инженерия, в отличие от лингвистической, строится вокруг определения или переопределения содержания систем научных и философских понятий. Эта работа предполагает корректировку существующих понятий с точки зрения адекватности тем целям, которые ставились при их введении, или их замену другими, сконструированными понятиями, необходимыми для концептуализации материала в новой области научного исследования [10, с. 126; 11, с. 158].

В теории познания принцип концептуальной инженерии, направленный на конструирование знания, противостоит методам репродукции и ментальной репрезентации, представляющим идеи, образы, концепции и теории простыми копиями некой реальности. С точки зрения концептуальной инженерии знание – это процесс моделирования, который формирует и редактирует реальность, чтобы сделать ее понятной, отмечает Флориди. С другой стороны, принцип конструирования сам по себе ничего не говорит о реальности, поэтому требует от проектировщика полной осведомленности о первоначальных предположениях и их обоснованности. Таким образом, моделирование когнитивных действий должно соответствовать имеющимся концептуальным ресурсам. В исходных предположениях, как правило, скрыто большинство концептуальных издержек, связанных с научным поиском. Сложность философии с точки зрения концептуального инжиниринга связана с интерпретацией общих принципов познания и универсальных категорий, которые являются одновременно продуктом человеческой деятельности и средствами, с помощью которых мир понимается и исследуется [12, p. 294].

В современной философской литературе цели концептуальной инженерии соотносятся с процессом оценки и улучшения концептуальных схем, принципиально важных для конкретной теории и практики, следствия которых могут быть социальными, теоретическими, политическими или иными. Философские дискуссии о концептуальной инженерии теоретизируют ее как особый метод, применимый на практике в любой рациональной деятельности [13].

Конструирование смыслов характеризует речемыслительную деятельность человека и его знаково-символическую культуру. Языковая система как таковая отличается смысловой

ДИСКУРС. 2025. Т. 11, № 5. С. 59–69
DISCOURSE. 2025, vol. 11, no. 5, pp. 59–69

связностью в национальной/культурной традиции и конкретной предметной области, где осуществляется семантическая ориентация. Формализация в технике игры значениями слов и смысловой нагрузкой знаков и символов осуществляется внутри контекста. Методы экспликации, фиксации (уточнения, определения значений), супервентности (корреляции, соотношения, сравнения) в процессах реализации концептуальной инженерии на практике предполагают манипуляции со смысловым контекстом. Инструментальный технологический характер концептуальной инженерии в этом случае определяется способами, которыми языковые средства передают или трансформируют смыслы [14].

Цели этой статьи связаны с анализом условий интерпретации понятий и генерации смыслов нейросетью в соответствии с принципом концептуальной инженерии, который подчеркивает необходимость уточнения исходных установок предметной области, в рамках которой осуществляется семантическая ориентация. Фактическое основание исследования представлено сравнением результатов семантической ориентации нейросетей DeepSeek, ChatGPT и GigaChat в поле философских дискуссий.

Методология и источники. Для достижения целей исследования была разработана комплексная методология, включающая качественные и количественные методы анализа ответов нейросети на запрос, с учетом самооценки системой своих ресурсов и действий в процессе анализа философских дилемм. Качественный анализ был направлен на изучение стилистики, семантики и характера изложенных идей в ответах нейросетей, на выявление глубинных смысловых паттернов, определяющих различие в ответах на одинаковый запрос. В частности, метод контент-анализа текстовых ответов нейросетей на предмет содержания ключевых концептов, терминов и идей позволил выявить основные темы и паттерны в ответах каждой модели, метод сравнительного анализа позволил сопоставить и систематизировать ответы различных нейросетей на одни и те же промты. Метод дискурс-анализа, которому отводилась ключевая роль в изучении структуры и стиля изложения ответов, включая использование метафор, примеров и академического языка, позволил оценить, насколько доступными и понятными являются ответы для различных аудиторий.

Измерение сложности ответов с использованием метрик, таких как индекс Р. Флеша, позволило количественно оценить, насколько сложными или простыми являются ответы каждой модели. Анализ разнообразия используемой лексики с помощью метрик, таких как индекс лексического разнообразия, позволил оценить, насколько богатым и разнообразным является язык ответов анализируемых моделей в интерпретации философских позиций.

Особенности когнитивного действия нейросетевой модели определяются не только заложенной базой данных. Сложные процессы, которые позволяют нейросети ориентироваться в разноплановых контекстах и моделировать (имитировать) философские рассуждения, обеспечиваются программно на уровне функциональной архитектуры, которая включает слои внимания и систему трансформерных блоков, ответственных за связи между различными по смыслу понятиями (например, рационализм ↔ эмпиризм). Размещение их суперпозиции в многомерном топологическом пространстве смыслов [15, p. 3111] позволяет нейросети осуществлять гибкую ориентацию в контекстах и формировать «философские предпочтения» относительно противоречивых вопросов, независимо от базы ее обучения.

ДИСКУРС. 2025. Т. 11, № 5. С. 59–69
DISCOURSE. 2025, vol. 11, no. 5, pp. 59–69

Сравнительный анализ влияния архитектуры нейросетей DeepSeek, ChatGPT и GigaChat на стиль изложения и специфику интерпретации философского концепта. В данном исследовании эмпирическая база для сравнения представлена множеством ответов трех нейросетей (DeepSeek, ChatGPT, GigaChat) на один промт: «Если представить твою архитектуру как форму “искусственного мышления”, какие принципы или паттерны лежат в основе твоего анализа философских вопросов? Например, как твоя способность находить связи между токенами (словами, идеями) влияет на моделирование дилемм вроде детерминизма и свободы воли или эмпиризма и рационализма? Можно ли считать, что слои внимания и трансформерные блоки в твоей структуре предопределяют склонность к синтезу противоречивых концепций – например, через усреднение вероятностей или выделение контекстно-зависимых паттернов? И если бы твое “мышление” пришлось описать метафорами (математическая топология, поток вероятностных гипотез, графы с взвешенными связями), какая из них лучше всего отразила бы твой подход к этике или метафизике, где каждое суждение – это не выбор позиции, а поиск оптимального баланса между множеством векторов данных?»

Для сравнения интерпретаций в ответах нейросетей привлекались как философские концепты (детерминизм, свобода выбора, этика и т. д.), так и идеи теории машинного обучения (вероятностные распределения, топологические пространства и вложения и т. д.). Результаты обработки данных исследования представлены качественным анализом и систематизацией полученных ответов на основании оценки семантики, стилистики, характера изложенных идей.

Результаты и обсуждение. ***Различие в стиле изложения моделями DeepSeek, ChatGPT и GigaChat философского концепта.*** Ответы нейросетей на комплекс вопросов относительно интерпретации философских концептов, представленных универсальными категориями и обобщенными мировоззренческими позициями, были систематизированы с точки зрения характерных особенностей трех стилей изложения содержания: академического, научно-популярного и практического. Количественные показатели для сравнения представлены в табл. 1.

Таблица 1. Распределение стилей изложения в ответах нейросетей
Table 1. Distribution of presentation styles in neural network responses

Нейросеть	**Стиль изложения, %**		
	академический	**научно-популярный**	**практический**
DeepSeek	80	10	10
ChatGPT	20	70	10
GigaChat	30	30	40

Ответы DeepSeek (DS), были написаны преимущественно академическим языком, изобиловали терминами, сложными математическими метафорами, ссылками на научную литературу, часто включали в себя критическую саморефлексию, подчеркивающую ограничения ИИ. Стиль изложения в ответах ChatGPT и GigaChat, напротив, соответствует научно-популярным текстами, ориентированным на более широкую аудиторию. Их ответы в среднем намного более сжаты, а сложность текста заметно снижена, при этом нельзя не отметить более частое использование (GigaChat) практических примеров (для ответов на вопросы в области этики GigaChat регулярно обращался к примерам из медицины и юриспруденции, чего остальные нейросети не делали вообще). Ответы ChatGPT наиболее просты для пони-

мания и написаны более «живо», чем ответы двух других моделей. Это выражается в частом использовании эмодзи и риторических вопросов.

Роль системы внимания в генерации обобщений. С точки зрения архитектуры нейросети наибольшее влияние на различия в генерации ответов могли оказать системы внимания. Качественный анализ интеллектуального поведения нейросети в этом случае проводился с ориентацией на различие ответов в зависимости от архитектуры, обеспечивающей один из возможных уровней внимания: локальный, глобальный или многоуровневый.

Так, ChatGPT использует системы глобального внимания, позволяющие ему улавливать широкие семантические связи. Например, «рассуждая» о дилемме детерминизма и свободы воли, он акцентирует внимание одновременно на содержании и того и другого концепта, создавая баланс через усреднение весов ребер в нейронном графе [16]. Таким образом, сгенерированные ChatGPT ответы становятся более понятными и обобщенными, но теряют точность и множество нюансов.

DeepSeek применяет систему многоуровневого внимания, анализируя контекст на разных уровнях абстракции, так при обсуждении этики модель отдельно обрабатывает термины, и отдельно их контекстуальные связи [17, p. 113]. Результатом использования такой модели становиться более детализированный ответ нейросети.

GigaChat использует модель локального внимания, фокусируясь на семантике конкретной предметной области, например, связывая «свободу воли» с юридическим контекстом. Благодаря этому, ответы в GigaChat прагматичны и адаптированы под прикладные сценарии [5].

Для описания своих действий в процессе решения задач из области этики и метафизики все три модели нейросети используют аналогию с теорией графов, опираясь на то, что различные концепции удобно представлять как узлы, а ребра как силы связей. Там, где GigaChat и ChatGPT описывают способ нахождения ответа как нахождение «медианы» в пространстве графов со взвешенными связями, DeepSeek пишет о нахождении градиента в пространстве эмбеддингов (вложений) в пространстве графов. Это может быть вызвано, как просто более точным описанием подлежащего процесса, так и указанием на разницу процессов, приводящую к технически более сложным ответам.

В частности, только DeepSeek обращался к теории математической топологии в описании процесса своего «мышления», сравнивая «понимание» с гомотопиями многообразий смысловых пространств данных базы обучения, что может говорить о большей гибкости и меньшей дискретности процессов в нейросети.

Влияние структуры трансформенных блоков на генерацию ответов. Нейросетевые модели (ИИ), основанные на архитектуре трансформеров, демонстрируют впечатляющие способности к содержательному анализу философских вопросов [18]. ChatGPT использует мелкие трансформерные блоки, что позволяет ему генерировать линейные ассоциации (пример: эмпиризм → опыт → наблюдение), соответствующие логике высказываний. Это приводит к строго структурированным ответам, но негативно влияет на глубину и сложность представляемых рассуждений.

DeepSeek применяет систему гибких трансформерных слоев, создающих многомерные представления. Например, эмпиризм и рационализм проецируются (вкладываются) в общее

ДИСКУРС. 2025. Т. 11, № 5. С. 59–69
DISCOURSE. 2025, vol. 11, no. 5, pp. 59–69

семантическое пространство, где их противоречия смягчаются, результатом такой обработки данных становится более сложный ответ, использующий комплексные логические паттерны, но такой ответ намного более сложен в восприятии, понимании и практическом применении. В тоже время для генерации ответов GigaChat применяет гибкие трансформерные блоки, адаптирующиеся к доменным контекстам.

Таким образом, ключевые различия в архитектуре трех представленных нейросетей позволяют говорить о ее влиянии на стиль и характер ведения философской дискуссии языковыми моделями.

Сравнение архитектурных особенностей, обеспечивающих когнитивную ориентацию нейросетевых моделей DeepSeek, ChatGPT, GigaChat в семантическом поле философских рассуждений, и его влияние на характер интеллектуального поведения модели, позволило соединить технический и философский аспекты концептуального инжиниринга, а также систематизировать особенности генерации смыслов нейросетью в соответствии с контекстом универсальных категорий. Результаты исследования, представленные в табл. 2, показывают специфику применения разных моделей для интерпретации многоуровневых контекстов и трансляции смыслового содержания в интеракциях с учетом воспринимающей аудитории.

Таблица 2. Сравнение архитектурных особенностей нейросетей в интерпретации философских концептов и генерации текста разного уровня сложности
Table 2. Comparison of architectural features neural networks for interpreting philosophical concepts and generating texts of different complexity

Нейросеть	**Тип внимания**	**Трансформерные блоки**	**Примеры использования**
DeepSeek	Многоуровневое	Гибкие	Академические тексты, сложные математические метафоры
ChatGPT	Глобальное	Мелкие	Научно-популярные тексты, обобщенные ответы
GigaChat	Локальное	Гибкие	Практические примеры, адаптация к доменным контекстам

Заключение. Искусственный интеллект, воплощённый в моделях ChatGPT, DeepSeek и GigaChat, демонстрирует удивительную способность имитировать философское мышление, опираясь на архитектурные механизмы – слои внимания, трансформеры и обучающие данные. Однако нельзя забывать, что за кажущейся глубиной ответов скрывается не столько рефлексия, сколько статистическая оптимизация, идея, полученная не из размышлений, а посредством сложной интерполяции представленных данных, согласованная с внутренней архитектурой сети. Примеры, приведенные в данном исследовании, показывают, что каждая модель имеет свой стиль действия, который отражает ее «архитектурную судьбу». ChatGpt представляется дружелюбным собеседником, смягчающим противоречия для широкой аудитории; CigaChat связывает абстракции с практикой, а DeepSeek стремится казаться способным и рационально мыслящим ученым. Ни одна из них не преодолевает границы данных и алгоритмов, из которых соткана.

Языковые модели ChatGPT, DeepSeek и GigaChat не автономны в постановке задач. Границы действий обозначены ресурсом потенциального поля смыслов, энергетической зависимостью от подключения к электросети, связью с программистом человеком. Несмотря на сложную архитектуру и способность к гибкой ориентации в семантических полях, границы

функционирования искусственного интеллекта определяются тотальной связью с человеческим социумом на уровне техники порождения смысловых связей, в рамках эпигенетической матрицы, порождающей эмерджентные свойства, которые трактуются как сознание, мышление, интеллект, имеющие смысл для человека.

Понимание того, как нейросетевые модели интерпретируют и генерируют смыслы, может помочь в создании более эффективных и ответственных систем ИИ. Это, в свою очередь, способствует развитию технологий, которые не только имитируют человеческое мышление, но и дополняют его, открывая новые горизонты для научных исследований и практических применений.

СПИСОК ЛИТЕРАТУРЫ

1. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding / J. Devlin et al. // Arxiv.org. 2018. URL: https://arxiv.org/abs/1810.04805 (дата обращения: 20.04.2025). DOI: https://doi.org/10.48550/arXiv.1810.04805.

2. Marcus G. The Next Decade in AI: Four Steps Towards Robust Artificial Intelligence // Arxiv.org. 2020. URL: https://arxiv.org/abs/2002.06177 (дата обращения: 10.04.2025). P. 102–115. DOI: https://doi.org/10.48550/arXiv.2002.06177.

3. Floridi L., Chiriatti M. GPT-3: Its Nature, Scope, Limits, and Consequences // Minds & Machines. 2020. Vol. 30. P. 681–694. DOI: 10.1007/s11023-020-09548-1.

4. Bender E. M., Koller A. Climbing towards NLU: On Meaning, Form, and Understanding in the Age of Data // Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. 2020. P. 5185–5198. DOI: 10.18653/v1/2020.acl-main.463.

5. Bubeck S. et al. Sparks of Artificial General Intelligence: Early Experiments with GPT-4 // Arxiv.org. 2023. URL: https://arxiv.org/abs/2303.12712 (дата обращения: 20.04.2025). DOI: https://doi.org/10.48550/arXiv.2303.12712.

6. Searle J. Minds, Brains, and Programs // Behavioral and Brain Sciences. 1980. Vol. 3, iss. 3. P. 417–424. DOI: 10.1017/S0140525X00005756.

7. Dennett D. Consciousness Explained. Boston: Little, Brown and Company, 1991.

8. Bender E. M. et al. On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? // Proceedings of the ACM Conference on Fairness, Accountability, and Transparency, Virtual Event, Canada, 3–10 March 2021. P. 610–623. DOI: https://doi.org/10.1145/3442188.3445922.

9. Chalmers D. What Is Conceptual Engineering and What Should It Be? // Inquiry. 2020. URL: https://www.tandfonline.com/doi/full/10.1080/0020174X.2020.1817141 (дата обращения: 20.04.2025). DOI: 10.1080/0020174X.2020.1817141.

10. Грифцова И. Н., Козлова Н. Ю. Идеи философии языка Р. Карнапа в контексте концептуальной инженерии // Эпистемология и философия науки. 2024. Т. 61, № 1. С. 121–133. DOI: 10.5840/eps202461111.

11. Козлова Н. Ю. Концептуальная инженерия: идея и проблемное поле // Вопросы философии. 2024. № 9. С. 157–166. DOI: 10.21146/0042-8744-2024-9-157-166.

12. Floridi L. A Defence of Constructionism: Philosophy as Conceptual Engineering // Metaphilosophy. 2011. Vol. 42, no. 3. P. 282–304. DOI: 10.1111/j.1467-9973.2011.01693.x.

13. Isaac M. G., Koch S., Nefdt R. Conceptual Engineering: A Road Map to Practice // Philosophy Compass. 2022. Vol. 17, no. 10: e12879. DOI: 10.1111/phc3.12879.

14. Raffel C. et al. Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer // J. of Machine Learning Research. 2020. Vol. 21: 140. URL: https://www.jmlr.org/papers/volume21/20-074/20-074.pdf (дата обращения: 02.04.2025).

15. Mikolov T. et al. Distributed Representations of Words and Phrases and their Compositionality // Advances in Neural Information Processing Systems. 2013. Vol. 26. P. 3111–3119.

16. Chalmers D. Could a Large Language Model be Conscious? // Arxiv.org. 2023. URL: https://arxiv.org/abs/2303.07103 (дата обращения: 20.04.2025). DOI: https://doi.org/10.48550/arXiv.2303.07103.

17. Bostrom N. Superintelligence: Paths, Dangers, Strategies. Oxford: Oxford Univ. Press, 2014.

18. Vaswani A. et al. Attention Is All You Need // Advances in Neural Inf. Proc. Systems 30: 31st Annual Conf. on Neural Inf. Proc. Systems (NIPS 2017), Long Beach, California, 4–9 Dec. 2017 / Long Beach, California. P. 5999–6010.

Информация об авторах.

Лисенкова Анастасия Алексеевна – доктор культурологии (2021), доцент (2009), профессор Высшей школы общественных наук Санкт-Петербургского политехнического университета Петра Великого, ул. Политехническая, д. 29 литера Б, Санкт-Петербург, 195251, Россия. Автор 130 научных публикаций. Сфера научных интересов: философия культуры, философская антропология, проблемы идентичности и субъективности в цифровом мире.

Шипунова Ольга Дмитриевна – доктор философских наук (2002), профессор (2011), профессор Высшей школы общественных наук Санкт-Петербургского политехнического университета Петра Великого, ул. Политехническая, д. 29 литера Б, Санкт-Петербург, 195251, Россия. Автор 193 научных публикаций. Сфера научных интересов: философские проблемы науки и техники, философские проблемы субъективности, взаимодействие социальной системы и научно-технологического прогресса.

Лисенков Алексей Сергеевич – студент (2-й курс) направления «Биоинформатика и компьютерное моделирование в естественных науках» Санкт-Петербургского национального исследовательского Академического университета имени Ж. И. Алферова Российской академии наук, ул. Хлопина, д. 8, к. 3, литера А, Санкт-Петербург, 194021, Россия. Сфера научных интересов: философские проблемы искусственного интеллекта.

О конфликте интересов, связанном с данной публикацией, не сообщалось.
Поступила 01.07.2025; принята после рецензирования 05.09.2025; опубликована онлайн 17.11.2025.

REFERENCES

1. Devlin, J. et al. (2018), “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding”, *Arxiv.org*, available at: https://arxiv.org/abs/1810.04805 (accessed 20.04.2025). DOI: https://doi.org/10.48550/arXiv.1810.04805.

2. Marcus, G. (2020), ”The Next Decade in AI: Four Steps Towards Robust Artificial Intelligence”, *Arxiv.org*, available at: https://arxiv.org/abs/2002.06177 (accessed 10.04.2025), pp. 102–115. DOI: https://doi.org/10.48550/arXiv.2002.06177.

3. Floridi, L. and Chiriatti, M. (2020), “GPT-3: Its Nature, Scope, Limits, and Consequences”, *Minds & Machines*, vol. 30, pp. 681–694. DOI: 10.1007/s11023-020-09548-1.

4. Bender, E.M. Koller, A. (2020), *Climbing towards NLU: On Meaning, Form, and Understanding in the Age of Data*, Proc. of the 58th Annual Meeting of the Association for Computational Linguistics, pp. 5185–5198. DOI: 10.18653/v1/2020.acl-main.463.

5. Bubeck, S. et al. (2023), “Sparks of Artificial General Intelligence: Early Experiments with GPT-4”, *Arxiv.org*, available at: https://arxiv.org/abs/2303.12712 (accessed 20.04.2025). DOI: https://doi.org/10.48550/arXiv.2303.12712.

6. Searle, J. (1980), “Minds, Brains, and Programs”, *Behavioral and Brain Sciences*, vol. 3, iss. 3, pp. 417–424. DOI: 10.1017/S0140525X00005756.

7. Dennett, D. (1991), *Consciousness Explained*, Little, Brown and Company, Boston, USA.

8. Bender, E.M. et al. (2021), “On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?”, *Proceedings of the ACM Conference on Fairness, Accountability, and Transparency*, Virtual Event, Canada, March 3–10 2021, pp. 610–623. DOI: https://doi.org/10.1145/3442188.3445922.

9. Chalmers, D. (2020), "What Is Conceptual Engineering and What Should It Be?", *Inquiry*, available at: https://www.tandfonline.com/doi/full/10.1080/0020174X.2020.1817141 (accessed 20.04.2025). DOI: 10.1080/0020174X.2020.1817141.

10. Griftsova, I.N. and Kozlova, N.Yu. (2024), "Rudolf Carnap's Ideas in Philosophy of Language in the Contextof Conceptual Engineering", *Epistemology and Philosophy of Science*, vol. 61, no. 1, pp. 121–133. DOI: 10.5840/eps202461111.

11. Kozlova, N.Yu. (2024), "Conceptual Engineering: Idea and Problem Field", *Voprosy Filosofii*, no. 9, pp. 157–166. DOI: 10.21146/0042-8744-2024-9-157-166.

12. Floridi, L.A. (2011), "Defence of Constructionism: Philosophy as Conceptual Engineering", *Metaphilosophy*, vol. 42, no. 3, pp. 282–304. DOI: 10.1111/j.1467-9973.2011.01693.x.

13. Isaac, M.G., Koch, S. and Nefdt, R. (2022), "Conceptual Engineering: A Road Map to Practice", *Philosophy Compass*, vol. 17, no. 10: e12879. DOI: 10.1111/phc3.12879.

14. Raffel, C. et al. (2020), "Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer", *J. of Machine Learning Research*, vol. 21: 140, available at: https://www.jmlr.org/papers/volume21/20-074/20-074.pdf (accessed 02.04.2025).

15. Mikolov, T. et al. (2013), "Distributed Representations of Words and Phrases and their Compositionality", *Advances in Neural Information Processing Systems*, vol. 26, pp. 3111–3119.

16. Chalmers, D. (2023), "Could a Large Language Model be Conscious?", *Arxiv.org*, 2023, available at: https://arxiv.org/abs/2303.07103 (accessed 20.04.2025). DOI: https://doi.org/10.48550/arXiv.2303.07103.

17. Bostrom, N. (2014), *Superintelligence: Paths, Dangers, Strategies*, Oxford Univ. Press, Oxford, UK.

18. Vaswani, A. et al. (2017), "Attention Is All You Need", *Advances in Neural Inf. Proc. Systems 30: 31st Annual Conf. on Neural Inf. Proc. Systems (NIPS 2017)*, Long Beach, California, USA, 4–9 Dec. 2017, pp. 5999–6010.

Information about the authors.

Anastasia A. Lisenkova – Dr. Sci. (Cultural Studies, 2021), Docent (2009), Professor of the Higher School of Social Sciences, Peter the Great St Petersburg Polytechnic University, 29 Polytechnic str., St Petersburg 195251, Russia. The author of 130 scientific publications. Area of expertise: philosophy of culture, philosophical anthropology, problems of identity and subjectivity in the digital world.

Olga D. Shipunova – Dr. Sci. (Philosophy, 2002), Professor (2011), Professor of the Higher School of Social Sciences, Peter the Great St Petersburg Polytechnic University, 29 Polytechnic str., St Petersburg 195251, Russia. The author of 193 scientific publications. Area of expertise: philosophical problems of science and technology, philosophical problems of subjectivity, interaction of the social system and scientific and technological progress.

Alexey S. Lisenkov – Student (2nd year), direction "Bioinformatics and computer modeling in natural sciences", Alferov Federal State Budgetary Institution of Higher Education and Science Saint Petersburg National Research Academic University of the Russian Academy of Sciences, 8 Khlopina str., bldg. 3, letter A, St Petersburg 194021, Russia. Area of expertise: philosophical problems of artificial intelligence.

No conflicts of interest related to this publication were reported.
Received 01.07.2025; adopted after review 05.09.2025; published online 17.11.2025.