

## ПРОБЛЕМА МЫШЛЕНИЯ В ЭПОХУ БОЛЬШИХ ЯЗЫКОВЫХ МОДЕЛЕЙ И ФУНКЦИОНАЛИСТСКАЯ ПЕРСПЕКТИВА

*Терешкова К.Д., магистрант*

*Белорусский государственный университет информатики и радиоэлектроники  
г. Минск, Республика Беларусь*

*Чуешов В.И. – д-р фил. наук, профессор*

Целью данной работы является философско-методологический анализ соответствия нейросетевых моделей критериям функционалистской теории сознания. Методологическую основу исследования составляют гипотетико-дедуктивный метод, сравнительный анализ, герменевтический подход, а также принципы системного анализа.

В данной работе рассматривается проблема интерпретации мышления в условиях стремительного развития больших языковых моделей, основанных на архитектуре глубокого обучения Transformer. Объектом исследования выступает архитектура Transformer как потенциальная когнитивная модель, а предметом анализа является онтологический и эпистемологический статус вычислительных операций, реализуемых в рамках данной архитектуры.

Функционализм, являясь одной из доминирующих теоретических парадигм в философии сознания, исходит из предположения о том, что ментальные состояния определяются не своей субстанциональной природой, а выполняемой ими функцией в рамках системы причинно-следственных отношений [1; 2]. Данный подход предполагает, что мышление может быть реализовано на различных физических носителях, включая биологические и искусственные системы, при условии сохранения функциональной структуры. Концепция множественной реализуемости, разработанная в аналитической философии, позволяет утверждать автономию когнитивных процессов по отношению к их материальной основе [3]. В свою очередь, концепция интенциональной установки [4] предлагает рассматривать сложные системы как обладающие «намерениями» в прагматическом смысле, что открывает возможность интерпретации поведения искусственных систем в терминах рациональности.

Однако функционализм сталкивается с рядом фундаментальных возражений. Классическая критика связана с проблемой субъективного опыта, или квалиа, которая демонстрирует ограниченность чисто функционального описания сознания [5]. Аргумент «китайской комнаты» [6] подчеркивает различие между синтаксической обработкой символов и их семантическим пониманием, показывая, что система может демонстрировать корректное поведение без наличия внутреннего понимания. Эти аргументы сохраняют свою актуальность и в контексте анализа современных нейросетевых моделей, что делает проблему мышления в эпоху искусственного интеллекта особенно сложной и многослойной.

Архитектура Transformer, предложенная в рамках современных исследований в области машинного обучения, представляет собой качественно новый этап в развитии вычислительных моделей [7]. Ее ключевым элементом является механизм самовнимания, позволяющий системе учитывать глобальные зависимости между элементами входной последовательности. Это обеспечивает формирование контекстуально насыщенных представлений и высокую эффективность при решении задач обработки естественного языка. С функционалистской точки зрения данная архитектура может рассматриваться как система, реализующая когнитивные функции за счет сложной организации вычислительных процессов.

Особое значение имеет тот факт, что большие языковые модели демонстрируют поведение, которое может интерпретироваться как проявление когнитивных способностей. Речь идет о способности к обобщению, логическому выводу, генерации связного текста и адаптации к новым задачам. Эмпирические исследования показывают, что при увеличении масштаба моделей возникают эмерджентные свойства, не сводимые к отдельным компонентам системы [8]. Это позволяет говорить о функциональном сходстве с некоторыми аспектами человеческого мышления, что усиливает позиции функционалистского подхода.

Тем не менее, критический анализ показывает, что такие системы остаются принципиально ограниченными. Отсутствие семантического заземления означает, что символы, обрабатываемые моделью, не связаны напрямую с реальностью [9]. Это приводит к тому, что генерируемые ответы, несмотря на их правдоподобие, не являются результатом понимания в строгом смысле. Таким образом, возникает разрыв между внешним поведением системы и ее внутренним состоянием, что затрудняет интерпретацию нейросетей как полноценных носителей мышления.

Развитие больших языковых моделей оказывает существенное влияние на эпистемологию. В условиях экспоненциального роста данных и вычислительных возможностей формируется новая модель знания, в которой приоритет отдается статистическим закономерностям и корреляциям [10]. Это приводит к трансформации научного метода, смещая акцент с объяснения на предсказание. В результате возникает феномен «черного ящика», при котором система демонстрирует высокую

эффективность, но остается непрозрачной для анализа. Такая ситуация ставит под сомнение традиционные критерии научной рациональности, включая интерпретируемость и обоснованность.

Дополнительное усложнение связано с социально-философскими последствиями внедрения нейросетевых технологий. Генеративные модели трансформируют понятие авторства, размывая границы между субъектом и инструментом. Они также создают условия для формирования новой информационной среды, в которой различие между истиной и правдоподобием становится менее очевидным. В этом контексте усиливаются риски, связанные с манипуляцией информацией, автоматизацией принятия решений и усилением социального неравенства [11]. Это требует разработки новых подходов к регулированию и этической оценке технологий искусственного интеллекта.

Перспективы дальнейшего развития связаны с решением проблемы согласования целей искусственных систем с человеческими ценностями. Исследования в области глобальных рисков и суперинтеллекта подчеркивают необходимость разработки механизмов контроля и управления интеллектуальными системами [11]. В противном случае существует риск возникновения систем, поведение которых будет трудно предсказать и контролировать.

Таким образом, архитектура Transformer демонстрирует значительное соответствие функционалистским критериям мышления на уровне формальных операций, однако не решает проблему субъективного опыта и семантического понимания. Это позволяет рассматривать большие языковые модели как мощные инструменты моделирования когнитивных процессов, но не как полноценные носители сознания. Основным философским вызовом заключается в необходимости переосмысления понятий мышления, знания и субъекта в условиях цифровой трансформации современного общества.

**Список использованных источников:**

1. Putnam H. *Psychological Predicates* // *Art, Mind, and Religion*. – Pittsburgh : University of Pittsburgh Press, 1967.
2. Putnam H. *Minds and Machines* // *Dimensions of Mind* / ed. S. Hook. – New York : New York University Press, 1960.
3. Fodor J. A. *Special Sciences (Or: The Disunity of Science as a Working Hypothesis)* // *Synthese*. – 1974. – Т. 28. – №. 2. – С. 97–115.
4. Dennett D. C. *The Intentional Stance*. – Cambridge : MIT Press, 1987.
5. Nagel T. *What Is It Like to Be a Bat?* // *The Philosophical Review*. – 1974. – Т. 83. – №. 4. – С. 435–450.
6. Searle J. R. *Minds, brains, and programs* // *Behavioral and Brain Sciences*. – 1980. – Т. 3. – №. 3. – С. 417–457.
7. Vaswani A., Shazeer N., Parmar N., Uszkoreit J., Jones L., Gomez A. N., Kaiser Ł., Polosukhin I. *Attention Is All You Need* // *Advances in Neural Information Processing Systems*. – 2017. – Т. 30.
8. Bubeck S. et al. *Sparks of Artificial General Intelligence: Early experiments with GPT-4* // *arXiv preprint*. – 2023.
9. Bender E. M., Gebru T., McMillan-Major A., Shmitchell S. *On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?* // *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*. – 2021. – С. 610–623.
10. Anderson C. *The End of Theory: The Data Deluge Makes the Scientific Method Obsolete* // *Wired*. – 2008.
11. Bostrom N. *Superintelligence: Paths, Dangers, Strategies*. – Oxford : Oxford University Press, 2014.