

Секция «Интеллект человека и искусственный интеллект: кто кого использует?»

Дуальность искусственного мышления: большие языковые модели и модели рассуждения в зеркале человеческой психологии

Научный руководитель – Горбунов Иван Анатольевич

Бурангулов Р.Р.¹, Стребкова К.Е.²

1 - Санкт-Петербургский государственный университет, Факультет психологии, Санкт-Петербург, Россия, *E-mail: burangulovbiz@gmail.com*; 2 - Санкт-Петербургский государственный университет, Факультет психологии, Санкт-Петербург, Россия, *E-mail: streb.karina@gmail.com*

Разработка искусственного интеллекта (ИИ) демонстрирует всё большую дихотомию между большими языковыми моделями (LLM) и системами формализованного рассуждения (Reasoning моделями). Эта дивергенция особенно интересна для психологической науки, поскольку отражает дуализм природы когнитивных процессов: противопоставление интуитивных связей аналитическому мышлению [3]. Создавая ИИ, человек моделирует свою когнитивную систему и переосмысляет её через призму созданных моделей. Цель настоящей статьи – сравнительный анализ архитектур LLM и Reasoning моделей для выявления их соответствий когнитивным процессам человека и формулирования перспектив взаимодействия человеческого и искусственного интеллекта. Как уже упоминалось, LLM и Reasoning модели принципиально различаются. LLM, такие как ChatGPT (OpenAI), Claude (Anthropic) и Grok (xAI), — нейронные сети, обученные на текстовых корпусах методом предсказания следующего токена [9]. Технология основана на имплицитном статистическом обучении, при котором система усваивает корреляции без эксплицитного представления о семантике и грамматике. Можно сказать, что LLM реализуют эмерджентный подход, где знание возникает из обобщения паттернов — аналогично тому, как сложные репрезентации формируются из множества простых ассоциаций [5]. Напротив, Reasoning модели (IBM Watson for Decision Support, Cys, Neural Theorem Provers и др.) интегрируют символические структуры и эксплицитные механизмы логического вывода, которые в том числе следят за своими рассуждениями [4]. Подобные модели реализуют подход, соответствующий тем парадигмам в психологии, где есть акцент на алгоритмических процессах в ментальном пространстве [2]. Продолжая дихотомию, можно заметить, что обе модели демонстрируют работу двух систем мышления, предложенных Д. Канеманом [3]. LLM схожи с Системой 1: быстры, автоматичны, оперируют без видимых усилий, параллельно обрабатывая множество ассоциаций и контекстуальных сигналов. Reasoning модели соответствуют Системе 2: они работают последовательно, медленно, с отслеживанием логических шагов и метакогнитивным контролем процесса рассуждения. Работа LLM характеризуется тремя феноменами, имеющими аналоги в человеческой психике. Во-первых, контекстуальная чувствительность, схожая с эффектами прайминга: небольшие изменения формулировки задачи или контекста могут радикально изменить ответ, показывая отсутствие стабильной внутренней модели реальности [10]. Во-вторых, склонность к конфабуляции — генерации правдоподобных, но фактически неверных объяснений при отсутствии необходимой информации, что напоминает аналогичный феномен из практики клинических психологов [6]. В-третьих, «ассоциативный дрейф» — тенденция к постепенному тематическому

смещению в процессе речепорождения, что находит отражение в феномене свободных ассоциаций в психике [7]. Reasoning модели отличает способность к декомпозиции сложных проблем на иерархию более простых подзадач, а также способность к построению причинно-следственных связей и моделированию сценариев, отвечая на вопросы типа «что, если...», что характеризует абстрактное мышление у человека. Особенно показательна их способность к критике собственных процессов рассуждения, что соответствует метакогнитивным навыкам.

Таким образом, погружение в архитектуры ИИ позволяет изолированно анализировать различные аспекты работы человеческого познания. Одновременно с этим происходит обратный процесс: регулярное взаимодействие с интеллектуальными системами постепенно трансформирует когнитивные привычки и ментальные модели самого человека. Так, использование искусственного интеллекта экономит когнитивные усилия, высвобождая ресурсы для более сложных и креативных задач [8].

Происходит своеобразная «когнитивная коэволюция», где человеческий интеллект адаптируется к взаимодействию с искусственным, перенимая его операциональные логики и эвристики. Этот процесс напоминает эффект, описанный Л. С. Выготским: внешние инструменты интериоризируются, становясь частью внутренней когнитивной архитектуры [1]. Искусственный интеллект, задуманный как инструмент, постепенно становится агентом формирования новых когнитивных структур.

Таким образом, исследования ИИ предоставляют психологии формализованные модели обработки информации и новые метафоры для концептуализации мышления. Параллельно психологическая наука обогащает разработку ИИ пониманием когнитивных ограничений, инсайтами о природе социального познания, а также теоретическими рамками для интерпретации эмерджентных свойств сложных систем. Эта взаимная инструментализация ведет к симбиотическим отношениям, где границы между «пользователем» и «инструментом» размываются. Человеческий и искусственный интеллект становятся взаимными зеркалами, через которые каждая сторона получает возможность увидеть и понять себя в новом свете.

Источники и литература

- 1) 1. Выготский Л. С. Развитие высших психических функций. М., 1960. 500 с.
- 2) 2. Fodor J. A. Minds without meanings: An essay on the content of concepts. MIT Press, 2015.
- 3) 3. Kahneman D. and Frederick S., 2007. Frames and brains: Elicitation and control of response tendencies // Trends in cognitive sciences. Т. 11. №2. С.45-46.
- 4) 4. Lake B., Baroni M. Generalization without systematicity: On the compositional skills of sequence-to-sequence recurrent networks // International conference on machine learning. PMLR. 2018. С. 2873-2882.
- 5) 5. McClelland J. L. et al. Placing language in an integrated understanding system: Next steps toward human-level performance in neural language models // Proceedings of the National Academy of Sciences. 2020. Т. 117. №. 42. С. 25966-25974.
- 6) 6. Mitchell R. et al. A model of cue integration as vector summation in the insect brain // Proceedings of the Royal Society B. 2023. Т. 290: 20230767.
- 7) 7. Rodríguez M. J. G. et al. Collusion detection in public procurement auctions with machine learning algorithms // Automation in Construction. 2022. Т. 133: 104047.
- 8) 8. Tortorella G. L. et al. How does artificial intelligence impact employees' engagement in lean organisations? // International Journal of Production Research. 2024. Т. 63. №3. С. 1011-1027.

- 9) 9. Vaswani A. et al. Attention is all you need // Advances in neural information processing systems. 30-31st Conference on Neural Information Processing Systems, Long Beach, CA, USA. 2017: <https://proceedings.neurips.cc/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf>
- 10) 10. Zhao H. et al. Explainability for large language models: A survey // ACM Transactions on Intelligent Systems and Technology. 2024. Т. 15. №. 2. С. 1-38.