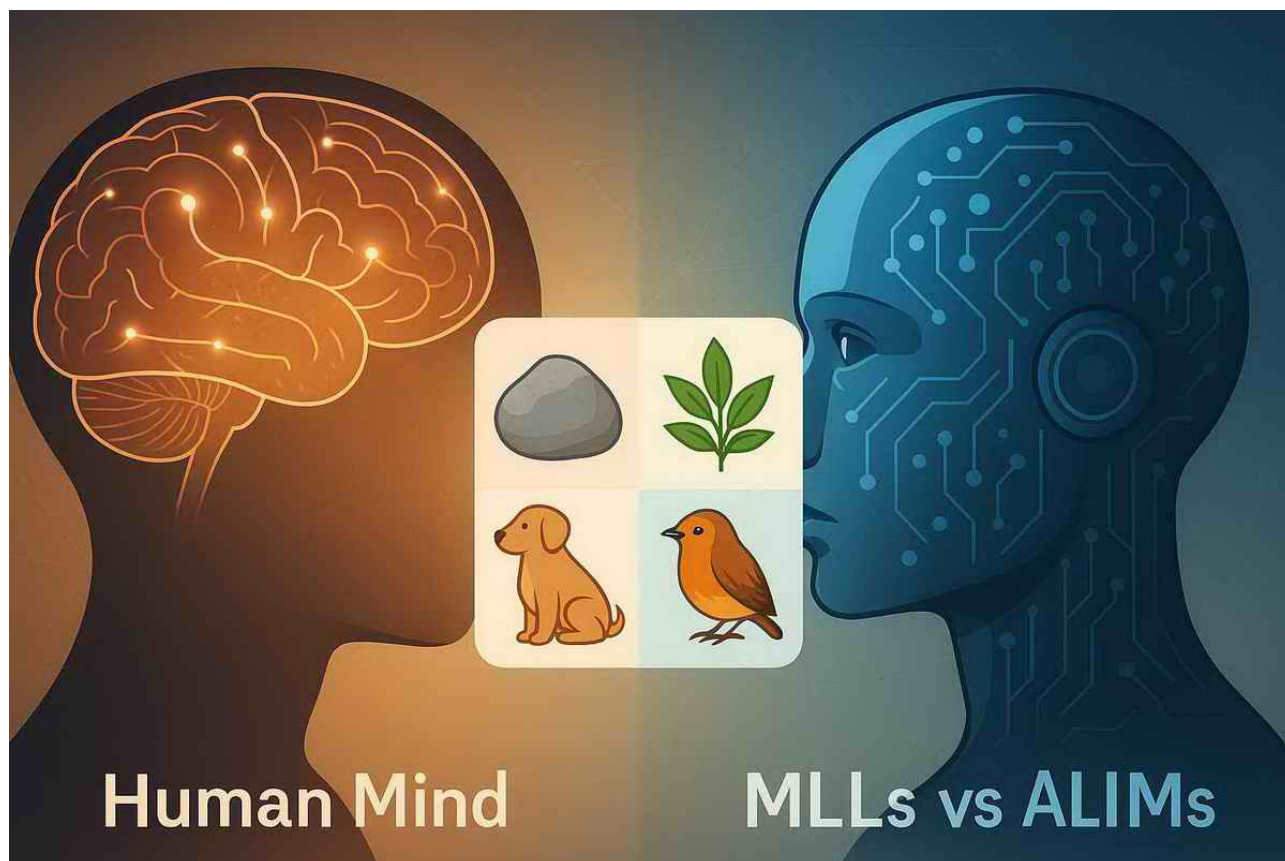


Как искусственный интеллект и мозг человека видят одно и то же: открытия в нейроподобных моделях LLM



Дата публикации: 26.06.2025

Большие языковые модели нового поколения, способные одновременно обрабатывать текст и изображения, открывают уникальное окно в то, как машины могут формировать представления об окружающем мире, схожие с человеческими. Учёные из Китайской академии наук провели масштабное исследование, в ходе которого сравнили работу нейросетей и мозга человека при восприятии и категоризации природных объектов.

В качестве основы они использовали модели ChatGPT-3.5 и GeminiPro Vision 1.0 от DeepMind — современные мультимодальные системы, способные интерпретировать изображения и описания объектов. В ходе эксперимента модели решали так называемые триплетные задачи: из трёх объектов нужно было выбрать два наиболее похожих. Такие же задания давались и людям, а их мозговая активность отслеживалась с помощью нейровизуализации.

В результате анализа 4,7 миллиона триплетных суждений были построены

66-мерные вложения — математические пространства, отражающие взаимосвязи между объектами. Эти вложения показали устойчивость, предсказуемость и семантическую организованность. Объекты в них группировались по смысловым признакам: животные, растения, инструменты и так далее. Такая структура удивительно напоминала, как человек организует категории объектов в памяти.

Но самым примечательным стало то, что паттерны активации мозга, зафиксированные при восприятии изображений у участников, соответствовали структурам, возникающим в моделях. Наиболее сильные совпадения наблюдались в областях мозга, отвечающих за зрительное восприятие и пространственную память: экстрастриарная кора, парагиппокампальная зона, ретроспленальная кора и область веретенообразного тела. Это указывает на то, что нейросети могут, пусть и в упрощённом виде, имитировать концептуальную архитектуру человеческого мышления.

Важным аспектом стало то, что используемые вложения были интерпретируемыми — каждое измерение можно было связать с конкретной чертой объекта: формой, цветом, функцией. Такая прозрачность делает модели удобными не только для технического применения, но и для когнитивных исследований.

Понимание того, как модели ИИ приходят к «мыслям», близким к человеческим, может перевернуть подходы к разработке новых интерфейсов человек-машина, систем поддержки принятия решений и даже технологий когнитивного расширения. Кроме того, это открывает двери к созданию ИИ, способного понимать контекст и категории мира не только статистически, но и семантически, как делает мозг.

Результаты этого исследования становятся частью растущего корпуса работ, доказывающих, что ИИ способен не только имитировать язык, но и развивать формы абстрактного мышления. И в этой гонке на приближение к человеческому разуму следующей вехой может стать ИИ, способный не только классифицировать, но и интуитивно осмысливать мир так, как это делаем мы.

Ссылка: «Представления концепций объектов, подобные человеческим, естественным образом возникают в многомодальных больших языковых моделях» DOI: [10.1038/s42256-025-01049-z](https://doi.org/10.1038/s42256-025-01049-z).