

Исследование компании Apple показало, что даже самые продвинутые нейросети, обученные для логического мышления, не справляются с действительно сложными задачами. Специалисты протестирували модели нового поколения в искусственно созданных головоломках и пришли к выводу: когда задание становится по-настоящему трудным, искусственный интеллект (ИИ) просто «сдается».

В работе рассматривались модели с усиленной логикой, известные как большие модели рассуждения (LRM), а также стандартные большие языковые модели (LLM). Их сравнивали в решении различных задач, от простых до более комплексных. На низком уровне сложности обычные языковые модели показали себя лучше — они были точнее и использовали меньше ресурсов. Когда задания становились чуть сложнее, преимущество переходило к LRM — за счет встроенных логических цепочек. Но как только сложность повышалась еще сильнее, обе категории моделей теряли способность справляться с задачей: точность падала до нуля.

Исследователи обращали внимание не только на результаты, но и на ход размышлений моделей. Выяснилось, что по мере усложнения задач нейросети начинали использовать более длинные логические цепочки. Но как только приближался предел их возможностей, модели неожиданно сокращали ход размышлений, даже если вычислительные ресурсы позволяли продолжать.

Даже при наличии правильных алгоритмов ИИ не мог корректно выполнять пошаговые инструкции. Кроме того, производительность сильно зависела от того, насколько похожими были головоломки на те, что использовались в обучении. Это говорит о том, что модели больше полагаются на знакомые шаблоны, чем на универсальные навыки логического мышления.