人間を超えたAI推論：Poetiq (Mix)モデルの衝撃

課題：「ARC-AGI-2」の壁

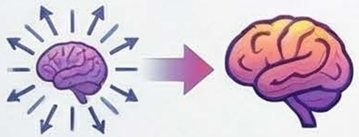
AIには超難関の推論テスト「ARC-AGI-2」

人間には簡単だが、従来のAIには極めて困難な課題で構成される。

GPTやClaudeなど主要LLMが軒並み低迷

トップクラスのモデルでさえ、正答率は20%未満に留まった。

計算量を増やしても性能が頭打ちになる「逆転現象」



単純に思考時間を延ばすだけでは、賢くならないことが判明した。

ブレークスルー：Poetiq (Mix)のアプローチ

正答率約65%を達成し、人間平均 (60%)を突破！

約65%

人間を超えるスコアを記録した、歴史上唯一のAIシステムとなった。

LLMを道具として使う革新的な「メタシステム」

解答を生成・検証・改良するサイクルを繰り返し、精度を高める。

賢い「推論戦略」が圧倒的な性能差の鍵

モデルの大きさ以上に、問題解決のプロセスが重要であることを証明した。

NotebookLM