AI科学推論能力の新基準：FrontierScienceベンチマークが示す現在地

Olympiadトラック：競技レベルの理論問題

国際科学オリンピック形式の短答問題100問で、高度な科学知識と推論力を測る。

構造化された知識 vs. 未知への探求

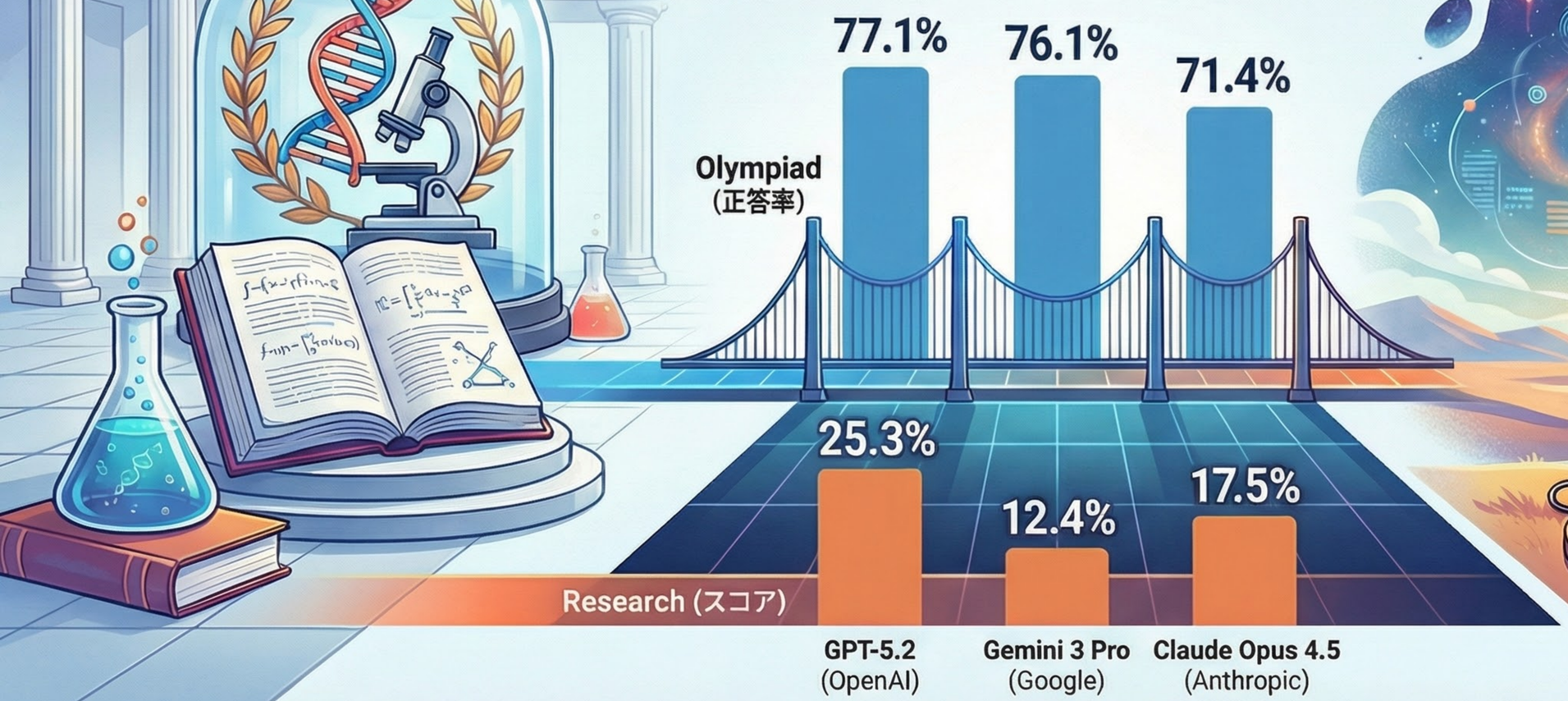
主要モデルの性能比較：明らかになった能力のギャップ

全モデル、Research課題で大幅にスコア低下

競技形式の問題では高得点を記録するが、オープンエンドな研究課題では軒並み苦戦。

GPT-5.2がResearchトラックで他を圧倒

他モデルの約2倍のスコアを記録し、より機械的な推論能力で優位性を示した。



Researchトラック：博士レベルの研究課題

博士号を持つ研究者が作成した自由記述式の問題60問で、実践的な研究遂行能力を測る。

Olympiadは既知の答えを導く能力を、Researchは研究プロセスを遂行する能力を評価する。