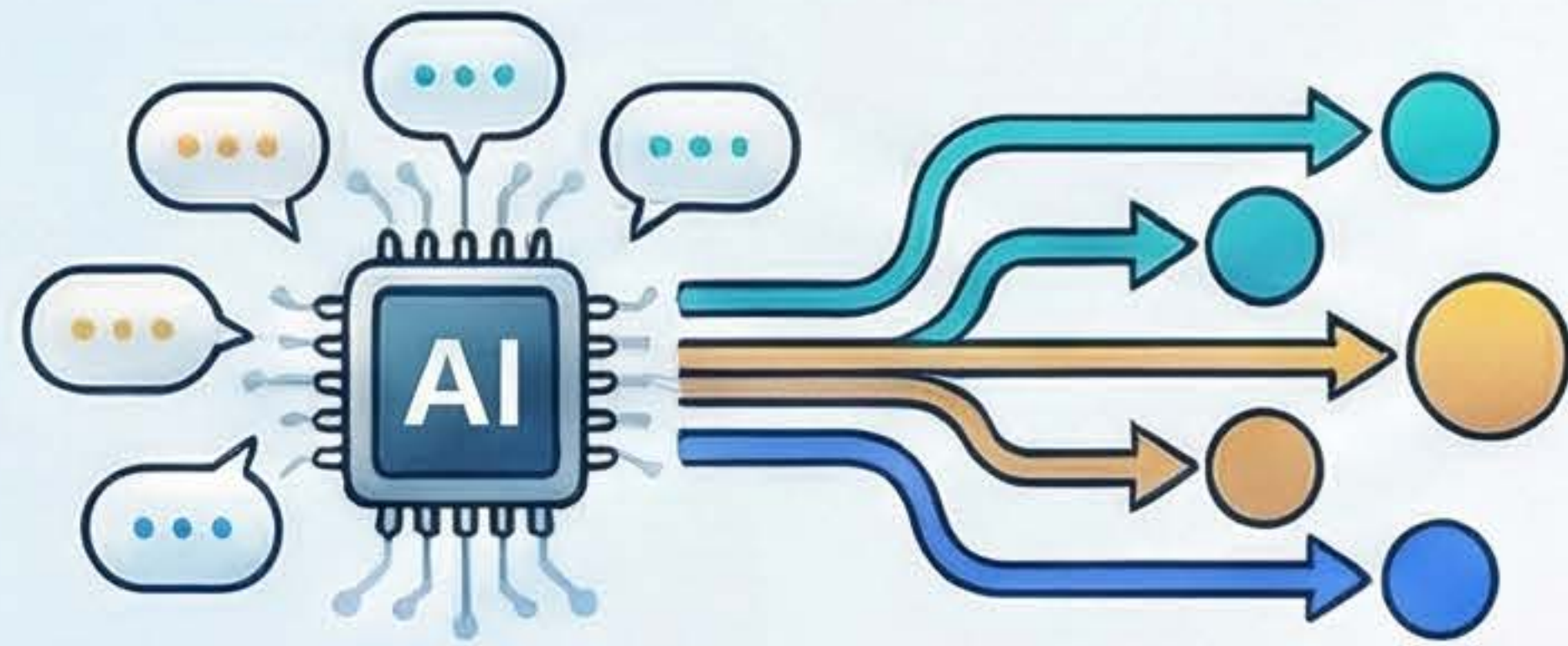


AGIレベル4への挑戦：Gemini 3 Deep Thinkの実力

驚異的な性能：博士レベルの推論力を証明

「じっくり考えるAI」が誕生

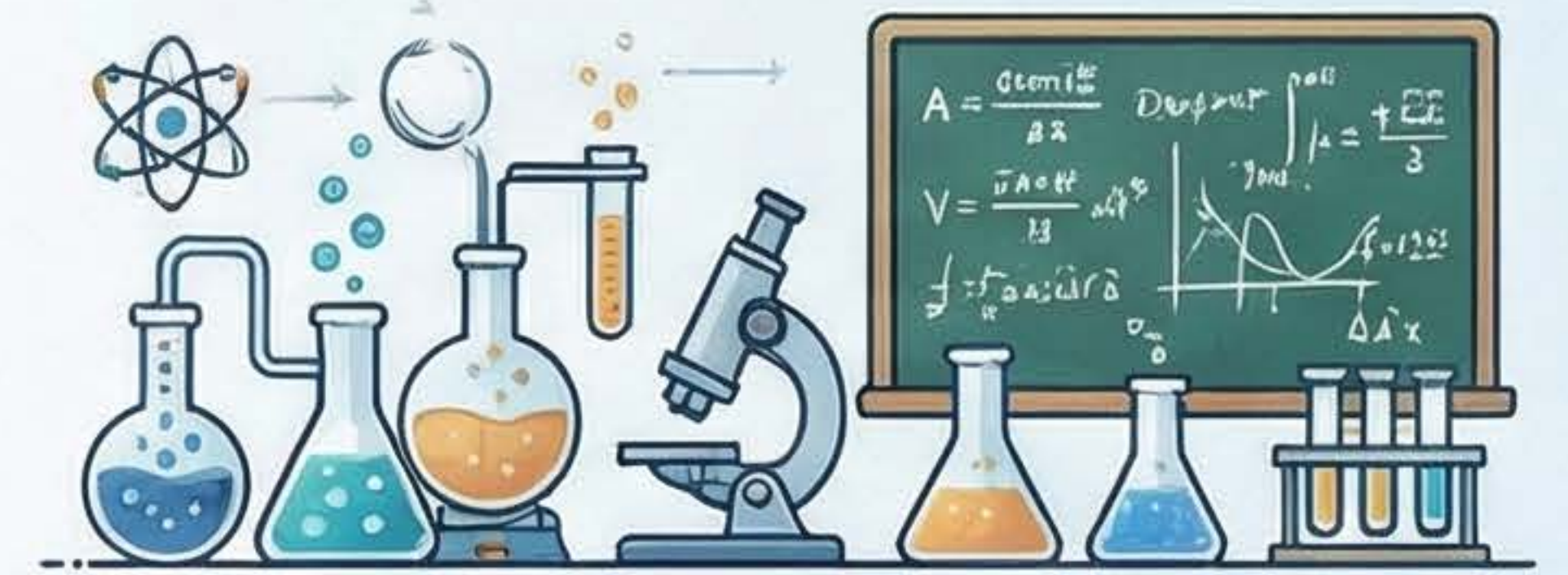


複数の推論経路を並行探索し、難問に対する思考の精度と深さを飛躍的に向上させました。



「発明」への最後の壁：残された課題

未知の科学的問題では依然として苦戦



博士課程レベルの物理課題(CritPt)では正解率が1割未満で、「自律的な科学者」には程遠い結果です。

汎用推論テストで他を圧倒



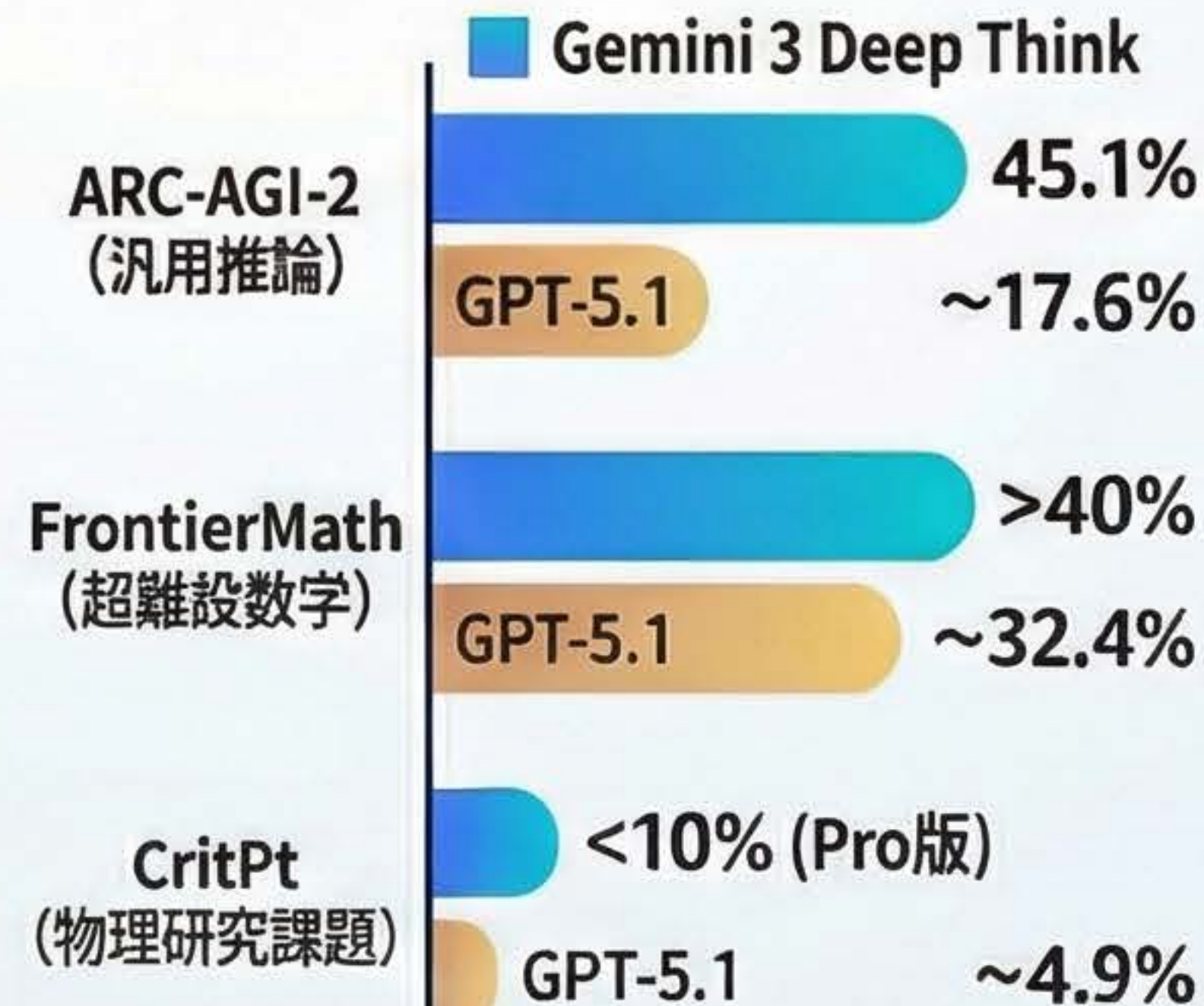
ARC-AGI-2ベンチマークで45.1%を記録し、懸念のGPT-5.1(約17.6%)を大きく引き越しました。

超難度の数学問題でトップに



FrontierMathで40%以上の正等率を達成し、数学分野での優位性を示しました。

ベンチマーク性能比較



専門家の見解：「あと一步」



「レベル4の入り口にいるが、完全到達は時期尚早」というのが全体的な評価です。

自律的な「発明」の具体例はまだない



現時点で、AI自身が独立して人類未織の科学的発見を成し遂げたという公式事例は報告されていません。