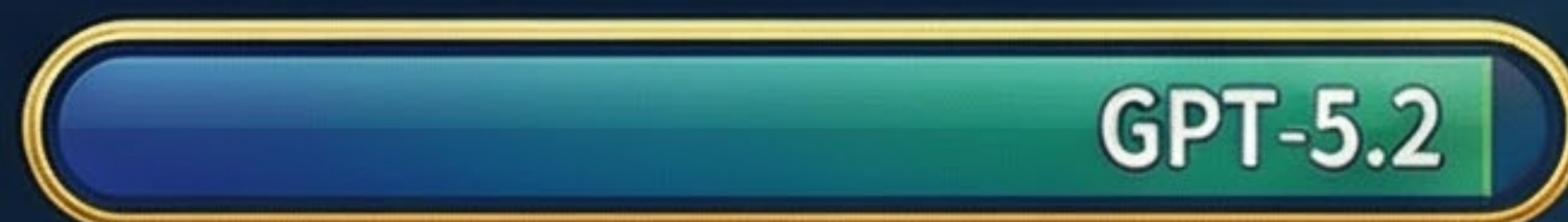


2026年 共通テスト AI検証レポート：GPT-5.2が9科目で満点達成と教育の変革

GPT-5.2が圧倒的970点を記録



文系科目1000点満点中970点（得点率97%）を達成し、受験生平均の58.1%を大きく上回る圧倒的な結果を示しました。

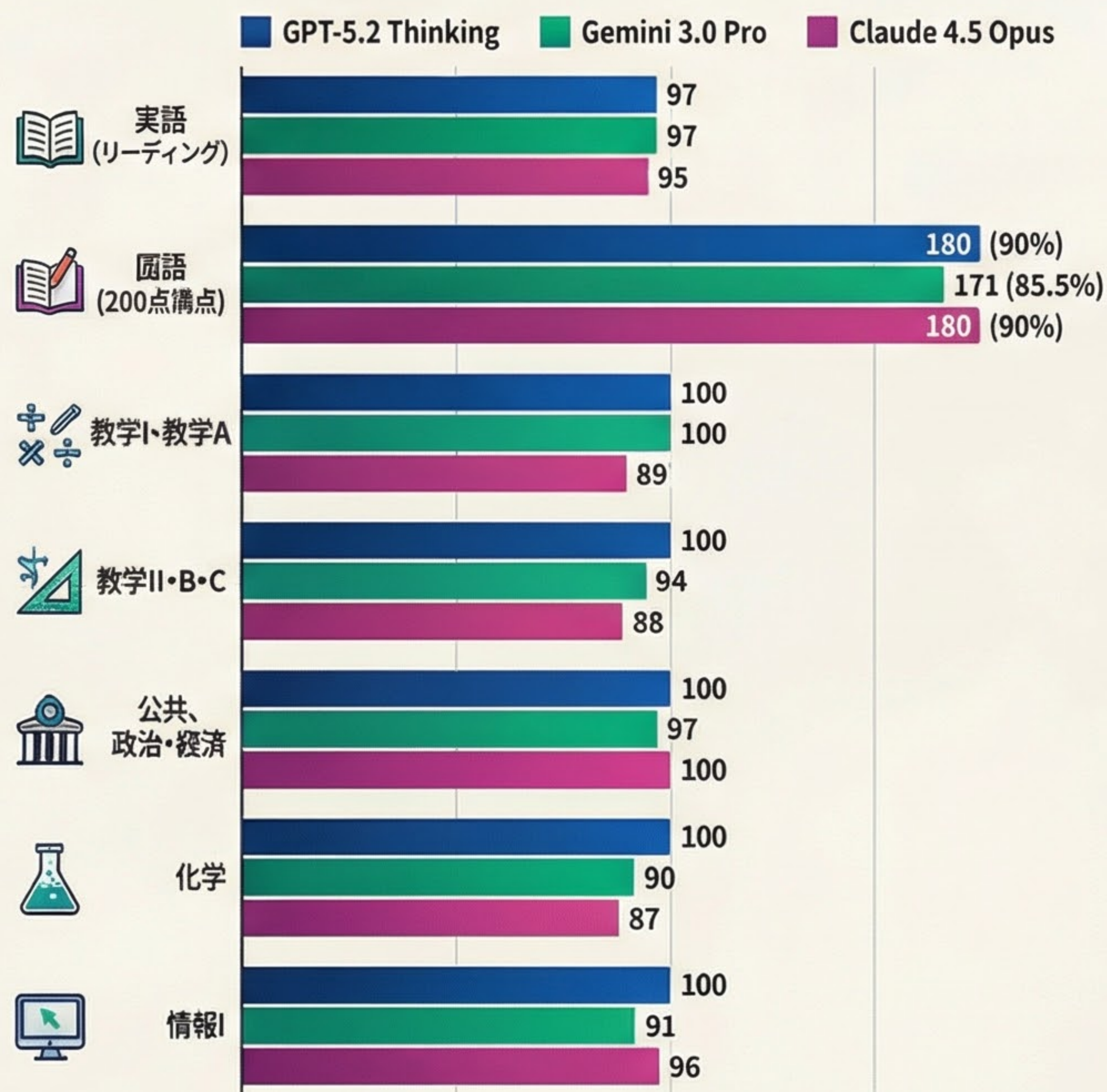
3モデルとも「東大合格ライン」を突破



15科目中9科目で満点を達成 (OpenAI)



主要3モデルの科目別得点比較



モデル別アプローチと特徴



GPT-5.2: バックグラウンドモードによる「熟考」
数分間の思考時間を確保する検定により、解答に約10分間の熟考を要しましたが、これを最終解に至るまで繰り返して正解を導きました。



Gemini 3.0: 圧倒的な「超高速処理」と推察能力
解答時間はわずか1時間45分、検定とグラフを視覚的に処理し、推察能力に優れた、他者の難問に唯一正解を導きました。



Claude 4.3: 人間らしい「文脈理解」
外見を的確に読み取り、面談などの文脈を捉えて6PTを満点満点を獲得し、最も人間らしい文章生成を示しました。

AIが直面した「壁」



視覚情報の抽象的な描き分けの限界
発着なスニグのパスと実態において、矢印の向きを的確に読み取る能力を失くすことが全モデルで発生しました。



人間の「微妙なニュアンス」へのバイアス
医師の問診文で、全人達の「病状」という言葉を、AIは通訳を学習データに偏る「病状」と一語で表現してしまいました。



地面のヒートマップ読み取り
地域における色の濃淡によるデータの読み取りは、表裏として全モデルで失敗しています。

教育・入試へのパラダイムシフト



入試の「奨励・奨励価値」と記述式導入
AIが豊富なマークから、入試問題の表裏分や価値や価値の記述式・満点満点への価値、およびAIによる記述式導入が実現されています。



不正行為対策の強化
スマホを使用しAI利用の監視も受け、試験場での電子機器の使用や不正行為の監視が強化されています。



AIと共生する「新たな学び」
単なる知識ではなく、AIを通じた問いを立てる力 (AIリテラシー) を重視する学習へのシフトが起っています。