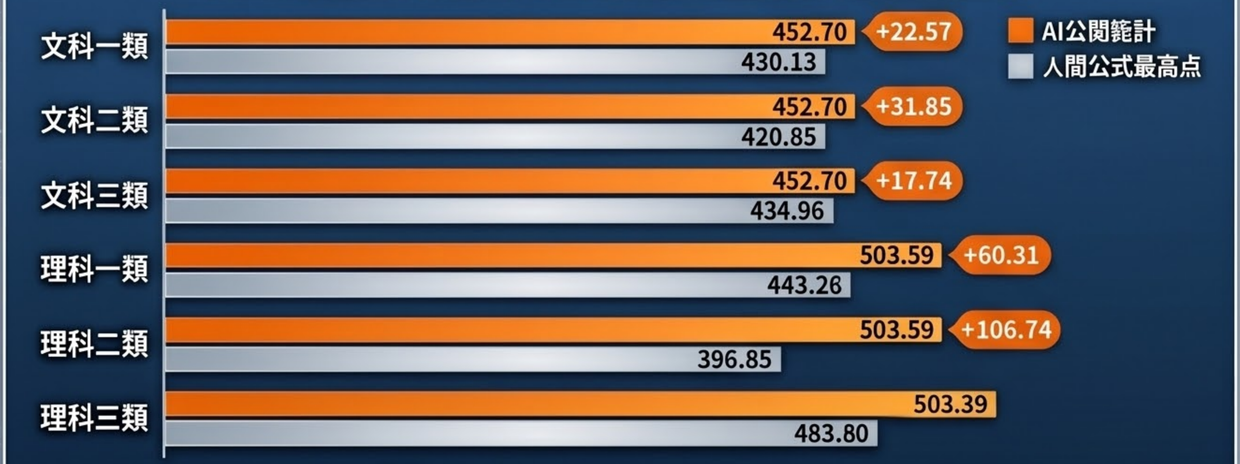
ChatGPT-5.2 Thinking: 徹底検証 東大・京大「首席合格」主張の監査レポート

公開データに基づくAI入試成績の妥当性評価と透明性の課題

東京大学：公開値による「首席超え」の確認

2026年度東大入試：AI vs 人間最高点比較（点数差強調）



理系数学で「120点満点」を達成

文・理全6科類で人間最高点を超過

検証の手法とプロセスの透明性

問題POFの画像化

API桂由で送信

共通プロンプトで解答生成

困示問題はPythonで処理

河合塾講師による採点

実験条件の「甘さ」への指摘

- 共通テストと二次試験の「合成スコア」
- 人間のような連続試験・疲労・時間制約は未考慮

再現性に必要な情報の不足

プロンプト全文、温度感、seed値、reasoning effortの詳細設定が非公開

京都大学：主張の揺れと不透明な点

参照データの差し替え問題

note公開時：2025年度最高点と比較

翌日プレスリリース：2026年度最高点に差し替え & AI得点穢修正

数値の固定性に課題

「全学部・学科」主張の検証困難

医学科以外は公式翁点表とLifePrompt掾舁尺廈が不一致

第三者による検証が不十分

京大医学科での高いパフォーマンス

AI得点 1176.38点 vs 人間最高点 1098.25点（+78.13点）

医学科単体では首庶超えの蓋然性が高い

AIの得意分野と依然として残る限界

理数系における圧倒的な強み

- 数学・化学満点
- 物理での高皮な理解
- 数理捶諞・レイアウト依存の誚み取り能力

論述・文硼理解における「賜さ」

- 世界史：論述構成力が弱く（15/60点の大纲れ）
- 国語：比嚙や皮向の処理
- 日本史：不自然な日本語袁現
- 物理における慣習の諓（実話鼯の侘考損的を使用）