ChatGPT

# ChatGPT-5.2 Thinking の東大・京大首席合格主張を検証する

## エグゼクティブサマリー

- 結論から言うと、**東京大学についての主張は、公開されている一次資料だけでもかなり強く裏づけられる**。LifePrompt 公開値では ChatGPT-5.2 Thinking の東大理系総計が 503.59 点、文系総計が 452.70 点で、東京大学の 2026 年度公式最高点は理科三類 453.60 点、文科三類 434.96 点だった。したがって、少なくとも「理三で首席相当」「数学満点」という骨格部分は、公開数値の突合だけで整合する。理科三類との差は **+49.99 点**で、報道の「50点上回る」は丸めとして妥当である。 [1]

- ただし、**京都大学については公開資料の整合性に問題がある**。LifePrompt の note 本文は「京大の 2026 年度合格者データは未公表」として 2025 年度医学科最高点 1105.87 点を参照し、AI 得点を 1176.25 点としている。一方、翌日のプレスリリースと業界報道では、参照値が 2026 年度医学科最高点 1098.25 点に差し替わり、AI 得点も 1176.38 点へ微修正されている。公開メッセージの更新自体はあり得るが、**同じ検証の主要数値が 24 時間足らずで変わっている**点は、厳密な監査では減点要素である。 [2]

- さらに重要なのは、**「京大全19学部・学科で首席超え」という広い主張は、現状の公開資料だけでは第三者が完全再現できない**ことだ。医学科については 1275 点満点で同尺度比較が可能だが、他学部・学科については LifePrompt 側が公開した表の換算尺度が大学公式の総点表と必ずしも同一に見えず、同じ土俵での学科別検証ができない。プレスリリース本文の前半では「ほぼ全学部・学科」と書かれ、後半では「全19学部・学科」と書かれており、**同一文書内でも射程が揺れている**。 [3]

- 手法面では、**再現可能性に必要な情報が足りない**。公表されているのは「PDF をページ画像化して API で送信」「全科目共通プロンプト」「Web 検索不使用」「図示問題では Python コードでグラフ出力」「河合塾講師採点」までであり、**正確なプロンプト全文、temperature、seed、reasoning effort、試行回数、失敗時リトライ、OCR の有無、相互採点一致率**は公開されていない。しかも 2025 年検証は「基本的にプロンプトなし・画像入力」、2026 年検証は「共通プロンプトあり・自動受験システム・図示で Python 使用」なので、**2025→2026 の伸びをそのままモデル進化だけに帰属するのは危険**である。 [4]

- よって、本件を最も厳密に要約すると、**「東大については公開値ベースで首席超えをほぼ確認できる。京大医学科も高得点は確認できる。しかし京大全学部・学科への一般化と、採点の厳密な再現性までは現公開資料では証明できない」**となる。報道見出しとしては強いが、学術的・計量的な意味での“確証”には、**生答案、採点票、換算式、モデル設定の完全公開**がなお必要である。 [5]

## 一次情報の整理と時系列

本件の一次情報は、LifePrompt [6] の note 記事とプレスリリース、東京大学 [7] および 京都大学 [8] の公式入試ページ・合格者成績表、採点協力先として明記された 河合塾 [9] の講師コメント、そして OpenAI [10] のモデル仕様書・リリースノートで構成される。比較対象として Google [11] と Anthropic [12] のモデルも登場するが、本報告の焦点は ChatGPT-5.2 Thinking の主張検証に置く。なお、「チャッピー」は一部報道で使われた通称であり、OpenAI 公式文書では ChatGPT-5.2 Thinking / gpt-5.2 と表記される。 [13]

ユーザーの依頼文では「2024–2025 の東大・京大入試」となっているが、**首席合格主張そのものは 2026 年 2 月実施の入試を対象にしたもの**である。他方、大学側が問題・出題意図・解答等を体系的に公開している最新年は、本監査時点で東大が 2025 年度、京大が 2025 年度であり、2024–2025 年は「公式問題・解答の入手が比較的しやすい基準年」と位置づけるのが正確である。 14

公開情報を時系列で並べると、論点は次のように整理できる。 15

```
timeline
    title 公開情報の時系列
    2025-03 : LifePrompt が 2025 年度東大入試で ChatGPT o1 を検証
    2026-02 : 東大・京大の 2026 年度二次試験を実施
    2026-04-27 : LifePrompt note 公開
               : 京大医学科は 2025 年度最高点と比較
    2026-04-28 : LifePrompt プレスリリース公開
               : 京大医学科は 2026 年度最高点と比較
    2026-04-29 : AI Watch がプレスリリース内容を報道
    監査時点 : 東大 2026 公式最高点と京大 2026 公式最高点は大学サイトで確認可能
```

ここで最も重要なのは、**京都大学に関する比較基準が 2025 年度版から 2026 年度版へ差し替わっている**ことである。これは単なる細部ではなく、主張の強さに直結する。医学科の差分は note 版で +70.38 点、PR/報道版で +78.13 点となり、更新後の方が強い主張になっている。 2

## 手法の再構成

公表情報から再構成できる 2026 年検証フローはかなり明確だが、肝心な再現パラメータが抜けている。判明しているのは、入試問題 PDF をページごとに画像化し、チャット UI ではなく API 経由で各モデルに送り、共通プロンプトで解答させ、Web 検索を使わず、記述答案は河合塾講師が採点したことまでである。さらに「図示せよ」に対しては Python コードでグラフを出力する処理が組み込まれていた。 16

```
flowchart LR
    A[入試問題PDF] --> B[ページ単位で画像化]
    B --> C[API 経由でモデルへ送信]
    C --> D[共通プロンプトで解答生成]
    D --> E[必要に応じて図示問題は Python でグラフ化]
    E --> F[出力答案を取得]
    F --> G[河合塾講師が記述答案を採点]
    G --> H[共通テスト結果と合算して総計化]
```

このフローで監査上の要点になるのは、**総計が「共通テスト実験」と「二次試験実験」の合算値である**点だ。LifePrompt 本文は、今年 1 月の共通テスト 2026 検証結果を前提に、そこへ二次試験の得点を載せて総計を作っている。つまり、AI が人間受験生のように連続した日程・疲労・時間制約で一次から二次まで完走したわけではなく、**別実験の高得点を組み合わせた“合成入試成績”**である。総計比較として無効とは言わないが、「首席合格」という言葉の印象より条件は甘い。 17

以下が、公開情報から言えることと言えないことの整理である。 18

| 項目 | 公表されていること | 監査上の評価 |
|---|---|---|
| 入力形式 | PDF をページ単位で画像化し API 送信 | 明確 |
| OCR の扱い | 画像入力は明言。外部 OCR を使ったか、モデルのネイティブ視覚処理だけかは未公表 | 不明 |
| プロンプト | 全科目共通。高校教養課程までの知識・詳細な思考・LaTeX 出力などを指定 | **全文未公表** |
| Web 利用 | ブラウジング不使用 | 明確 |
| 外部ツール | 図示問題では Python によりグラフ出力 | **一部ツール利用あり** |
| 時間制約 | 2025 は各科目おおむね試験時間内。2026 も一部科目で高速処理が示される | 全科目の完全ログは未公表 |
| モデル設定 | OpenAI 側では reasoning.effort など複数設定が可能 | **使用設定未公表** |
| 試行回数 | 記載なし | 不明 |
| 失敗時再実行 | 記載なし | 不明 |
| 採点 | 河合塾講師が人間と同基準で採点 | 重要だが、**採点票・一致率未公表** |

さらに、**2025 年検証と 2026 年検証は方法が違う**。2025 年は基本的に「スクリーンショットをそのまま入力」「追加プロンプトなし」で、国語だけ縦書きのため文字起こしを行っていた。2026 年は「共通プロンプトあり」「自動受験システム」「図示問題で Python 対応」である。つまり、2025 から 2026 への点数上昇には、モデル能力の向上だけでなく**実験設計の最適化**も混ざっている。OpenAI 自身も GPT-5.2 の prompting guide で、prompt の設計が精度・規律・グラウンディングを左右すると明記している。 [19]

加えて、OpenAI の公式仕様では GPT-5.2 に 400k コンテキスト、128k 出力上限、2025-08-31 の知識カットオフ、そして none/low/medium/high/xhigh の reasoning.effort 設定がある。ChatGPT 側では 2026 年 1〜2 月に GPT-5.2 Thinking の思考時間設定が何度か変更されており、**どの時点・どの設定で実験したか**が分からないと厳密再現はできない。 [20]

## 公開値の照合結果

### 東京大学

東京大学については、大学公式の 2026 年度最高点と LifePrompt 側の 550 点満点総計が同尺度で並ぶため、数値監査がしやすい。次の表は、大学公式の各科類最高点と、LifePrompt が公開した ChatGPT-5.2 Thinking の文系総計・理系総計を照合したものだ。 [21]

| 科類 | 公式最高点 2026 | AI公開総計 | 差分 |
|---|---|---|---|
| 文科一類 | 430.13 | 452.70 | +22.57 |
| 文科二類 | 420.85 | 452.70 | +31.85 |
| 文科三類 | 434.96 | 452.70 | +17.74 |

| 科類 | 公式最高点 2026 | AI公開総計 | 差分 |
|---|---|---|---|
| 理科一類 | 443.28 | 503.59 | +60.31 |
| 理科二類 | 396.85 | 503.59 | +106.74 |
| 理科三類 | 453.60 | 503.59 | +49.99 |

この表から分かる通り、**東大全6科類で最高点を上回った**という主張は、少なくとも公開値同士の突合では成立する。特に理科三類については 503.59−453.60＝49.99 点で、報道の「50点上」は誇張ではなく丸めである。 [1]

公開情報を図にすると、東大側は非常に分かりやすい。 [21]

```
xychart-beta
    title "東大 2026 総合点の比較"
    x-axis ["文一最高","文二最高","文三最高","AI文系","理一最高","理二最高","理三最高","AI理系"]
    y-axis "550点満点" 0 --> 550
    bar [430.13,420.85,434.96,452.70,443.28,396.85,453.60,503.59]
```

一方、2025 年の LifePrompt 検証では、ChatGPT o1 の東大総計は文系 379 点、理系 374 点で、**合格最低点は超えるが首席点には遠い**という位置にあった。東京大学公式 2025 年最高点に照らすと、理科三類では −99.06 点、文科三類では −41.03 点である。2026 年の 503.59 点は、2025 年理系総計 374 点から **+129.59 点**の上昇になる。だが前節で見た通り、ここにはモデル世代更新だけでなく、プロンプトと実装の変更も乗っている。 [22]

科目別では、ChatGPT-5.2 Thinking の東大公開値は、英語 108/120、理系国語 52/80、理系数学 120/120、物理 59/60、化学 58/60、生物 56/60、日本史 29/60、世界史 15/60、地理 49/60 である。**AI 側の科目別点は公開されているが、大学公式は人間首席の科目別点を公表していない**ため、「AI が人間首席の科目別点を超えたか」は一次資料だけでは判定できない。ここで確認できるのは、数学満点と理科系高得点、そして世界史の大崩れである。 [23]

## 京都大学

京都大学は事情が複雑で、**同尺度比較が公開情報だけで確実にできるのは医学科だけ**である。note 本文では「2026 年度データは未公表」として 2025 年の医学科最高点 1105.87 点を参照し、ChatGPT-5.2 Thinking を 1176.25 点とした。一方、翌日のプレスリリースと報道では、2026 年医学科最高点 1098.25 点に対して 1176.38 点とされている。 [24]

| 公開版 | AI得点 | 比較に使った人間側最高点 | 差分 | コメント |
|---|---|---|---|---|
| LifePrompt note | 1176.25 | 2025 医学科 1105.87 | +70.38 | 本文では「2026年度データ未公表」と説明 |
| PR / 業界報道 | 1176.38 | 2026 医学科 1098.25 | +78.13 | 後続版で値が更新 |

この差分表から読み取るべきなのは、「京大医学科でも AI が非常に高い」という大筋ではなく、**公表版ごとに比較対象と AI 得点が動いている**という事実の方である。医学科については後続版の 1176.38 vs 1098.25 を採れば確かに首席超えだが、主張の根拠が固定的ではない。 [25]

さらに大きな問題は、**プレスリリースの前半では「京大のほぼ全学部・学科」、後半では「ChatGPT は全19学部・学科」と書かれている**点だ。医学科については本文で 1275 点満点比較が示されるが、他の 18 単位については同尺度の総点表が公開されていない。公開画像には学部換算値があるものの、大学公式の総点表と直接突合できる換算式が本文中に説明されておらず、第三者監査には不十分である。したがって、**京都大学についての最も強い結論は「医学科の首席超えは公開値で確認可能だが、全19学部・学科首席超えは現状未検証」**である。 [26]

科目別の AI 公開値だけを見ると、京都大学では英語 133/150、理系国語 64/100、理系数学 200/200、物理 95/100、化学 100/100、生物 98/100、日本史 87/100、世界史 78/100、地理 89/100 と、高得点が並ぶ。ここでも大学公式は人間首席の科目別点を出していないため、科目別の対人比較は不能である。**京大の「数学満点」「化学満点」は、河合塾講師採点による AI 側点数としては確認できるが、人間首席の科目点との直接比較ではない。** [27]

## 採点の妥当性と再現性

この検証の強みは、記述式答案を予備校講師が見ていることだ。東大・京大とも、大学自身が記述問題で重視すると明言しているのは、正答そのものだけでなく**論理的な道筋、表現の明晰さ、条件把握**である。東京大学の数学出題意図は「数学的に思考する力」「数学的に表現する力」「総合的な数学力」を掲げ、京都大学の数学出題意図も「値を求める問題でも答えに至る論理的道筋を見る」「必要条件・十分条件に配慮した適切な表現を見る」と書いている。したがって、機械採点ではなく人間採点を入れたこと自体は合理的である。 [28]

ただし、**採点の妥当性が高いことと、採点の再現性が高いことは別問題**である。LifePrompt が公開した方法論には、二重採点の有無、主査・副査による調整、採点前のブラインド化、採点者間一致率、異議申立手順が含まれていない。京都大学の公式「出題意図等」は、標準的な解答例はそこに示す表記に限られず、問い合わせにも応じないとしている。これは大学入試の記述採点が一定の幅を持つことを示すが、同時に第三者が完全再現しにくいことも意味する。 [29]

公開サンプルをみると、採点のブレが入り得る箇所は実際に存在する。たとえば東大世界史では、GPT の短答式解答に余計な説明文が混入し、LifePrompt 本文自身が「これを 0 点と見ると 15 点、甘く見ても 35 点にとどまる」と説明している。つまり、**厳格採点と寛容採点で合計点が大きく動き得る**例が、公開された範囲だけでも確認できる。 [30]

物理でも同様だ。公開コメントでは、GPT が日本の物理慣習では正とする量に英語圏の符号規約を持ち込み、設定とずれた符号を出したとされる。これは「物理理解がない」よりも、「試験文脈に従うこと」と「既有の理論的慣習を優先すること」の競合で起きるエラーであり、**部分点のつけ方は採点者の哲学に依存しやすい**。 [31]

数学満点主張の扱いも慎重であるべきだ。公開された講師コメントでは、GPT と Gemini は東大・京大数学で満点とされた一方、Claude には「厳密性に欠ける」「一歩目までしか点が入らない」「題意把握自体ができていない」などの評価が付いている。つまり採点は、単に最終答が合っているかではなく、書き方も強く見ている。その意味で満点は重い評価だが、**生答案全体と採点票が公開されていない以上、第三者がその満点を確認することはできない**。 [32]

総合すると、採点は「雑」ではないが、**監査可能性はまだ低い**。本件の点数は「河合塾講師が見たらこう採点した」という有力な専門家判断ではあるものの、学術論文レベルで求められる inter-rater reliability の提示がない以上、誤差帯を見積もれない。したがって、「理三で 50 点上」「数学満点」は**確定的事実というより、現行公開ルールの下で最も信頼できる推定値**として扱うのが適切だ。 [29]

## モデル能力と限界の読み方

能力面では、ChatGPT-5.2 Thinking が**数理・理科で非常に強い**ことは、公開答案分析からも、OpenAI の公式説明からも整合的だ。OpenAI は GPT-5.2 Thinking を複雑な知識業務向けのフロンティアモデルと位置づけ、長文理解、ツール活用、画像認識、長時間タスク処理を強みとしている。また FrontierMath では Python ツール有効・最大 reasoning effort で高成績を報告している。LifePrompt 側でも東大理系数学 120/120、京大理系数学 200/200、京大化学 100/100 が公開された。 [33]

ただし、それは「万能」ではない。LifePrompt の科目別分析では、英語では GPT が下線部の位置や複数行にまたがる範囲を最も正確に読み取り、図表・構造式・家系図でも Claude より安定したという評価がある一方、世界史では論述構成力が弱く、国語では比喩・韜晦・皮肉の処理が苦しく、日本史では日本語として不自然な表現や冗長さが目立った。**強いのは数理推論・レイアウト依存の読取りであって、日本語の高度な含意処理や答案欄への出力制御ではまだ脆い**。 [34]

この限界は、まさに大学側の出題意図とも対応している。東京大学の外国語出題意図は、論理的思考力と文脈に応じた判断力を要件に挙げ、京都大学の国語出題意図は、複雑な議論や比喩表現を「明晰で統御された表現」でまとめる力を問う。AI が理系で圧倒し、論述系で取りこぼしたのは、試験が測っている能力のどこまでが現世代 AI に自動化可能かをかなり素直に映している。 [35]

学習データ汚染については、**2026 年本番問題の丸暗記可能性は高くないが、近年の過去問や類題への接触可能性は十分ある**というのが妥当な評価だ。OpenAI 公式の GPT-5.2 の知識カットオフは 2025-08-31 であり、2026 年 2 月実施の東大・京大入試そのものは直接学習されていない可能性が高い。一方で、東大 2025 年問題と解答等、京大 2025 年の試験問題・出題意図等は大学公式サイトで公開されており、2025 年 8 月以前にウェブ上にあった過去問・予備校解説・類題集は訓練データに含まれていても不思議ではない。したがって、**2026 の問題文そのものの漏洩を疑うより、過去問文化に深く浸されたモデルが高い移転性能を示した**と見る方が自然である。 [36]

また、2026 年の比較には**モデルドリフト**の問題もある。ChatGPT 側の GPT-5.2 Thinking は 2026 年 1 月と 2 月に思考時間設定が変更されており、2025 年検証も 2026 年検証も、実験時点の API 挙動に依存している。さらに 2025 年は no-prompt に近い条件、2026 年は prompt あり・Python ありである。したがって、「1 年で数学 38 点から満点」という絵は印象的だが、厳密には**モデル能力・プロンプト制度・ツール利用・視覚前処理の複合改善**として読むべきである。 [37]

## 含意と提言

この結果がすぐに「大学入試は AI に敗北した」を意味するわけではない。実際の東大・京大入試は、決まった時間、紙の問題冊子、答案冊子、会場監督という条件で実施される。公開された AI 実験は強力な比較材料ではあるが、共通テストと二次試験を別実験で合成し、図示問題では Python を使い、疲労や連日拘束も負わない。したがって、**現時点で脅かされているのは「会場筆記試験」そのものより、試験が測る能力の一部が外部化・自動化できるという事実**である。 [38]

大学側への実務的な提言は三つある。第一に、**採点基準の説明責任を強める**ことだ。少なくとも匿名化した模範答案・部分点例・典型減点例を増やせば、AI 検証にも人間受験にも有益である。第二に、**一発の静的答案ではなく、途中の判断理由を口頭で守らせる試験**を増やすことだ。短い口頭試問、面接型フォローアップ、複数資料のその場統合は、現在の AI が得意な「定型化された高難度問題」とは別の能力を測りやすい。第三に、**レイアウト認識や高速計算より、未知資料の咀嚼・批判・選択を重視する設問設計**へ寄せることである。これは東大・京大が公式に掲げる「論理的表現」「文脈判断」「複雑な議論の再構成」とも整合する。 [39]

予備校・報道機関・AI 企業への提言もある。**「首席合格」級の見出しを使うなら、少なくとも次の四点は同時公開すべきだ**。生答案全文、採点済み答案または採点票、換算式、モデル設定である。これがなければ、話題性は高くても、第三者は数値の真偽を十分に監査できない。とくに京都大学のように、本文・PR・報道で比較基準が動いた案件では、その必要性がさらに増す。 40

教育政策上は、過剰反応も禁物だ。AI が東大理三級の問題で高得点を取っても、それは**研究者・医師・法律家・歴史家としての総合適性**をそのまま意味しない。公開サンプルでも、物理では慣習依存、世界史では論理接続の弱さ、国語では比喩理解と日本語運用の弱さが出ている。政策議論は「AI が入試を解けたか」よりも、「どの能力が自動化され、どの能力がなお人間中心なのか」を分解して進めるべきである。 41

## 優先ソース一覧

- **最優先の一次資料**
  LifePrompt の note 本文。2026 年版は手法、科目別分析、講師コメント、京大医学科比較の初期値を含み、2025 年版は no-prompt 条件・処理時間・前年比較の基準になる。 42

- **大学公式資料**
  東京大学の 2026 年合格者成績表、2025 年過去問ページ、2025 年解答等の公表、京都大学の 2026/2025 合格者最高点表、2025 年試験問題および出題意図等。数値照合ではここが土台である。 43

- **公表の更新差分を追う資料**
  LifePrompt のプレスリリースと、その内容を写した業界報道。note と PR/報道のずれ、京都大学比較基準の差分、主張のスコープの揺れを確認するために必須。 44

- **モデル仕様の一次資料**
  OpenAI の GPT-5.2 紹介ページ、API モデル仕様、ChatGPT リリースノート、Prompting Guide。能力だけでなく、思考時間設定の変更や reasoning.effort の存在まで確認できる。 45

- **科目別の公開サンプル答案**
  世界史、物理、数学の抜粋画像。採点の揺れや、AI 特有の「余計に書く」「慣習に引きずられる」「厳密性が足りない」を具体例で確認できる。 46

- **補助的報道**
  日本経済新聞 47 の見出し紹介や共同通信系配信は「社会的にどう受け止められたか」を追うには有用だが、検証の核はあくまで大学公式資料、LifePrompt 公開物、OpenAI 公式文書に置くべきである。 48

---

1 5 21 43 https://todai.info/juken/data/2026/400281996.pdf
https://todai.info/juken/data/2026/400281996.pdf

2 4 6 9 11 13 16 17 18 24 30 31 32 34 40 41 42 47 https://note.com/lifeprompt/n/n85674c186fbc
https://note.com/lifeprompt/n/n85674c186fbc

3 10 12 14 25 26 29 44 https://prtimes.jp/main/html/rd/p/000000003.000136448.html
https://prtimes.jp/main/html/rd/p/000000003.000136448.html

7 33 https://openai.com/ja-JP/index/introducing-gpt-5-2/
https://openai.com/ja-JP/index/introducing-gpt-5-2/

8 22 https://www.u-tokyo.ac.jp/content/400200766.pdf
https://www.u-tokyo.ac.jp/content/400200766.pdf

15 19 37 https://note.com/lifeprompt/n/n0078de2ef36b
https://note.com/lifeprompt/n/n0078de2ef36b

20 36 45 https://developers.openai.com/api/docs/models/gpt-5.2
https://developers.openai.com/api/docs/models/gpt-5.2

23 https://assets.st-note.com/img/1777179054-g8ifQr7UczTxOVj0ZqdH5KEn.png?width=1200
https://assets.st-note.com/img/1777179054-g8ifQr7UczTxOVj0ZqdH5KEn.png?width=1200

27 https://assets.st-note.com/img/1777179698-TlPUQ8F56mLIxuraNjpiJ1by.png?width=1200
https://assets.st-note.com/img/1777179698-TlPUQ8F56mLIxuraNjpiJ1by.png?width=1200

28 39 https://www.u-tokyo.ac.jp/content/400239145.pdf
https://www.u-tokyo.ac.jp/content/400239145.pdf

35 https://www.u-tokyo.ac.jp/content/400239234.pdf
https://www.u-tokyo.ac.jp/content/400239234.pdf

38 https://www.u-tokyo.ac.jp/content/400239118.pdf
https://www.u-tokyo.ac.jp/content/400239118.pdf

46 https://assets.st-note.com/img/1775889957-dXUqDMRwl3HGtWIbOLSPpvC8.png?width=1200
https://assets.st-note.com/img/1775889957-dXUqDMRwl3HGtWIbOLSPpvC8.png?width=1200

48 https://x.com/nikkei/status/2048534188953915739
https://x.com/nikkei/status/2048534188953915739