

GPT-5 Pro が Mensa Norway 公開 IQ テストで 148 点？公開版スコアの裏側と解釈ポイント



Genspark

Aug 13, 2025

要旨（エグゼクティブサマリー）

- 結論: 2025 年 8 月 12 日、AI 監視サイト「Tracking AI」の IQ ページにおいて、OpenAI の GPT-5 Pro が Mensa Norway の公開オンライン IQ テストで「148」と表示されていることは画面証拠とともに確認できました。ページ自体の更新時刻も同日（EDT）である旨が示されています。ただし、これは“公開・練習用”の非監督・非公式テストであり、問題の公開性や言語化提示などにより AI スコアが過大化する「汚染（contamination）」リスクがあるため、汎用的な“知能”の証明としては限定的です。Tracking AI 自身の「オフライン（非公開）テスト」では、同モデルは Mensa Norway よりも低いスコア（116）で表示され、指標の差異が確認できます。[TrackingAI1](#) [TrackingAI2](#)
- 重要な補足: Mensa Norway のオンラインテストは公式入会審査の代替ではなく、練習・参考用の無料オンライン版。Mensa International も公式サイトで“練習チャレンジ”と位置づけており、信頼区間や監督条件が整った医療・教育現場向けの専門 IQ 検査とは異なるため、AI 能力の一般化可能な指標としては慎重な解釈が求められます。[test.mensa.no3](#) [Mensa International4](#)
- モデルの存在と命名: OpenAI は 2025 年 8 月に GPT-5 と GPT-5 Pro を正式発表。Pro は拡張推論（長時間思考・高難度タスク）を特徴とする上位バリエーションで、ChatGPT の Pro プランや企業プラン向けに提供されると案内されています。OpenAI の発表文は IQ ではなく各種ベンチマーク（例: GPQA）を強調しており、Mensa Norway の“148”は OpenAI の一次発表ではなく、Tracking AI 等の外部検証に属するデータです。[OpenAI5](#) [OpenAI6](#) [CNBC7](#)

証拠（一次情報の可視化）

- Tracking AI の当該ページ（2025 年 8 月 12 日 EDT 更新）スクリーンショットに、ガウス分布上の「IQ Test Results」図表と「Rank by Test Source（Mensa Norway vs

Offline)」バーが表示され、OpenAI GPT-5 Pro“148”およびオフラインではより低い数値（約 116）が確認できます。掲載は毎週、複数の Verbal/ Vision モデルを定期テストするという運用説明とともに示されています。TrackingAI¹

今回の主張は何か（整理）

- 主張の核: 「ChatGPT-5 Pro（=OpenAI の GPT-5 Pro）が、Mensa Norway オンライン IQ テストで 148 を記録」との公表があった。
- 確認状況: Tracking AI の最新 IQ ページにおける図表・更新時刻・モデルラベルに基づき、GPT-5 Pro= 148 の掲示を確認。加えて「Rank by Test Source」で Mensa Norway とオフラインテストの差が可視化されています。TrackingAI¹

Mensa Norway のオンラインテストとは（性質と制約）

- 位置づけ: Mensa Norway が提供する 35 問 25 分の視覚マトリクステスト（公開オンライン版）は、認知能力の“目安”であり、専門家が監督し、標準化・バッテリー化された臨床的な IQ 検査の代替ではありません（公式入会審査にも代用不可）。平均は 100、測定範囲は概ね 85–145 と案内されています。test.mensa.no³
- Mensa International の案内: “Mensa IQ Challenge”は難易度が高い練習チャレンジとして提供され、同一人が複数回受ければスコアが揺れうること、繰り返し挑戦可能であることが明記されています。監督・再現性・標準化の点で、正式な臨床 IQ テストとは異なる性質です。Mensa International⁴

Tracking AI のテスト方法（Verbal 化と Vision、再質問ルール）

- 実施形式: Tracking AI は“Vision モデルには画像そのものを提示、Verbal モデルには問題を言語化して提示（verbalized）”と説明。回答拒否が起きた場合の再質問（最大 10 回）ルールなど、運用面の注記も記載されています。これは同一問題でも表現の違いやテキスト化の手順が成績を左右し得ることを意味し、AI 評価系としての系統誤差が混入しうる点に注意が必要です。TrackingAI²
- オフラインテストの併用: Tracking AI は“Offline（非公開）”の項目も導入し、公開問題の学習混入（いわゆる“汚染”）影響を緩和する意図がうかがえます。実際、同サイトの図表では「Mensa Norway で高得点、Offline では低め」という乖離が GPT-5 Pro（Vision）でも見て取れます。TrackingAI¹

外部の言及・二次資料

- BinaryVerseAI のまとめ(2025 年 8 月 12 日付更新): 公開 Mensa Norway で GPT-5 Pro が“天才域（ ≥ 140 ）超え”とするリーダーボードを掲示し、表形式で Mensa/Offline 双方のスコアを併記。GPT-5 Pro は Mensa=148、Offline=116 と記

載しつつ、公開問題の“学習データへの流入”や言語化手順の差が精度に影響する“汚染”問題を詳述しています。これは“公開セットでは高く、非公開では低め”という傾向を裏づける報告です。BinaryVerseAI⁸

- Reddit の議論: GPT-5 Pro の“148”画像が共有される一方、モデルルータの不具合、Pro 版(有料/一部アクセス)のテスト再現性、IQ という指標の AI 評価への妥当性、問題リークの可能性など、懐疑や留保が多数表明されています。これは“話題性はあるが、評価指標としての限界が目立つ”というコミュニティ感覚を示します。Reddit⁹

参考図版 (BinaryVerseAI 掲載イメージ)

- AI IQ テストの歴史や言語化プロンプトの課題を図示 (公開問題がウェブ上に漂う状況や、言語化の語順・用語選択が正答率を数%単位で左右し得る旨)。評価設計の難しさを視覚的に説明しています。BinaryVerseAI⁸

画像:

BinaryVerseAI⁸

BinaryVerseAI⁸

OpenAI の公式アナウンス (モデルの実在と“IQ”の扱い)

- GPT-5 と GPT-5 Pro の公式発表: OpenAI は GPT-5 を「これまでで最も高性能な統一モデル」として発表し、Pro 版は拡張推論で難問に強い上位バリエーションとして提供。強調されるのは GPQA のような学術・専門ベンチマークであり、Mensa や“IQ 148”は OpenAI 公式の主張ではありません (外部サイト側の評価)。OpenAI⁵
- 提供スキームと媒体報道: GPT-5 は広く ChatGPT に展開、Pro/Team/Enterprise/Edu に段階的に提供。CNBC などローンチを報道しています。とはいえ、OpenAI 発表文に「Mensa IQ」の直接的な記載はありません。OpenAI⁶ CNBC⁷

“148”が意味すること／しないこと

- 意味すること:
 - Mensa Norway の公開オンライン版という特定の設定では、GPT-5 Pro (Vision) が非常に高い正答率を示し、Tracking AI の可視化では IQ スケール換算で“148”と表示されうる、という事実関係。ページ更新時刻を含め、当日更新で掲示されていました。TrackingAI¹
- 意味しないこと:
 - “IQ 148 = 人間の汎用知能 148 に匹敵”を意味するわけではありません。テ

ストは公開・練習用で、問題の事前露出、言語化プロンプトの恣意性、回答再試行、非監督、モデルの視覚/言語ルート選択など、多数の交絡要因があります。臨床的・監督下の専門 IQ 検査の代替ではないことは Mensa 側も明確にしています。test.mensa.no³ Mensa International⁴

- OpenAI が“IQ 148”を公式に主張したわけではありません。OpenAI は多様なベンチマーク（GPQA 等）を提示しており、Mensa を使った“IQ”を公式評価軸にしていません。OpenAI⁵
- モデルの全機能や安全性、長期記憶、事実整合性、ツール使用、マルチモーダル運用などの包括的性能を直接表す指標でもありません。コミュニティでも IQ は AI 評価として不適切・限定的との議論が強いです。Reddit⁹

なぜ“公開 Mensa > オフライン”になりやすいのか（技術的背景）

- 公開問題の学習混入: Frontier モデルは大規模ウェブコーパスで学習されるため、公開問題が事前曝露されている可能性。これが正答率を押し上げる“汚染”要因となり、一般化可能な推論力と切り分けにくくなります。BinaryVerseAI⁸
- 言語化 (Verbalization) バイアス: 画像問題を LLM に提示する際の文章化は、語順・属性（色→形→向きの順など）・参照の一貫性が正答率に数%単位で影響し得る“プロンプト工学”の領域。比較の公平性を担保するには厳密なプロトコルが必要です。BinaryVerseAI⁸ TrackingAI²
- 監督・再現条件: オンライン練習テストは回答補助や反復試行を技術的に排除しづらく、AI 評価に関しても API/GUI、モデルルータの挙動、ビジョン/テキストの切替など、細部が結果に影響します。Mensa International⁴ Reddit⁹

関連データ（BinaryVerseAI 表より抜粋）

- 公開 Mensa Norway と Offline の両スコアが併記されたモデル比較（2025 年 8 月更新）。以下は同記事の表の一部抜粋です。BinaryVerseAI⁸

モデル	Mensa Norway	Offline
OpenAI GPT-5 Pro	148	116
OpenAI o3	135	116
Gemini 2.5 Pro	134	115
OpenAI o3 Pro	133	109
Claude-4 Opus	117	118
Claude-4 Sonnet	119	107

モデル	Mensa Norway	Offline
Grok-4	121	103
o3 Pro (Vision)	104	100
Gemini 2.5 Pro (Vision)	99	97
o3 (Vision)	97	94
注: 同記事は“公開セットは高得点に寄りやすい、Offline はより厳しめに出る”という傾向の背景（公開問題の学習流入、言語化の難しさ）を詳述しています。BinaryVerseAI ⁸		

実務的な読み替えと推奨

- 研究・評価の観点:
 - Mensa Norway オンラインでの“148”は話題性があるが、探索的・参考的な指標に留め、非公開・監督下ベンチや専門評価（GPQA、医療・法務・エンジニアリングの高難度セット）を主評価軸に据えるのが妥当。OpenAI 自身も後者を主に強調しています。OpenAI⁵
- 導入・PoC の観点:
 - 実ビジネスの評価は、ドメイン特化の検証課題（経済価値のある推論・設計・監査・安全要件）で A/B テストを行い、ルーティング、ツール使用、コンテキスト制御、ガバナンスを統合した総合スコアで意思決定することを推奨します。OpenAI⁶
- 公共コミュニケーションの観点:
 - “IQ 148”の見出しはキャッチーだが、テストの性質と限界、オフライン/公開の差、非監督である点を合わせて説明し、過度な一般化を避けるべきです。test.mensa.no³ Mensa International⁴

オープンな論点（今後の確認事項）

- 再現性: 第三者が同等条件（Vision 提示・同一問題セット・同一制約）で同スコアを再現できるか。Tracking AI は毎週の定期検証を標榜しており、今後の推移観測が有益です。TrackingAI¹
- バリエーションの明確化: ChatGPT-5 Pro（UI 上の“Pro”アクセス）と、API/バックエンドの“GPT-5 Pro（Vision/Verbal）”の差分とルーティングの透明性。テスト対象の厳密なモデル ID 管理が望まれます。Reddit⁹
- 問題セットの健全性: 公開問題の更新や非公開テストの拡充、言語化手順の標準化（プロンプト仕様書化）により、比較可能性と妥当性を高める必要があります。

BinaryVerseAI⁸ TrackingAI²

総括

- 事実面では、Tracking AI は 2025 年 8 月 12 日更新のページ上で GPT-5 Pro（特に Vision）が Mensa Norway 公開オンラインテストで“148”と表示される状況を示しており、図表・更新時刻からも同日付の掲示が確認できます。これは“公開・練習用”かつ“非監督”のテストにおける結果であり、公開問題の学習混入や言語化手順の影響、オフラインテストとの乖離が指摘されます。したがって、この数字は“その設定における上限寄りの成績”を示すもので、即“人間同等の汎用知能 148”を意味するわけではありません。OpenAI の公式評価も IQ ではなく、実務的・学術的ベンチマークを中心に据えています。結論として、ニュース価値は高いが、能力の一般化を論じるにはより厳密な非公開ベンチや監督下評価の継続が不可欠です。

TrackingAI¹ BinaryVerseAI⁸ test.mensa.no³ Mensa
International⁴ OpenAI⁵ OpenAI⁶ CNBC⁷ Reddit⁹

Appendix: Supplementary Video Resources



【ChatGPT 5 完全解説】 GPT-5 Thinking・Pro との違い！小枝 ...

1 day ago



"IQ テスト"攻略法であなたの IQ を 20 上げます

Mar 24, 2020



新発表「GPT-5」は何がすごい？使いこなす方法を徹底解説 ...

4 days ago

もっと詳しく

1

trackingai.org

2

www.trackingai.org

3

test.mensa.no

4

www.mensa.org

5

openai.com

6

openai.com

7

www.cnn.com

8

binaryverseai.com

9

www.reddit.com