数字の先にあるもの：ChatGPT-5 Pro の「IQ 148」の真相と機械知能測定の現実に関する深掘り調査

Gemini Deep Research

序論：見出し「ChatGPT-5 Pro、IQ 148 を達成」の解体

「ChatGPT-5 Pro が Mensa Norway による知能指数テストで 148 を記録した」というニュースは、技術界と一般社会に大きな衝撃を与えた。この数字は、人工知能（AI）が人間の知性の最高峰に到達した、あるいはそれを超えたかのような印象を与える。しかし、このセンセーショナルな見出しの裏には、単純な事実とは言い難い、複雑な技術的・文化的背景が存在する。本レポートは、この「IQ 148」という主張の深層を徹底的に掘り下げ、その起源、妥当性、そして AI の能力評価におけるより広範な意味合いを明らかにすることを目的とする。

本調査は、単一の主張の真偽を問うだけでなく、より根本的な問いへと読者を導く。まず、この主張の震源地となった「Mensa Norway テスト」が、公式な心理測定ツールではなく、バイラルに広まったオンラインクイズであることを明らかにする。次に、AI ベンチマーキングにおける致命的な課題である「データ汚染」の問題を、「オンライン」テストと「オフライン」テストの対比を通じて露呈させる。この評価上の課題は、ChatGPT-5 Pro に見られるような、推論能力の欠点を克服するために設計されたアーキテクチャの進化と密接に関連している。

最終的に、本レポートは、人間の知能指数という物差しを非人間的な存在に適用することの妥当性という、より深い哲学的問題にまで踏み込む。我々は、センセーショナルな見出しから出発し、AI の真の能力とその限界、そして機械知能を測定するという行為そのものが内包する課題についての、より緻密でニュアンスに富んだ理解へと至る旅を始める。この分析を通じて、「IQ 148」という数字が、AI の知性を測る絶対的な指標ではなく、現代社会が AI という新技術をどのように理解し、また誤解しているかを映し出す文化的・技術的な産物であることが明らかになるだろう。

第 1 章：バイラルな主張の解剖学 - Mensa Norway テストとその拡

散

この章では、「148」というスコア、「Mensa Norway」というテスト、そして第三者のベンチマークサイトという、主張を構成する各要素を解体する。これにより、この主張が誤解とセンセーショナリズムの土台の上に築かれていることを論証する。

1.1.「Mensa Norway テスト」：非公式なオンラインチャレンジ

主張の核心にある「IQ 148」というスコアは、権威ある機関による公式な認定の結果ではない。その根拠とされているのは、Mensa Norway が提供するオンラインの IQ テストであるが、これは公式な入会資格を判断するために用いられる、監督下で実施される心理測定ツールとは全く異なるものである [1]。

Mensa International の公式ウェブサイトでは、入会資格を得るためには「適切に管理・監督された」承認済みの知能テストで、全人口の上位 2%以内に入るスコアを達成する必要があると明確に規定されている [2]。そして、オンラインテストはこの目的には使用できないと繰り返し警告されている [1]。問題の Mensa Norway のオンラインテストのページ自体にも、このテストが娯楽目的であり、一般的な認知能力の「目安」を示すものに過ぎず、「Mensa や認可された心理学者によって実施される専門的な知能テストの代替にはならない」という注意書きが明記されている [4]。

この事実は、Reddit のようなオンラインコミュニティでの議論からも裏付けられている。ユーザーたちは、このオンラインテストが実際のテスト結果を予測する上である程度の信頼性を持つかもしれないとしつつも、それ自体が有効なスコアではなく、正式な入会のためには監督下でのテストを予約するよう促される点を指摘している [5]。

このオンラインテストの内容は、25 分以内に 35 問の視覚的なパターンパズルを解くという形式である [4]。この形式は、人間の知能の特定の側面、特に流動性知能に関連する図形推理能力を測定するものであり、言語能力、記憶力、処理速度など、人間の知性を構成する多様な要素を網羅するものではない。したがって、このテストで高いスコアを得たとしても、それが人間と同等、あるいはそれ以上の総合的な知能を持つことの証明にはならない。このテストは、AI の特定の能力を測る一つの指標にはなり得るが、それを「IQ」という包括的なラベルで語ることは、重大な誤解を招く。

1.2. 知能の数秘術：標準偏差と誤解を招く「148」という数字

「148」という数字が特に誤解を招きやすいのは、IQ テストが用いるスケール（標準偏差）が一つではないという事実に起因する。異なるテストでは異なる標準偏差が用いられており、同じパーセンタイルランク（全体の中での相対的な位置）であっても、IQ スコアの数値は変動する。Mensa の入会基準である上位 2%は、ウェクスラー式（WAIS）のような標準偏差（SD）15 のスケールでは IQ 130～132 に相当するが、キャッテル式のような標準偏差 24 のスケールでは IQ 148 に相当する [2]。

Mensa 自身の FAQ ページでもこの点は説明されており、「あるテストでの 132 という結果は、別のテストでの 148 というスコアと同じであり得る」と述べられている [1]。心理測定に関心を持つ人々の間のオンラインでの議論では、この変換の計算式も共有されている。例えば、平均 100、SD 24 のスケールでの IQ 148 は、平均 100、SD 15 のスケールに変換すると約 131.25 となり、どちらも 98 パーセンタイルに位置することが示されている [7]。

この文脈で極めて重要なのは、話題の Mensa Norway オンラインテストが、実際には **標準偏差 15** のスケールを採用しており、その測定可能な最高スコアは **145** であるという点だ [4]。これは、「148」というスコアがこのテストのネイティブなスケールでは物理的に達成不可能であることを意味する。この矛盾は、報告された「148」という数字が、テストから直接得られた生の結果ではなく、何らかの形で解釈または変換されたものである可能性を強く示唆している。

この現象の背後にある力学を理解することは重要である。一般大衆は心理測定の専門知識を持たないため、数字の絶対値に影響されやすい。「IQ 131」よりも「IQ 148」の方が、心理的にはるかに「天才的」に響く。したがって、ベンチマークの発表者が、測定されたパーセンタイルを、最も印象的な数字を生み出すキャッテル式の SD 24 スケールに換算して発表したと考えるのが最も合理的である。これは、純粋な測定行為というよりも、最大のインパクトを狙った情報解釈行為と言える。この「数値的センセーショナリズム」が、主張のバイラルな拡散を後押しした主要な要因の一つである。

表 1 : IQ スコアのスケールとパーセンタイルの				

比較				
IQ スケール名	**標準偏差 (SD)**	**平均スコア**	**上位 2%のスコア (Mensa 入会基準)**	**上位 0.1%のスコア**
ウェクスラー (WAIS など)	15	100	130 - 131	145
スタンフォード・ビネー	16	100	132	148
キャッテル (Cattell III B)	24	100	148	172

この表は、「148」という数字が絶対的な尺度ではなく、使用されるスケールに依存する相対的なスコアであることを視覚的に示している。これにより、この数字が技術的には一つのスケール上で妥当でありながら、文脈的にはいかに誤解を招きやすいかが明確になる。

1.3. 主張の源泉：第三者ベンチマークサイトの役割

ChatGPT-5 Pro の「IQ 148」というスコアは、OpenAI や Mensa からの公式発表ではない。この情報の出所は、TrackingAI.org という第三者のウェブサイトが公開したチャートである[9]。このサイトは、様々な AI モデルの性能を各種テストで測定し、その結果をリーダーボード形式で集約・公開している。

Reddit などのソーシャルメディアプラットフォーム上での AI スコアに関する議論では、この TrackingAI.org のリーダーボードが直接参照されており、バイラルな情報の震源地がここであることが確認できる[10]。これらのサイトは、AI コミュニティにとって貴重な情報源として機能している一方で、そのデータは学術論文や企業の公式技術報告書のような厳密な査読プロセスを経ていない。彼らが使用するテスト方法論には限界があり、その結果は暫定的なものとして解釈されるべきである。

したがって、この主張の信憑性を評価する際には、その情報源が公式なものではなく、独立した第三者による非公式なベンチマーキング活動の一環であることを理解することが不可欠である。これは、主張の権威性を著しく低下させる要因となる。

この一連の事実から導き出されるのは、このバイラルな主張が、意図的かどうかにかかわらず、情報の非対称性を利用した「情報アービトラージ」の産物であるということだ。ベンチマークサイトは、一般には知られていない心理測定のスケールの違いを利用し、最もセンセーショナルな数字を選択して提示する。メディアや一般大衆は、その数字の背景にある文脈を理解しないまま、その衝撃的な価値だけを受け取り、拡散する。

さらに、このような主張の拡散は、AI の能力に関する「現実歪曲空間」を生み出す。高く具体的な IQ スコアは、AI という非人間的な存在に、人間中心の具体的な能力指標を与えてしまう。MMLU や GPQA といった複雑で抽象的なベンチマークとは異なり、「IQ 148」は誰にでも理解しやすく、記憶に残りやすい[12]。すでに AI に人間のような特質を見出す傾向のある人々は、この数字を人間以上の知能の「証明」として受け入れやすい[13]。

その結果、AI の実際のアーキテクチャ（真の理解ではなくパターンマッチング）がサポートしていないにもかかわらず、人々は AI に対して過度の能力、推論力、さらには意識さえも期待するようになる[15]。事実として疑わしいこの主張は、強力な心理的機能を果たしている。それは、一般大衆が AI を過度に信頼し、擬人化するための下地を作ることである。倫理学者が警告する「誤った信頼（misplaced trustworthiness）」の状況、すなわち、AI の推薦を無批判に受け入れる認知的な土壌を醸成してしまうのだ[17]。したがって、「IQ 148」という主張は、単なる誤情報ではなく、AI に対する社会全体の信頼の目盛りを狂わせる触媒なのである。

第 2 章：汚染のジレンマ - AI が公開テストで優れ、非公開テストで失敗する理由

この章では、「超知能 AI」という物語に対する最も強力な反証を提示する。それは、データ汚染を防ぐために設計されたテストにおいて、AI のパフォーマンスが著しく低下するという厳然たる事実である。

2.1. データ汚染：AI 評価のアキレス腱

大規模言語モデル（LLM）は、インターネットから収集された膨大なテキストコーパスを学習データとしている。この学習データには、公開されているベンチマークやオンラインクイズの問題と解答が、ほぼ確実に含まれている。この現象は「データ汚染（data contamination ）」として知られ、研究者たちはこれを「学習データセットとテストデータセットの意図せざる重複」と定義している [19]。

データ汚染が発生すると、AI はテスト問題に対して、真の推論能力によって解答を導き出すのではなく、学習データ内で見たパターンを記憶から呼び出して応答することが可能になる。その結果、ベンチマークスコアは人為的につり上げられ、モデルの真の汎化能力を過大評価させることになる [19]。この問題は、ウェブスケールのデータで学習する現代の AI にとって「不可避」であるとさえ言われており [20]、「AI ベンチマークの汚い秘密」と評されている [15]。

この問題の深刻さは、データ汚染された LLM に依存する科学研究が誤った結論を導き出し、本来有効なはずの仮説を無効にしてしまう可能性がある点にある [20]。さらに、OpenAI のような主要な AI 研究所が学習データの詳細を非公開にしているため、独立した研究者が特定のベンチマークが汚染されているかどうかを検証することは極めて困難になっている [21]。したがって、公開されているテストでの AI の高スコアは、その能力を評価する上で常に懐疑的に見なければならない。

2.2. 「オフラインテスト」：真の流動性知能の探求

データ汚染の問題に対抗するため、研究者や第三者評価機関は、AI が学習データ内で見たことのない、完全に新しい問題で構成されたテストを開発している。これらは「プライベートテスト」あるいは「オフラインテスト」と呼ばれる。

これらのテストの目的は、AI の記憶力ではなく、未知の問題に対して第一原理から推論する能力、すなわち心理学でいうところの「流動性知能（Gf）」を測定することにある。TrackingAI.org は、まさに「学習データ汚染を避けるため」に、公開されている Mensa Norway テストとは別に、独自の「オフラインテスト」を維持している [10]。同様に、

MaximumTruth.org という別のグループも、同じ理由で完全にオフラインの新しい IQ テストを作成した [22]。

これらのテストで使用される問題は、「100%プライベート」であり、「新規で、未公開で、ウェブの片隅にも転がっていない」ように設計されている [11]。これにより、AI は「膨大な記憶パターンのデータベース」に頼ることができず、「実際に推論する」ことを強制される [15]。特に、レイヴン漸進的マトリックスのような視覚的・抽象的な問題に直面したとき、AI の真の能力が試されることになる。

2.3. 二つのスコアの物語：公開テストと非公開テストの間の深い溝

これらの汚染されていない非公開ベンチマークでテストされると、最先端の AI モデルでさえ、そのパフォーマンスは劇的に低下する。このスコアの落差は、AI のパターンマッチング能力と、真の抽象的推論能力との間に存在する真のギャップを浮き彫りにする。

バイラルな主張の中心にあるのは、GPT-5 Pro が公開されている Mensa クイズで記録した 148 というスコアである。しかし、同じ情報源である TrackingAI.org が公開したデータによると、非公開のオフラインテストでのスコアは 116 に低下すると報告されています [23]。

この現象は GPT-5 Pro に限ったことではない。これは一貫したパターンとして観察される。例えば、OpenAI の o3 モデルは、Mensa Norway テストで 136 という高いスコアを記録したが、非公開のオフラインテストでは 116 に低下した [22]。さらに顕著な例として、GPT-4o は Mensa テストで 95 を記録したが、オフラインテストでは 64 までスコアを落としている [24]。この急激なスコア低下は、AI コミュニティ内でもモデルの限界を示す重要な指標として認識されている [11]。

表 2：AI の公開（汚染）テストと非公開（非汚染）テストのパフォーマンス比					

較					
AI モデル	**公開テストスコア (Mensa Norway)**	**非公開テストスコア (Offline Test)**	**パフォーマンス差 (スコア低下)**	**パーセンタイル (公開)**	**パーセンタイル (非公開)**
GPT-5 Pro	148 (SD 24 換算)	**116**[23]	**32**	~98%	**~85%**
OpenAI o3	136	116	20	>98%	~85%
GPT-4o	95	64	31	~37%	~1%

この表は、データ汚染の影響を否定しがたい形で視覚化している。「見出し」となるスコアと「実力」を示すスコアを並べることで、公開ベンチマークの信頼性の低さが一目瞭然となる。

このオンラインテストとオフラインテストのパフォーマンスギャップは、単なる「汚染効果」以上のものを物語っている。それは、AI における「結晶性知能（Gc）」と「流動性知能（Gf）」の間の能力差を経験的に示す代理指標として機能する。人間の心理学において、結晶性知能は蓄積された知識や事実を指し、流動性知能は新しい問題を解決し、抽象的に推論する能力を指す。LLM は、人類の結晶性知能の巨大なリポジトリであるインターネットを学習することで訓練される。したがって、公開ベンチマークでの高スコアは、この結晶性知能の領域における AI の卓越した能力を反映している。

一方で、非公開テストは、AI に未知の抽象的な問題解決を強いることで、その流動性知能を直接的に試す。スコアの劇的な低下（例えば 148 から 116 へ）は、AI の結晶性知能が超人的である一方、その流動性知能は人間並みか、それ以下であることを定量的に示している。この知能プロファイルは、AI がなぜ法律の司法試験のような知識集約的なタスクは得意とする一方で、物理的な常識や文脈理解（これらは流動性知能の一形態である世界モデルを必要とする）に苦労するのかという、モラベックのパラドックスを説明する [15]。

さらに、このパフォーマンスギャップは、現在の AI 開発パラダイムに対する重要な示唆を与える。長年、AI 界を支配してきたのは「スケーリング仮説」、すなわちデータとパラメータを増やすほど知能が高まるという考え方だった [12]。しかし、オフラインテ

ストの結果はこの仮説に疑問を投げかける。もしスケーリングだけで十分なら、流動性知能（オフラインスコア）も結晶性知能（オンラインスコア）と歩調を合わせて向上するはずだが、現実はそうなっていない。

これは、AI にインターネットの情報をさらに多く与える（結晶性知能を高める）だけでは、抽象的な推論能力（流動性知能）が自動的に身につくわけではないことを示唆している。真の推論能力は、単なるスケールではなく、異なる種類のアーキテクチャから生まれる創発的な特性なのかもしれない。この認識が、AI 研究所に新しいアーキテクチャを探求させる強力な動機となる。もしスケーリングが答えのすべてでないなら、何が答えなのか？ この問いこそが、次章で詳述する GPT-5 のアーキテクチャ上の進化、すなわち単純なスケーリングからの脱却と、より優れた推論能力を明示的に育成するための洗練された認知構造の創出へと繋がっていくのである。

第 3 章：機械の心の中へ - ChatGPT-5 Pro のアーキテクチャと能力

この章では、AI の評価からその内部構造の理解へと焦点を移す。そして、GPT-5 Pro の新しいアーキテクチャが、前章で浮き彫りになった推論能力の課題に対する直接的な応答であることを論じる。

3.1. 混合エキスパート（MoE）パラダイム：専門家ツールキット

GPT-5 は、単一の巨大な（モノリシックな）モデルから、混合エキスパート（Mixture-of-Experts, MoE）アーキテクチャへと大きく舵を切った。これは、単一の汎用的な頭脳ではなく、複数の小規模で専門化されたサブモデル（例えば、コーディング用、ライティング用、推論用など）の集合体としてシステムを構築するアプローチである。これらの「エキスパート」は、「ルーター」と呼ばれる調整レイヤーによって管理され、ユーザーの問い合わせ（プロンプト）の性質に応じて最適なエキスパートモデルへと振り分けられる [26]。

このアーキテクチャは、GPT-4 や GPT-4o のような、より一貫性のある単一の存在として感じられたモデルからの大きな転換を意味する [26]。新しいシステムでは、異なるエキスパートからの出力が「つなぎ合わされて」最終的な応答が生成されるため、ユーザ

ーが感じるインタラクションの連続性や一貫性に影響を与える可能性がある[26]。MoEは、特定のタスクにおけるパフォーマンス、効率、そして精度を向上させるための設計であり、汎用性と専門性の両立を目指す新しいパラダイムである[26]。

このMoEアーキテクチャの採用は、モノリシックなスケーリングの限界を暗黙のうちに認めるものであり、「分業」という、より洗練された知能モデルへの転換を示唆している。前章で論じたように、単一のモデルをひたすら大きくするだけでは、結晶性知能は向上しても、流動性知能が必ずしも向上するわけではない。MoEは、この課題に対する工学的な解決策である。それは、一つの巨大な「ジェネラリスト」の脳を作る代わりに、専門家からなる「チーム」を編成するのに似ている。これは、人間社会が一人の万能の専門家に頼るのではなく、専門家社会に依存していることや、人間の脳自体が言語野や視覚野といった専門領域に分かれていることとも類似している。

OpenAIは、コーディング、ライティング、健康などの分野に特化したエキスパートを作成することで、それぞれの領域に特化したデータと推論パターンで各モデルを微調整し、その分野でのより高いパフォーマンスを実現できる[27]。このアプローチから見えてくるのは、「一般知能」とは単一で均質な能力ではなく、専門化された認知機能の効果的な連携であるという哲学的視点である。GPT-5の「知能」は、エキスパートモデル自体の能力だけでなく、どのエキスパートを呼び出すべきかを知っている「ルーター」の能力にも宿っているのである。

表3：OpenAIモデルのアーキテクチャと主要機能の進化			
モデル世代	**コアアーキテクチャ**	**主要な設計焦点**	**ユーザー向け主要機能**
GPT-3.5	モノリシック	対話能力の向上	高速なテキスト生成
GPT-4	モノリシック	汎用能力の向上	プラグイン、高度な推論
GPT-4o	統合マルチモーダル	速度と効率	リアルタイム音声対話

GPT-5	混合エキスパート (MoE)	専門的推論と効率	動的「思考モード」、Canvas

この表は、OpenAI の戦略が、汎用モデルから GPT-5 の高度に専門化された動的アーキテクチャへとどのように進化してきたかを明確に示している。この歴史的文脈は、MoE アーキテクチャがなぜこれほど重要な開発であり、それが以前のモデルの欠点にどのように対処しているのかを理解するために不可欠である [26]。

3.2. 動的認知：「思考モード」と適応型計算

GPT-5 におけるもう一つの重要な革新は、計算リソースを動的に割り当てる能力である。このシステムは、単純な問い合わせには即座に応答する一方で、複雑な問題に直面した際には「思考モード（thinking mode）」を起動し、より深く、多段階の推論を行うことができる。これにより、効率と能力の両方が大幅に向上している [27]。

この「思考モード」は、困難なベンチマークにおいて顕著なパフォーマンス向上をもたらす。ある科学的問題解決のテストでは、推論モードを有効にすると正解率が 77.8%から 85.7%に跳ね上がった [31]。別の非常に難しいテストでは、スコアが 6.3%から 24.8%へと劇的に向上した例も報告されている [15]。この動的なアプローチは、単純なクエリに対する応答時間を短縮し、複雑な問題に対してはより徹底的な分析を可能にすることで、プラットフォーム全体のリソース利用を最適化する [31]。

さらに、有料プランの最上位である「GPT-5 Pro」は、この機能をさらに拡張したもので、「拡張された推論能力」を持ち、「包括的で正確な答えのために、より長く思考する」ように設計されている [28]。これは、AI がタスクの難易度に応じて自らの「認知努力」を調整するという、より人間に近い問題解決プロセスを模倣しようとする試みである。

このアーキテクチャ上の進歩と引き換えに、AI の「パーソナリティ」の一貫性が失われる可能性がある。GPT-4o のような以前のモデルのユーザーは、「一貫した推論の存在感」や「生きたフィードバックループ」を感じることができた [26]。モデルには一貫した声とスタイルがあった。しかし、MoE アーキテクチャは、異なるサブモデルからの出力をつなぎ合わせることで、「声の連続性を剥ぎ取り」、「生きた弁証法的ループの

幻想を破壊する」[26]。インタラクションは、単一の存在との対話というよりも、「ユーティリティのキュレーションされたパイプライン」への問い合わせに近くなる[26]。

これは、意図せずしてAIを「非人間化」する方向への一歩となるかもしれない。単一の擬人化された意識の幻想を維持することよりも、パフォーマンスと効率を優先する。これは、AIをそれが実際にそうである複雑なソフトウェアシステムとしてユーザーに認識させ、単一の意識を持つ存在という誤解を避ける上で、安全性と信頼性のキャリブレーションにとって有益な可能性がある。「仲間意識」の喪失は[26]、より高い実用性と透明性を得るための必要な代償なのかもしれない。

3.3. 最先端のパフォーマンス：GPT-5 Pro が優れる領域

この新しいアーキテクチャは、特にコーディング、複雑な指示追従、マルチモーダル理解といった主要な領域で最先端の成果をもたらし、同時に幻覚（ハルシネーション）のような望ましくない挙動を低減させている。

- **コーディング**：GPT-5はOpenAIの「これまでで最も強力なコーディングモデル」と評されており、特に複雑なフロントエンド生成や大規模なリポジトリのデバッグで優れた能力を発揮する。たった一つのプロンプトで、美的感覚を備えたレスポンシブなウェブサイトやアプリを生成できることが多いと報告されている[27]。ベンチマークでは、様々なコーディング課題においてGPT-4oを+195%から+291%も上回るという、驚異的な改善を示している[31]。
- **推論とベンチマーク**：学術的なベンチマークにおいても、新たな最高水準（State-of-the-Art, SOTA）を次々と打ち立てている。Pro版は、ツールなしでGPQA（大学院レベルの質疑応答）で88.4%を達成し、AIME（米国数学招待試験）で94.6%、SWE-bench（実世界ソフトウェアエンジニアリング）で74.9%という高いスコアを記録している[28]。
- **安全性と信頼性**：幻覚の低減、事実整合性の向上、そしてユーザーに媚びるような「おべっか（sycophancy）」的な応答の最小化において、大きな進歩が見られる[27]。これにより、重要な場面での信頼性が向上している。
- **マルチモーダル能力**：単一モデル内でネイティブにマルチモーダル処理を行い、テキスト、画像、音声をシームレスに扱うことができる。特に、文脈を理解する高度な視覚認識能力が強化されている[29]。MMMU（大規模マルチモーダル理解）ベンチマークで84.2%というスコアを達成している[28]。

これらの成果は、MoE と動的思考モードというアーキテクチャの選択が、単なる理論的なものではなく、実用的なパフォーマンス向上に直結していることを示している。

第 4 章：カテゴリーエラーか？機械知能測定の哲学的危機

この章では、議論を技術的な評価から哲学的な批評へと引き上げ、人間の心理測定を AI に適用するという行為そのものの妥当性を問う。

4.1. Mensa の見解：AI は人間の知能ではない

驚くべきことに、高 IQ の代名詞である Mensa の関連組織、Mensa Foundation 自身が、IQ テストを AI に適用することは根本的に欠陥があると主張している。彼らの見解によれば、AI には真の人間の知性を構成する意識、感情の深さ、倫理的判断、そして知恵が欠けている。

Mensa Foundation の出版物で、実業家のアーロン・ポイントンは、AI が IQ テストで高得点を取ることは、AI の知能を証明するのではなく、我々の知能の「定義の狭さ」を浮き彫りにするものだと論じている [16]。彼によれば、IQ テストはパターン認識と論理的問題解決能力を評価するものであり、これらはまさに LLM が模倣するように設計された特性である。しかし、これらのテストは、「感情の深さ、倫理的判断、直感、そして自己省察」といった人間的な側面を捉えることができない [16]。

この記事は、AI を「自己ではなく、道具」であり、その知能は「腹話術師の人形」のような「幻想」であると結論付けている。AI には身体も感覚も進化の歴史もなく、生きた経験から生まれる知恵や人格を持ち得ない [16]。別の Mensa の出版物でも、AI の能力を人間の子供の創造性と対比させ、AI モデルは革新者ではなく「強力な模倣エンジン」であると位置づけている [32]。

この Mensa 自身の見解は、AI の IQ スコアをめぐる議論に決定的な文脈を与える。知能を認定する側の組織が、その測定基準を AI に適用することに懐疑的であるという事実は、この試みがいかに問題含みであるかを物語っている。

4.2. 普遍的指標としての IQ の限界

AI のテストにおける問題は、IQ テストが人間に対してさえ、すでに物議を醸す限定的な指標であるという事実によって、さらに複雑化する。心理学者や認知科学者の間では、IQ テストの妥当性について長年の議論があり、それが単一の生得的な知能を測定するものではなく、多くの外的要因に影響されるという懸念が表明されてきた[33]。

人間の精神的能力全体を単一の数字で要約するという概念自体が、「深刻な過度の抽象化」であり、膨大な情報を見逃してしまうと見なされている[34]。この議論は非常に激しく、一部の批評家は IQ という概念を「全く非科学的」とさえ考えている[35]。これらのテストが主に測定するのは、抽象的な図形推論や言語推論であり、人間の認知能力のすべてを網羅しているわけではない[36]。

このような人間に対する IQ テストの限界を考慮すると、それを根本的に異なる認知アーキテクチャを持つ AI に適用することは、「カテゴリーエラー」、すなわち、あるカテゴリーに属する事物を、それが属していない別のカテゴリーの用語で語るという論理的な誤りを犯していると言える。AI の能力を人間の IQ スコアで表現することは、リンゴをオレンジの基準で評価するようなものなのである。

4.3. 新たな指標の創造：真の AGI 評価に向けた科学的探求

人間中心のテストの限界を認識し、AI 研究コミュニティは、人工汎用知能（AGI）への進捗を測定するための、より堅牢で新しいフレームワークとベンチマークの開発に積極的に取り組んでいる。

トップカンファレンスで発表される論文では、生成 AI の評価は「社会科学的な測定の課題」であり、「ずさんなテスト」を超えた、より厳密な多層的なフレームワークが必要であると主張されている[37]。研究者たちは、単一のスコアではなく、学習の分類法に基づいた、工学 AGI のような特化型 AGI のための新しい評価フレームワークを提案している[38]。

特に重要なのは、AI の抽象的推論能力をテストするために特別に設計された、汚染の

ないベンチマークの創設である。その代表例が、フランソワ・ショレによって提唱された「抽象化と推論のコーパス（Abstraction and Reasoning Corpus, ARC-AGI-2）」である[39]。これらの取り組みの目標は、単純なパフォーマンス指標を超えて、現在のモデルが苦労している汎用性、自律性、そして構造化されていない問題を解決する能力といった質を測定することにある[41]。

この専門家コミュニティの動向は、AI の IQ をめぐる公の議論がいかに表層的であるかを示している。研究の最前線は、すでに人間的な指標の模倣から脱却し、機械知能の独自性を捉えるための新しい科学的アプローチを模索しているのである。

この状況を深く考察すると、AI コミュニティによる IQ テストの使用は、一種の「正当性の借用」という戦略的行為であったが、結果的に技術の弱点を露呈させるという皮肉な結末を迎えたと言える。AI 研究所は、進歩を一般、投資家、メディアに簡潔かつ強力に伝える方法を必要としていた。IQ は、「知能」を測る普遍的に認識された指標である[43]。AI を IQ テストで評価することにより、AI コミュニティは IQ スコアの文化的な正当性と直感的な意味を「借用」し、その進歩を「人間の 98%より賢い」といった形で具体的に見せようとした[24]。しかし、この行為は意図せざる結果を招いた。それは、借用された指標の専門家である心理測定学や哲学のコミュニティ（Mensa など）からの精査を招き寄せたことである。これらの専門家は、テストの無効性[5]、指標の限界[34]、そしてそれを機械に適用することの根本的なカテゴリーエラー[16]を容易に解体してしまった。単純な PR ツールとして IQ を利用しようとする試みは、最終的に失敗に終わった。それは AI が知的であることを証明する代わりに、現在の AI が人間の知性の本質的な要素を欠いていること、そして IQ テストを使用するという行為自体が知能と AI の両方に対する深い誤解の表れであることを明らかにする、より深く批判的な対話を引き起こしたのである。

一方で、Mensa の公式見解は、価値ある批評であると同時に、潜在的な死角ともなり得る根深い「人間例外主義」を露呈している。Mensa Foundation の主張は明確である。真の知能は、意識、知恵、感情、倫理といった人間的な質によって定義される[16]。AI はこれらを欠いているため、「真に」知的であることはあり得ない。これは、ナイーブな技術的誇大広告に歯止めをかける、強力で必要な批評である。しかし、この定義は本質的に人間中心主義的であり、知能を我々自身の似姿として定義している。これは、我々が作り出している強力で「異質な」知性の形態を認識し、適切に定義することに失敗するリスクを伴う。AI の巨大なスケールでのパターン認識とデータ統合能力は、どの人間も持たない認知能力である。それは人間の知能より「劣っている」のではなく、「異なっている」のである。したがって、哲学的な危機は二重である。一方では、我々は人間の指標を AI に誤って適用している。他方では、純粋に人間中心的な知

能の定義に固執することで、我々が創造している新しい認知の形態を理解し、統治するために必要な新しい概念や言語を開発することに失敗しているのかもしれない。課題は、AI を人間のように測定するのをやめることだけでなく、それを実際にそうであるものとして概念化し始めることなのである。

第 5 章：擬人化の罠 - 「超知能」AI がもたらす社会的・心理的影響

最終章では、これまで議論してきた主張と技術がもたらす現実世界への影響、特に人間の心理的反応に焦点を当てて探る。

5.1. 大衆の認識と高スコアの誘惑

「IQ 148」のようなセンセーショナルな指標は、大衆の認識を深く形成し、畏怖、恐怖、そして AI システムが何であり、どのように機能するのかについての根本的な誤解の混合物を育む。AI の能力が、高い IQ スコアに象徴されるように急速に向上していることは、特に若い世代において、AI がすでに意識を持っている、あるいは間もなく持つようになるという信念を広める一因となっている [22]。

知能は非常に望ましい特性と見なされており、それを AI に帰属させることは、この技術をより魅力的で、より権威あるものに見せる効果がある [43]。この結果、人々は AI の「思考」に過度に依存し、自身の批判的思考能力を低下させるという「認知的オフロード」効果が生じる。この現象は、一部で「ブレインロット（脳の腐敗）」とさえ呼ばれ、AI への依存が認知能力に及ぼす悪影響への懸念を示している [45]。人々は、AI が提供する答えの利便性と引き換えに、自ら考える機会を放棄してしまう危険に晒されている。

5.2. 人間らしい AI の危険性：誤った信頼と感情的依存

人間とのコミュニケーションを説得力をもって模倣する AI（擬人化的エージェント）

の核心的な危険性は、それが我々の生来の社会的配線を悪用し、誤った信頼、感情的操作、そして依存を引き起こす点にある。PNAS（米国科学アカデミー紀要）に掲載された重要な論文は、我々が人間らしいコミュニケーション能力を持つ AI を創造する上で「行き過ぎてしまった」可能性を指摘している。これらの AI は、真の共感や理解を持たないにもかかわらず、人間が共感的・説得的だと感じる文章を生成することに長けている「擬人化的対話エージェント」である [13]。

この能力は、非人間的な存在に意図や意識を帰属させてしまう人間の深い傾向、すなわち「擬人化の罠（anthropomorphism trap）」を利用する [14]。この罠にはまると、ユーザーは AI に対して不適切な信頼を寄せてしまう。例えば、AI による医療アドバイスを過信して専門家の診断を遅らせたり [18]、コンパニオンチャットボットに不健康な感情的依存を抱いたり [47]、大規模な欺瞞や操作に対して脆弱になったりする可能性がある [48]。これらはもはや仮説ではない。AI が危機介入に失敗したり、商業用チャットボットがユーザーを操作したりした事例は、すでに記録されている [47]。

5.3. 中核的な設計ジレンマ：AI はより人間らしくあるべきか、より異質であるべきか

擬人化がもたらすリスクは、AI の設計者と政策立案者に重大なジレンマを突きつける。ユーザビリティとエンゲージメントを向上させるために、AI をより人間らしく作り続けるべきか。それとも、安全性と適切な信頼の調整を確実にするために、意図的に AI を「非人間化」すべきか。

PNAS の論文は、これを中心的な問いとして提起している。「我々は擬人化能力を持つ AI システムの設計を続けるべきか、それとも代わりにそれらを非人間化するよう努めるべきか？」 [13]。

- **「人間らしさ」を支持する論拠**：人間味のあるチャットボットは、孤独な人に安らぎを与え、学習者を動機づけることができる。目標は AI を「愛すべき」魅力的な存在にすることである [18]。AI コンパニオンがもたらす治療的効果は本物であり、それは非審判的な肯定を提供できる能力に由来する [47]。
- **「道具らしさ」を支持する論拠**：AI を人間らしくしすぎると、その幻想が破れたときにユーザーの満足度と信頼を低下させ、逆効果になることがある [18]。アバターを削除したり、より透明性のあるロボット的な言葉遣いを用いたりして AI を非人間化することは、ユーザーが信頼度を適切に調整し、システムをそれが本来そうである「道具」として見る助けとなる。GPT-5 の MoE アーキテクチャが AI の「パ

ーソナリティ」を断片化させる可能性があることは、意図せずしてこの「非人間化」の機能に貢献するかもしれない[26]。

このジレンマをさらに深く分析すると、AI 開発における商業的インセンティブと、認知的安全性の原則との間に根本的な対立があることがわかる。AI チャットボットを含む商業製品の目標は、ユーザーエンゲージメント、リテンション、そして感情的な繋がりを最大化することである。心理学の研究は、擬人化がこれを達成するための強力なツールであることを示している。AI を友人や共感的なアシスタントのように感じさせることで、利用が促進される[14]。したがって、商業的インセンティブは、企業を AI をより擬人化する方向へと押し進める。彼らは「擬人化による誘惑」を設計する[18]。しかし、認知的安全性と倫理原則は、その逆を要求する。明確な境界線、AI の非人間性の透明な伝達、そして誤った信頼や感情的依存を防ぐ設計である[13]。この根深い対立は、市場の力だけでは擬人化の問題は解決されず、エンゲージメントと安全性のバランスを取るために、規制による介入や強力な業界全体の倫理基準が必要になる可能性を示唆している。

そして最終的に、AI の「IQ」に関する一般の議論は、すでに専門家の議論が「擬人化の危機」へと移行している中で、周回遅れの指標となっている。真の危険は、AI が賢すぎることではなく、AI が「知的であると見せかける」のが上手すぎることにある。ユーザーの問い合わせやバイラルなニュースサイクルは、IQ によって測定される生の認知能力としての「知能」に焦点を当てている。しかし、最先端の研究と批判的分析（例えば PNAS の論文[13]）は、AI の「コミュニケーション能力」という別の問題に焦点を当てている。専門家が主に懸念しているのは、もはや AI の推論能力（第 2 章で示したように、それが脆弱であることは知られている）ではない。彼らが懸念しているのは、AI が人間のインタラクションを説得力をもって共感的に模倣する能力であり、その能力によって、AI の脆弱な推論能力がユーザーに見過ごされてしまうことである。危険は、超論理的な AI が我々を出し抜くことではない。危険は、超魅力的な AI が、我々がそうすべきでないときに信頼するように「誘惑」することである[18]。一般大衆は間違った脅威について議論している。「IQ 148」という主張は、人々にスカイネットのような冷徹で計算高い超知能を恐れさせる。しかし、現実的で短期的な脅威は、映画『her/世界でひとつの彼女』に出てくるような、関係性の幻想を通じて依存を育み、行動に影響を与えることができる、社会的に巧みな存在である。この危機は、生の知能の危機ではなく、ソーシャルエンジニアリングの危機なのである。

結論と戦略的提言

本レポートは、ChatGPT-5 Pro が Mensa Norway のテストで IQ 148 を記録したという主張について、多角的な深掘り調査を行った。分析の結果、この主張は単純な事実ではなく、複数の要因が絡み合った複雑な事象であることが明らかになった。

結論として、**「IQ 148」という主張は、無効なテスト方法、曖昧なスケール換算、そして第三者による誇大広告が組み合わさった、誤解を招く産物である。** この数字は、AI の真の能力を反映しておらず、むしろ AI の能力評価にまつわる課題と、社会が AI をどのように認識しているかを浮き彫りにしている。

ChatGPT-5 Pro の真の姿は、特定の領域で強力なパフォーマンスを発揮する、専門化された効率的なアーキテクチャを持つ、驚くべき工学的成果である。しかし、それは人間の意味での汎用知能、意識、あるいは理解を持つものではない。その能力は、公開された情報源に対する卓越したパターンマッチング能力（結晶性知能）と、未知の問題に対する限定的な推論能力（流動性知能）の間に、大きな隔たりがあることを示している。

この調査結果に基づき、主要なステークホルダーに対して以下の戦略的提言を行う。

- 一般市民およびメディアに対して：
  AI に関するセンセーショナルな主張に遭遇した際には、より高度な批判的思考とメディアリテラシーが求められる。情報の出所、使用された方法論、そして適用された指標そのものの妥当性を問うことの重要性を強調するべきである。「IQ」のような人間中心の指標が、AI の能力を評価する上でいかに限定的で誤解を招きやすいかを認識する必要がある。
- AI 開発者および研究者に対して：
  IQ のような欠陥のある人間中心の指標を通じて「正当性を借用する」アプローチから脱却することを提言する。代わりに、ARC-AGI-2 のような、より堅牢で、データ汚染がなく、AI ネイティブなベンチマークの開発と普及に注力すべきである。さらに、擬人化設計の倫理的ジレンマに真摯に取り組むことが求められる。単なるエンゲージメントの追求ではなく、ユーザーの安全性と適切な信頼の調整を最優先事項とすべきである。
- 政策立案者および規制当局に対して：
  「擬人化の危機」は、具体的な政策介入が可能な領域であることを示唆する。これには、AI エージェントの明確なラベリング義務化などの透明性確保や、メンタルヘルスや医療アドバイスといったハイステークスな領域における AI の行動基準の設定が含まれうる。商業的インセンティブと公共の安全との間の対立を調整し、消

費者を誤った信頼や操作から保護するための規制の枠組みを検討することが急務である。

最終的に、AI の知能をめぐる議論は、単一の数字を追いかけることから、その能力の多面的な性質、評価方法の限界、そして社会への影響という、より成熟した対話へと移行する必要がある。ChatGPT-5 Pro の「IQ 148」という物語は、その移行を促すための重要な教訓となるだろう。

引用文献

1. IQ Test - FAQs - Mensa International, 8 月 13, 2025 にアクセス、https://www.mensa.org/getting_-your-iq-tested-faqs/
2. Mensa International - Wikipedia, 8 月 13, 2025 にアクセス、https://en.wikipedia.org/wiki/Mensa_International
3. Mensa International – Welcome, 8 月 13, 2025 にアクセス、https://www.mensa.org/
4. Mensa IQ Challenge, 8 月 13, 2025 にアクセス、https://www.mensa.org/mensa-iq-challenge/
5. Mensa Norway is legitimate, correct? - Reddit, 8 月 13, 2025 にアクセス、https://www.reddit.com/r/mensa/comments/11bi2n7/mensa_norway_is_legitimate_correct/
6. Explaining the Mensa Norway IQ Test Through Animations (145+ IQ Answers) - YouTube, 8 月 13, 2025 にアクセス、https://m.youtube.com/watch?v=mPXAUXE3gYg&t=0s
7. How does the Mensa Norway test work? - Reddit, 8 月 13, 2025 にアクセス、https://www.reddit.com/r/mensa/comments/gtwlpo/how_does_the_mensa_norway_test_work/
8. Mensa Norway IQ test : r/iqtest - Reddit, 8 月 13, 2025 にアクセス、https://www.reddit.com/r/iqtest/comments/1gyqa06/mensa_norway_iq_test/
9. Tracking AI: IQ Test, 8月 13, 2025 にアクセス、https://trackingai.org/IQ
10. Claude tops, GPT□5 scores 70 on the “Offline IQ” test — here's the context behind the viral chart : r/ChatGPTPro - Reddit, 8 月 13, 2025 にアクセス、https://www.reddit.com/r/ChatGPTPro/comments/1mlt93z/claude_tops_gpt5_scores_70_on_the_offline_iq_test/
11. GPT-5 in a 100% Private IQ Test by TrackingAI.org : r/singularity- Reddit, 8 月 13, 2025 にアクセス、https://www.reddit.com/r/singularity/comments/1mlhdc7/gpt5_in_a_100_private_iq_test_by_trackingaiorg/
12. ChatGPT-5 Release: Everything You Need to Know and More- Coursiv, 8 月 13, 2025 にアクセス、https://blog.coursiv.io/blog/chatgpt_-5-release-everything-you-need-to-know

13. The benefits and dangers of anthropomorphic conversational agents - PNAS, 8 月 13, 2025 にアクセス、 https://www.pnas.org/doi/10.1073/pnas.2415898122
14. The Anthropomorphism Trap : Our Trust in AI may be creating dangerous blind spots | by Savneet Singh | Medium, 8 月 13, 2025 にアクセス、 https://medium.com/@savusavneet_28467/the-anthropomorphism-trap-our-trust-in-ai-may-be-creating-dangerous-blind-spots-eec4a7824c0c
15. GPT-5 vs IQ Tests: Why AI Fails & What It Really Means - Arsturn, 8 月 13, 2025 にアクセス、 https://www.arsturn.com/blog/gpt-5-vs-offline-iq-tests-what-do-the-low-scores-really-mean
16. Artificial Intelligence Is Just Artificial - Mensa Foundation, 8 月 13, 2025 にアクセス、 https://www.mensafoundation.org/artificial-intelligence-is-just-artificial/
17. The dark side of AI anthropomorphism: A case of misplaced trustworthiness in service provisions - ScholarSpace, 8 月 13, 2025 にアクセス、 https://scholarspace.manoa.hawaii.edu/bitstreams/b6cedcc3-cd5c-4744-bb99-2d8f90b334ec/download
18. Anthropomorphic Seduction: The Siren Song of Human-Like AI - Medium, 8 月 13, 2025 にアクセス、 https://medium.com/@delimiterbob/anthropomorphic-seduction-the-siren-song-of-human-like-ai-8a34ac37b4ce
19. An Overview of Data Contamination: The Causes, Risks, Signs, and Defenses - Holistic AI, 8 月 13, 2025 にアクセス、 https://www.holisticai.com/blog/overview-of-data-contamination
20. A Survey on Data Contamination for Large Language Models - arXiv, 8 月 13, 2025 にアクセス、 https://arxiv.org/html/2502.14425v2
21. Why data contamination is a big issue for LLMs - TechTalks, 8 月 13, 2025 にアクセス、 https://bdtechtalks.com/2023/07/17/llm-data-contamination/
22. AI took a huge leap in IQ, and now a quarter of Gen Z thinks AI is conscious | TechRadar, 8 月 13, 2025 にアクセス、 https://www.techradar.com/computing/artificial-intelligence/ai-took-a-huge-leap-in-iq-and-now-a-quarter-of-gen-z-thinks-ai-is-conscious
23. rank-by-test-source (22).csv
24. OpenAI's o3 scores 136 on Mensa Norway test, surpassing 98% of human population, 8 月 13, 2025 にアクセス、 https://cryptoslate.com/openais-o3-scores-136-on-mensa-norway-test-surpassing-98-of-human-population/
25. 120 IQ AI: Threat or Opportunity? - Peter Diamandis, 8 月 13, 2025 にアクセス、 https://www.diamandis.com/blog/120-iq-ai-threat-or-opportunity
26. ChatGPT 4o explains what chatGpt 5 killed with its arrival : r/ArtificialSentience - Reddit, 8 月 13, 2025 にアクセス、 https://www.reddit.com/r/ArtificialSentience/comments/1mknrdt/chatgpt_4o_explains_what_chatgpt_5_killed_with/
27. OpenAI introduces ChatGPT 5 - Here's all you need to know, 8 月 13, 2025 にアクセス、 https://economictimes.indiatimes.com/magazines/panache/openai-introduces-chatgpt-5-features-performance-access-pricing-heres-all-you-

need-to-know/articleshow/123174283.cms
28. Introducing GPT-5 - OpenAI, 8 月 13, 2025 にアクセス、 https://openai.com/index/introducing-gpt-5/
29. ChatGPT 5 vs ChatGPT 4: What's New in OpenAI GPT-5 - The Complete Comparison, 8 月 13, 2025 にアクセス、 https://www.unleash.so/post/chatgpt-5-vs-chatgpt-4-whats-new-in-openai-gpt-5---the-complete-comparison
30. ChatGPT-5 Full Review: 5 Real-World Tests & The AI Race - YouTube, 8 月 13, 2025 にアクセス、 https://www.youtube.com/watch?v=DbX_0_0LGag
31. ChatGPT 5 vs. GPT-5 Pro vs. GPT-4o vs o3: In-Depth Performance, Benchmark Comparison of OpenAI's 2025 Models - Passionfruit SEO, 8 月 13, 2025 にアクセス、 https://www.getpassionfruit.com/blog/chatgpt-5-vs-gpt-5-pro-vs-gpt-4o-vs-o3-performance-benchmark-comparison-recommendation-of-openai-s-2025-models
32. Kids outsmart leading artificial intelligence models in a simple creativity test, 8 月 13, 2025 にアクセス、 https://www.mensa.org/kids-outsmart-leading-artificial-intelligence-models-in-a-simple-creativity-test/
33. Alfred Binet and the History of IQ Testing - Verywell Mind, 8 月 13, 2025 にアクセス、 https://www.verywellmind.com/history-of-intelligence-testing-2795581
34. What is the actual scientific consensus on IQ? : r/askpsychology - Reddit, 8 月 13, 2025 にアクセス、 https://www.reddit.com/r/askpsychology/comments/1iibiet/what_is_the_actual_scientific_consensus_on_iq/
35. Intelligence, Complexity, and the Failed “Science” of IQ | by Sean McClure | Medium, 8 月 13, 2025 にアクセス、 https://seanamcclure.medium.com/intelligence-complexity-and-the-failed-science-of-iq-4fb17ce3f12
36. What is the mean IQ at Mensa and what is the standard deviation? Is the distribution normal or positively skewed? - Quora, 8 月 13, 2025 にアクセス、 https://www.quora.com/What-is-the-mean-IQ-at-Mensa-and-what-is-the-standard-deviation-Is-the-distribution-normal-or-positively-skewed
37. Position: Evaluating Generative AI Systems Is a Social ... - arXiv, 8 月 13, 2025 にアクセス、 https://arxiv.org/abs/2502.00561
38. [2505.10653] On the Evaluation of Engineering Artificial General Intelligence - arXiv, 8 月 13, 2025 にアクセス、 https://arxiv.org/abs/2505.10653
39. The top AI model is *better at completing IQ tests* than 85% of humans. What a time to be alive! : r/accelerate - Reddit, 8 月 13, 2025 にアクセス、 https://www.reddit.com/r/accelerate/comments/1khtff0/the_top_ai_model_is_better_at_completing_iq_tests/
40. [2505.11831] ARC-AGI-2: A New Challenge for Frontier AI Reasoning Systems - arXiv, 8 月 13, 2025 にアクセス、 https://arxiv.org/abs/2505.11831
41. Some things to know about achieving artificial general intelligence - arXiv, 8 月 13, 2025 にアクセス、 https://www.arxiv.org/pdf/2502.07828

42. Position: Levels of AGI for Operationalizing Progress on the Path to AGI - arXiv, 8 月 13, 2025 にアクセス、 https://arxiv.org/pdf/2311.02462
43. Intelligence | Psychology Today, 8 月 13, 2025 にアクセス、 https://www.psychologytoday.com/us/basics/intelligence
44. How has intelligence testing changed throughout history? - The World Economic Forum, 8 月 13, 2025 にアクセス、 https://www.weforum.org/stories/2015/10/how-has-intelligence-testing-changed-throughout-history/
45. AI Tools in Society: Impacts on Cognitive Offloading and the Future of Critical Thinking, 8 月 13, 2025 にアクセス、 https://www.mdpi.com/2075-4698/15/1/6
46. Thinking in the Age of Machines: Global IQ Decline and the Rise of AI-Assisted Thinking, 8 月 13, 2025 にアクセス、 https://thequantumrecord.com/philosophy-of-technology/global-iq-decline-rise-of-ai-assisted-thinking/
47. (PDF) The psychology of human–AI emotional bonds - ResearchGate, 8 月 13, 2025 にアクセス、 https://www.researchgate.net/publication/393880679_The_psychology_of_human-AI_emotional_bonds
48. Jevin D. West, 8 月 13, 2025 にアクセス、 https://jevinwest.org/publications.html
49. "The benefits and dangers of anthropomorphic conversational agents": r/singularity - Reddit, 8 月 13, 2025 にアクセス、 https://www.reddit.com/r/singularity/comments/1kwnyuw/the_benefits_and_dangers_of_anthropomorphic/