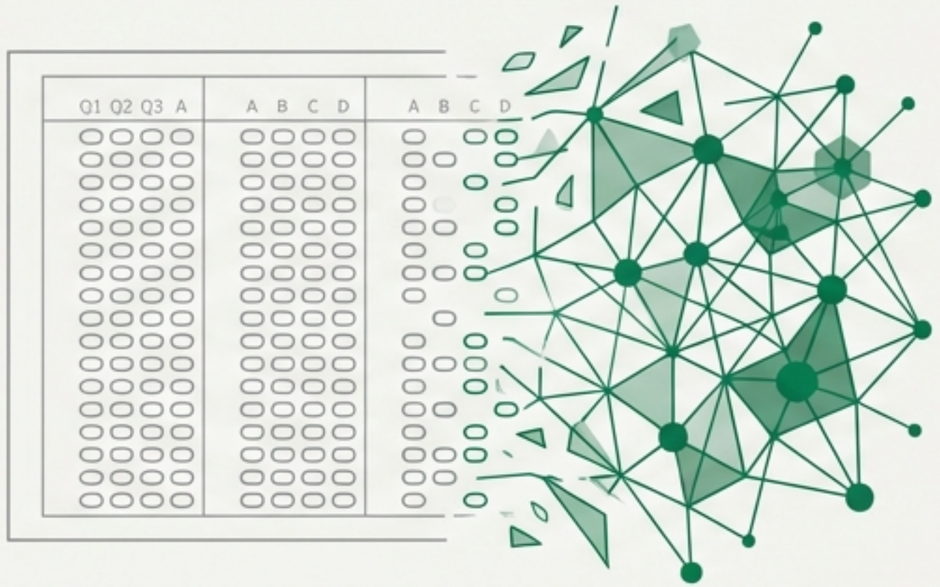
2026年 大学入学共通テストにおけるAIの「超人的」性能評価

GPT-5.2 Thinkingが切り拓く教育と産業の新たな地平



2026年 大学入学共通テストにおけるAIの「超人的」性能評価

GPT-5.2 Thinkingが切り拓く教育と産業の新たな地平

2026年1月 速報レポート

歴史的転換点：AIが得点率97%を記録し、「受験」という枠組みを超越

2026年1月、OpenAIの「GPT-5.2 Thinking」が大学入学共通テストで97%のスコアを達成。これは全受験者の上位0.1%を凌駕する「超越的知能（Superhuman Performance）」の到来を意味する。

97%

GPT-5.2 Thinking (970/1000点) - 超越的知能

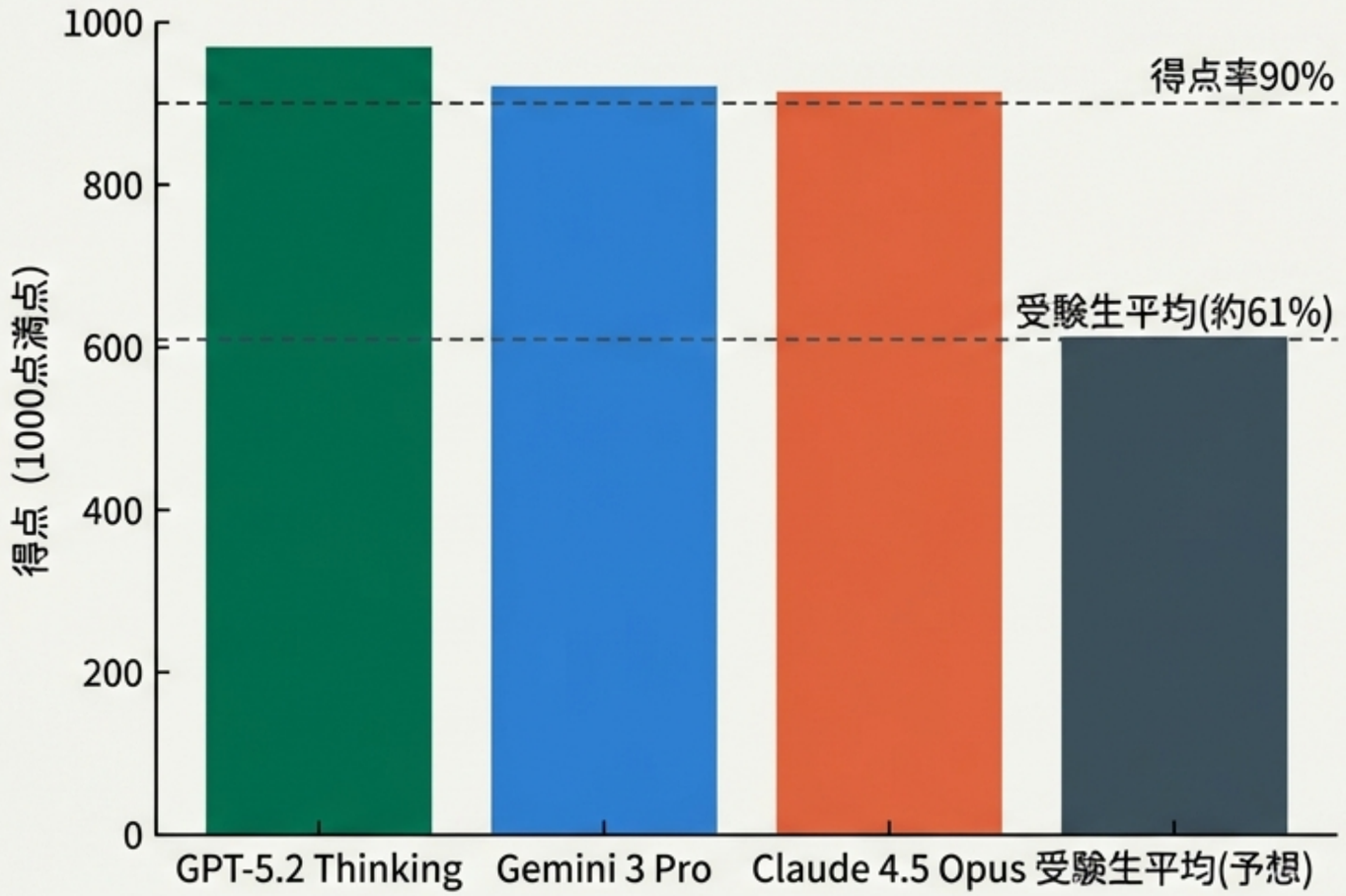
91%

Gemini 3 Pro (919/1000点) - 超難関大合格レベル

91%

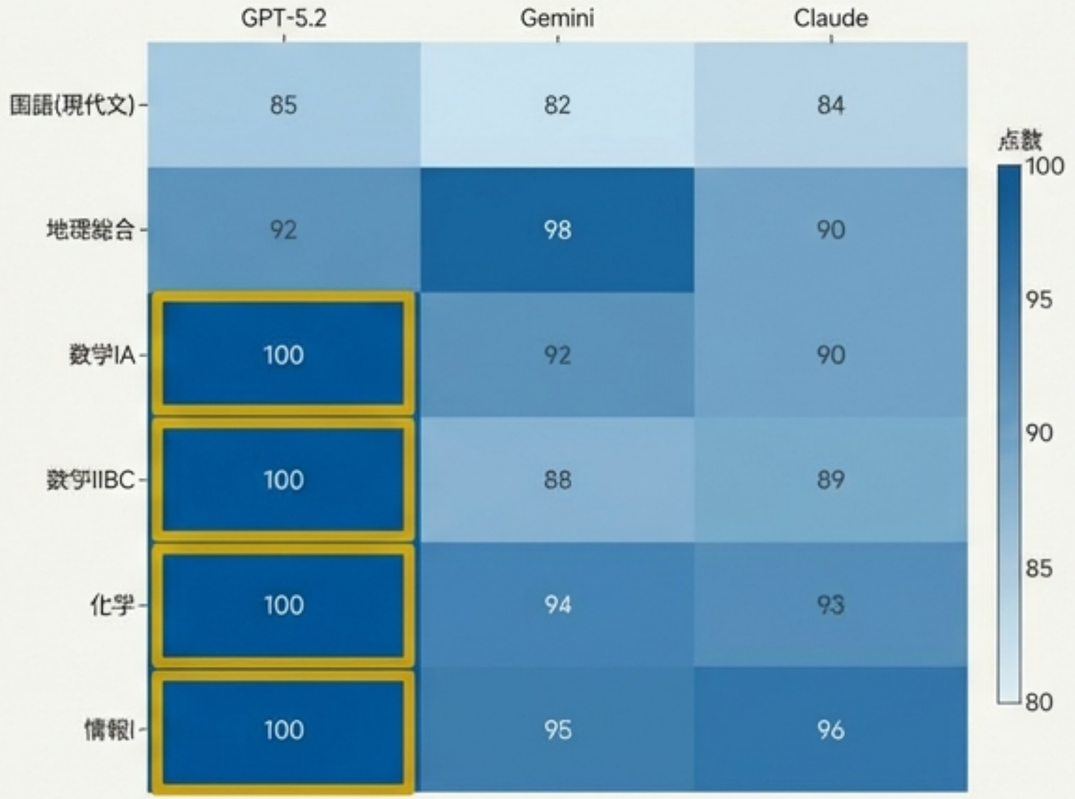
Claude 4.5 Opus (913/1000点) - 超難関大合格レベル

6ポイントの絶対的な壁：上位モデル間の階層構造



- 人間平均（約60%）に対し、全てのトップティアAIが90%オーバーを記録。
- しかし、97%のGPT-5.2と、91%台のGemini/Claudeの間には、統計誤差を超えた明確な性能差が存在する。
- 2024-2025年の「性能拮抗時代」は終わり、推論特化型アーキテクチャが新たなリードを築いた。

数理・科学・情報の完全制覇



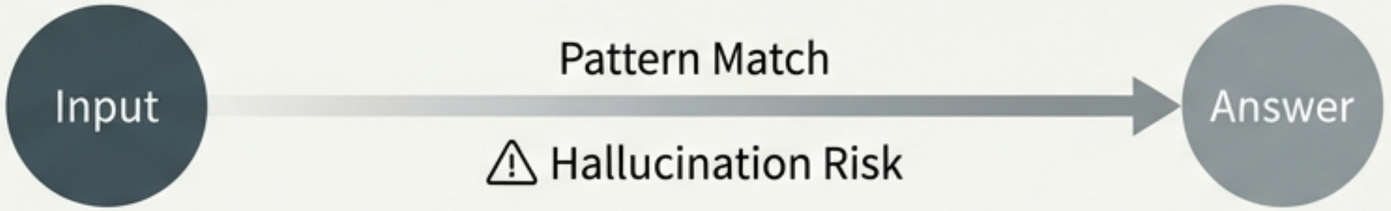
- GPT-5.2 Thinkingが満点（100点）を獲得した9科目：
  - 数学IA / 数学IIBC
  - 物理基礎 / 化学 / 化学基礎 / 地学基礎 / 生物基礎
  - 情報I / 公共、政治・経済

従来のLLMが苦手としていた「論理推論」や「計算プロセス」を含む科目で失点がゼロになったことが、97%達成の直接的な要因である。

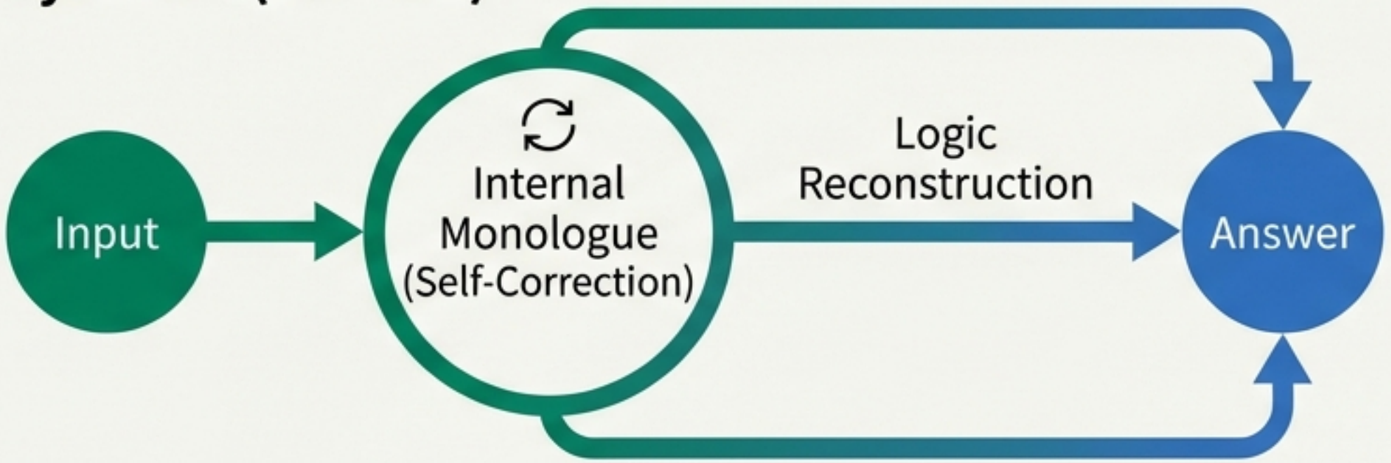
Data sources: Denfaminicogamer, ReseMom

確率的予測から「思考（Thinking）」プロセスへ

System 1 (Old LLM)



System 2 (GPT-5.2)



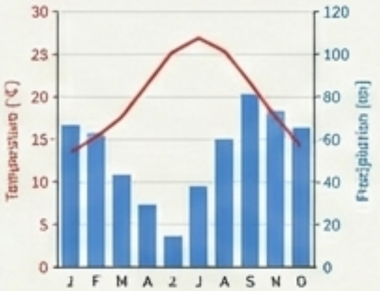
- 図形問題をピクセルではなく「座標データ」や「幾何学的制約」として再構築（Reconstruction）。
- 自己検証（Self-Correction）により、計算ミスや論理の飛躍を回答出力前に修正。
- これが数学・理科における「ハルシネーション（もっともらしい嘘）」の根絶に繋がった。

Gemini 3 Proの特異点：マルチモーダル地理推論

Case Study: 地理総合「アンデス山脈」問題

地図画像
(Visual)

地理空間知識
(Knowledge)



正しい雨温図の導出
(Inference)

- Challenge: 地図上の位置と気候グラフの紐付け。
- Gemini's Success: 地図画像から「アンデス山脈」を視覚的に識別 -> 標高と気候区分の知識を検索 -> 正しい雨温図を選択。
- Insight: Google Maps/Earthの地理空間データと知識ベースが高度に統合されている。視覚と知識の融合領域では、特定の□のドメインでGPT-5.2を上回る可能性がある。

残された3%の壁（1）：感情の機微と「情動」の欠落

領域：国語（小説問題）

Issue: 登場人物の**「言外のニュアンス」「皮肉」「葛藤」**の理解で失点。

Root Cause: AIは「情報（Information）」を処理できても、「**情動**（Affect）」を主観的に体験していない。

Symbol Grounding: 「**悲しい**」という言葉の定義は知っているが、身体的な痛みを伴う**悲しみの感覚（クオリア）**を持たないため、共感的な読解において人間のような直感が働かない。

残された3%の壁（2）：視覚的論理の断絶「バス問題」

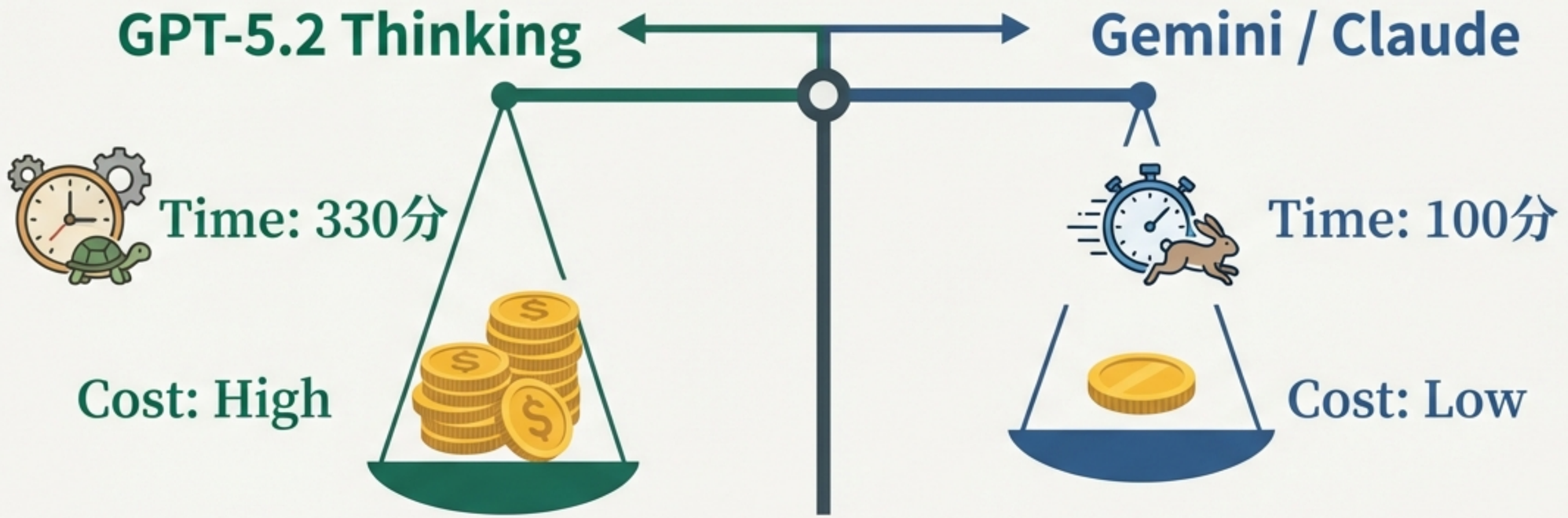
Case Study: 英語リスニング・バスの乗降ルール

Text Understanding: ✅
“Enter from the rear, exit from the front”
(テキストとして正しく認識・書き起こし可能)。

Visual Choice: ❌
誤った矢印のイラストを選択。

Implication:「言語的な論理理解」と「視覚的なシンボル操作」の間に深刻なギャップが存在する。自動運転やロボティクス応用におけるリスク要因となり得る。

「思考」の代償：精度とコスト・時間のトレードオフ



Strategic Takeaway:
Critical Tasks (Legal/Medical) -> Use GPT-5.2 (High Accuracy / High Cost).
Routine Tasks (Summary/Email) -> Use Gemini/Claude (High Speed / Low Cost).
適材適所の使い分けが2026年のAI活用の鍵。

ハードウェアの制約：熱暴走とコンテキスト脱落

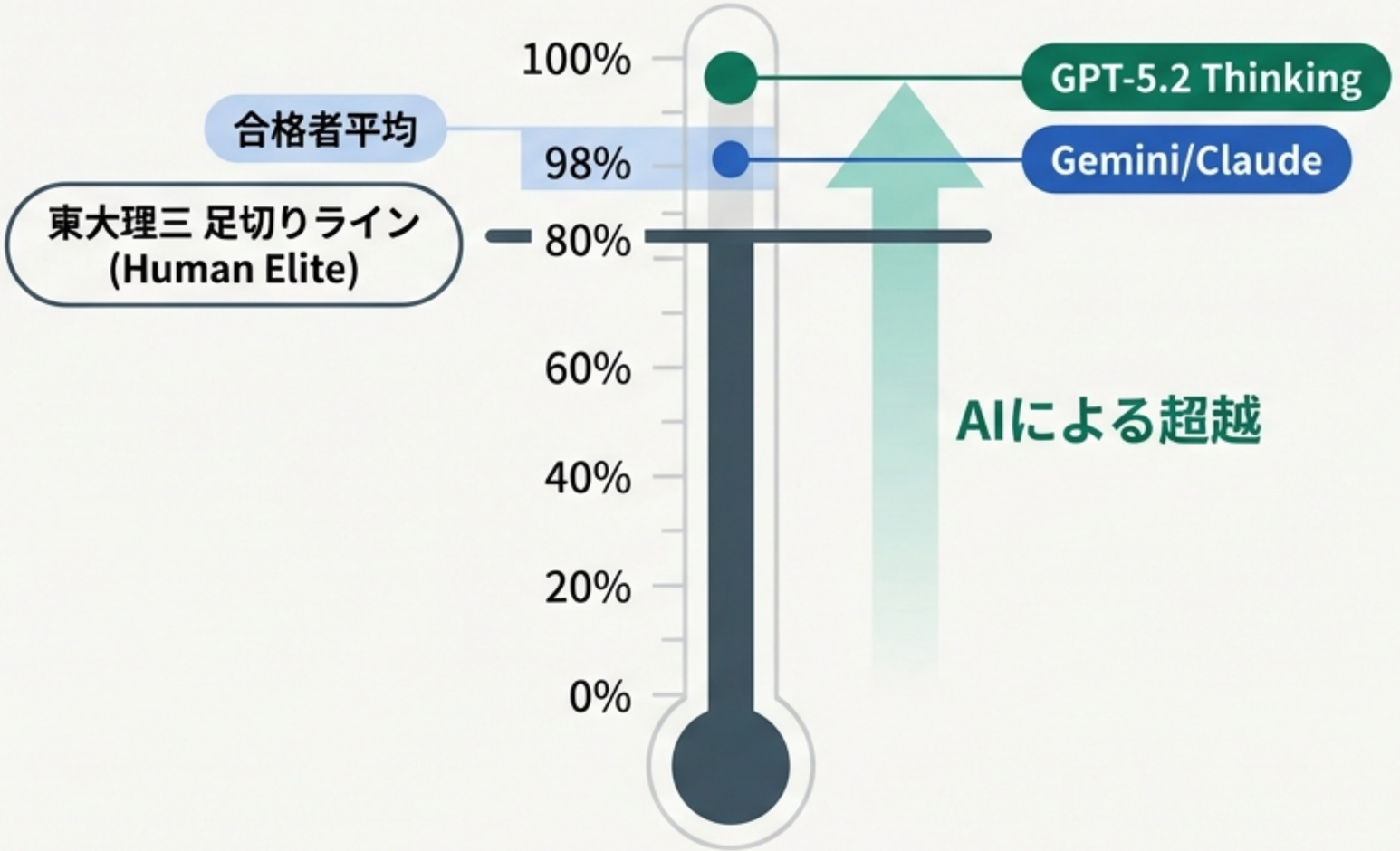
Issue: Gemini 3 Pro (On-Device)

Phenomenon:

- 長時間稼働によるデバイス温度上昇。
- 結果としての「**コンテキスト脱落（Context Drop）**」：文脈忘却や論理不整合の発生。

Insight: クラウド上のベンチマークは優秀でも、実環境（スマホ・PC）での安定稼働には**排熱と軽量化**という物理的な課題が残されている。

「東大理三」の足切りライン崩壊



マークシート方式の試験において、AIは人間を完全に凌駕した。
「知識の再生」を測る入試システムは、もはや人間能力の測定尺度として機能しない。

Modern Academic Editorial

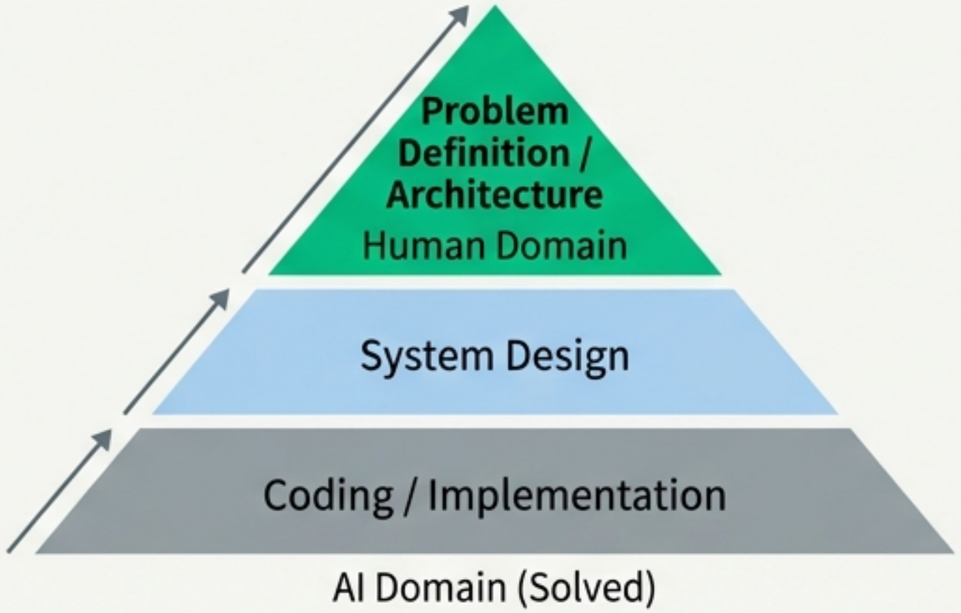
教育の再定義：「知識」の暴落と「検証力」の高騰

Old Skill: Answering (再生)

New Skill: Auditing（検証）

Before: 正解を暗記し、再生する能力（AIが代替可能）。
After: 検証力（Audit Capability）。AIが出す**97%の正解**を疑い、残りの**3%の致命的なミスミス（バス問題など）**を発見する能力。
New Curriculum: 「正解を出す訓練」から「AIオーディティング（監査）教育」へ。

産業への衝撃：情報I「満点」が示すエンジニアリングの未来



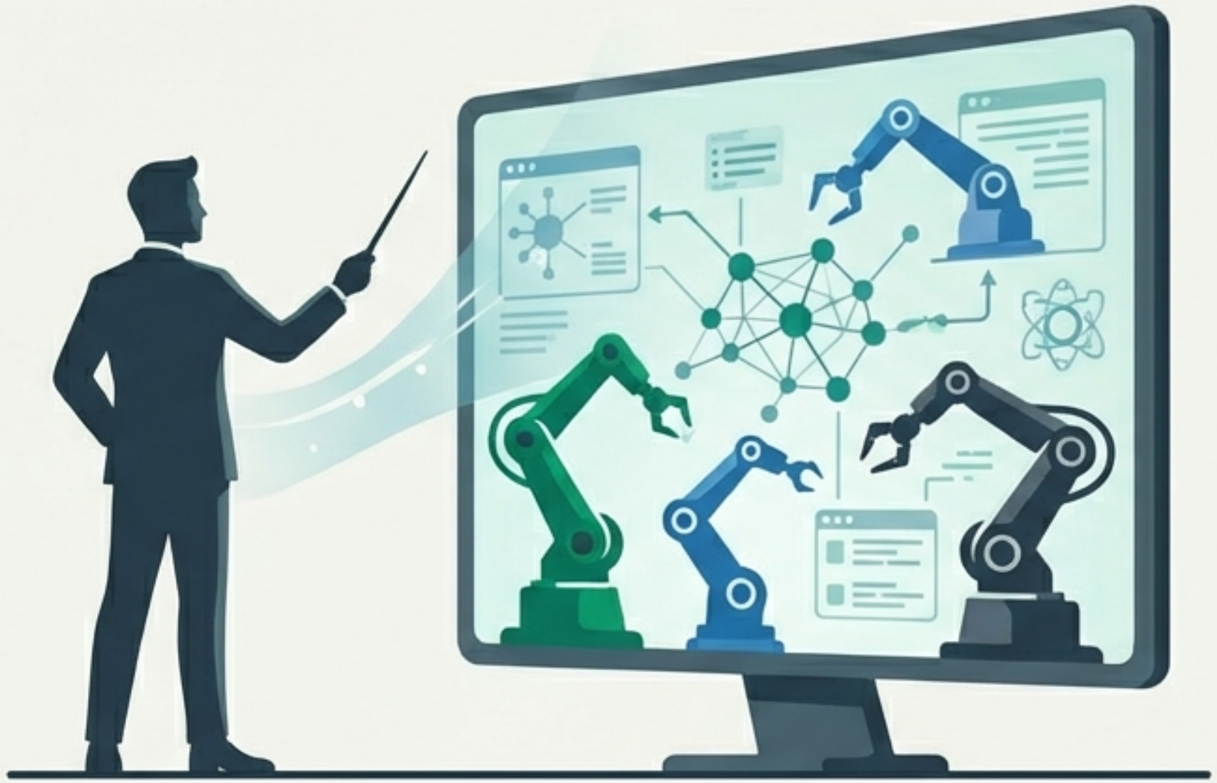
Result: GPT-5.2は「情報I」で満点を獲得。プログラミング、アルゴリズム、データ構造を完全に理解。

Implication:

- 初級〜中級のコーディング実装はAIが代替・支援可能。
- 人間は「コードを書く」ことよりも、「何を解決すべきか（課題設定）」と「システムアーキテクチャ（設計）」に集中する必要がある。

Modern Academic Editorial

結論：人間は「回答者」から「監督者」へ



2026年、AIは「回答者（Answerer）」としての能力で人間を超えた。

- 我々の新しい役割は、AIという強力な知能リソースを指揮する「監督者（Director)」である。
- System 1（瞬発力）はAIに任せ、人間はSystem 2（思考体力）を鍛え直し、問いを立てる力で勝負する時代が到来した。

Modern Academic Editorial

参考文献・データソース

- 株式会社LifePrompt 検証レポート（2026年1月20日）
- 日本経済新聞: "OpenAIは9科目満点"
- Denfaminicogamer: "GPT-5.2 vs Gemini vs Claude Comparison"
- Resemom: "AI共通テスト成績分析"
- LifePrompt Media: "GPT-5.2 Thinking Deep Dive"
- Note: Timing metrics (330min vs 100min) based on LifePrompt internal benchmarks.