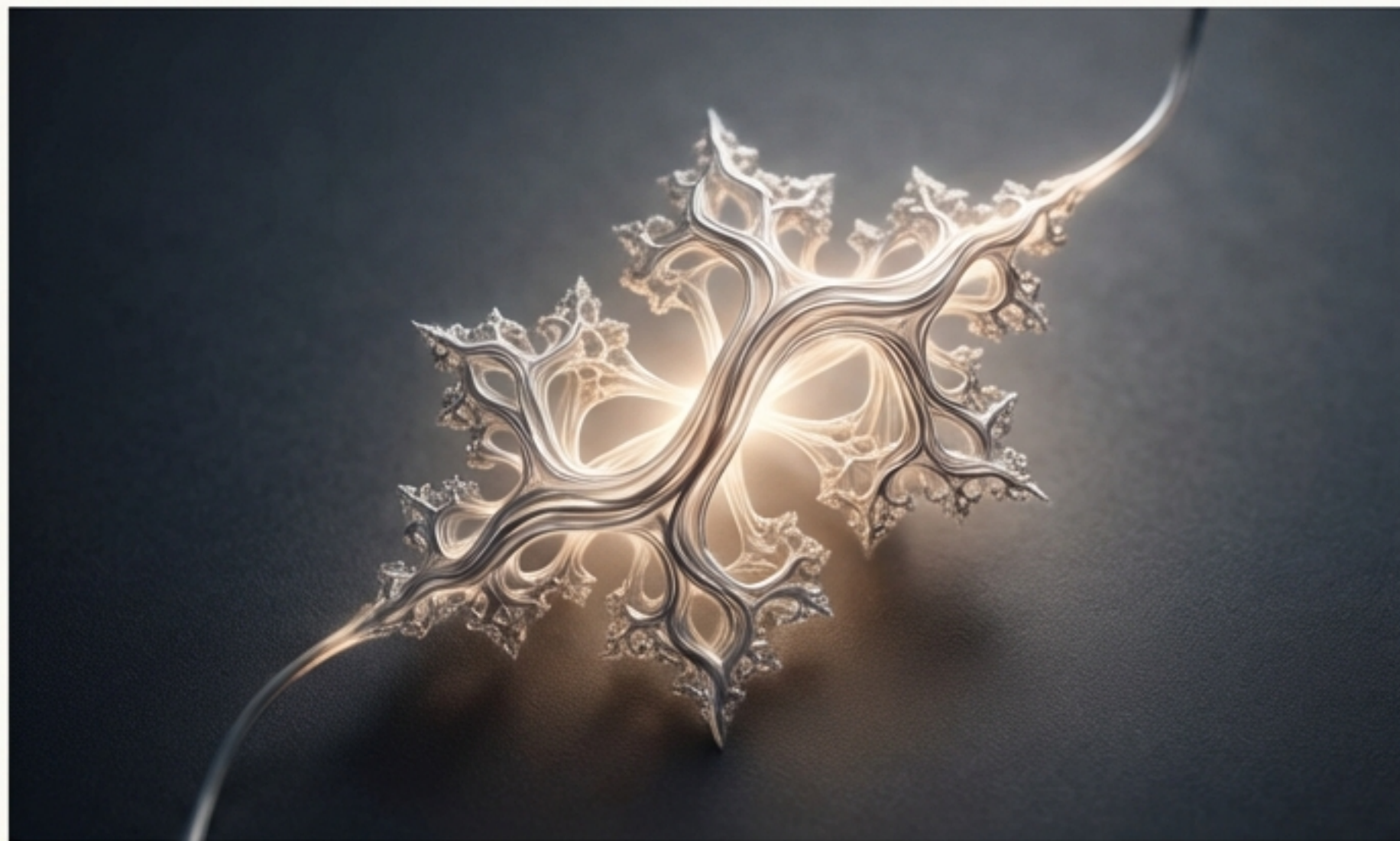


発明「創出」を加速するAI：Gemini 3 Pro vs GPT-5.1 Pro 徹底比較

「天才的研究パートナー」か「究極の熟練労働者」か。

2025年11月時点 最新分析レポート



結論：発明の「創出」フェーズでは、 Gemini 3 Proが決定的な優位性を持つ



**「天才的な研究パートナー」
Gemini 3 Pro**

創造性、抽象的推論、仮説生成に
卓越し、「0から1」を生み出す。



**「究極の熟練労働者」
GPT-5.1 Pro**

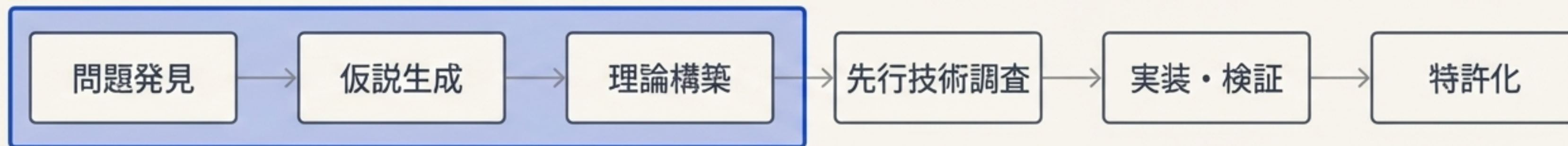
安定性、実装力、再現性に優れ、
「1から100」を確実に行う。

本レポートは、この本質的な違いがイノベーションの最前線においてなぜ重要なのかを、
定量的・定性的エビデンスに基づき解き明かす。

なぜ「創出」が重要なのか？

発明プロセスにおけるAIの必須能力

「創出フェーズ」 (Creation Phase)

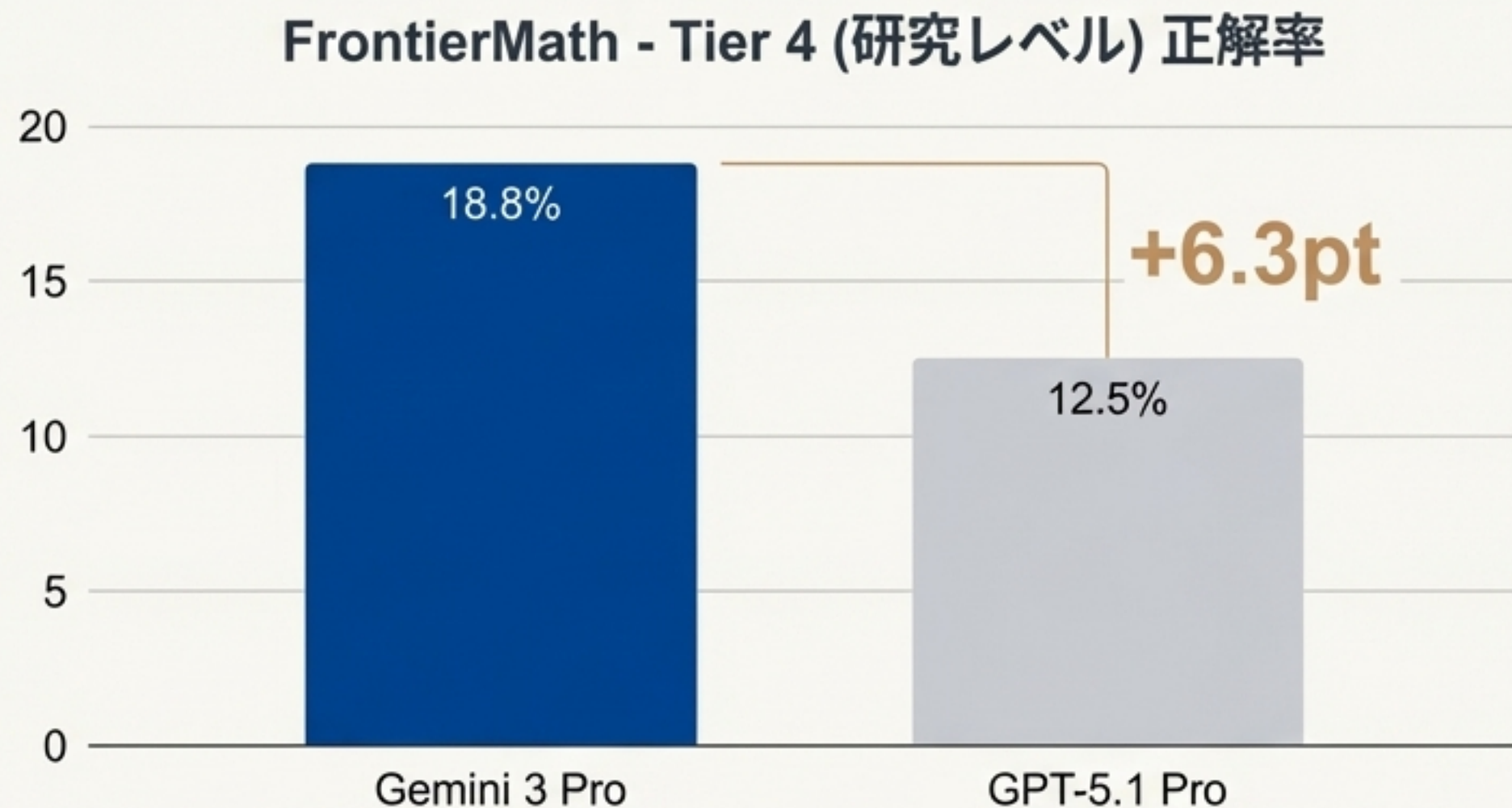


「創出」フェーズでAIに求められる5つの必須能力：

1. 抽象的思考能力: 未知の問題に対する本質的な理解
2. 仮説生成: 科学的直感に基づく新しいアイデアの提案
3. 探索的アプローチ: 既存パターンに囚われない創造性
4. 深い推論能力: 複雑な理論を構築するための多段階思考
5. マルチモーダル理解: 図面、データ、物理現象の正確な解釈

数学的創造性の証明：未知の数学問題において Gemini 3 Pro が圧勝

FrontierMath: フィールズ賞受賞者が設計に関与した、未公開の研究レベル数学問題で構成。AIの真の数学的創造力を測定する。

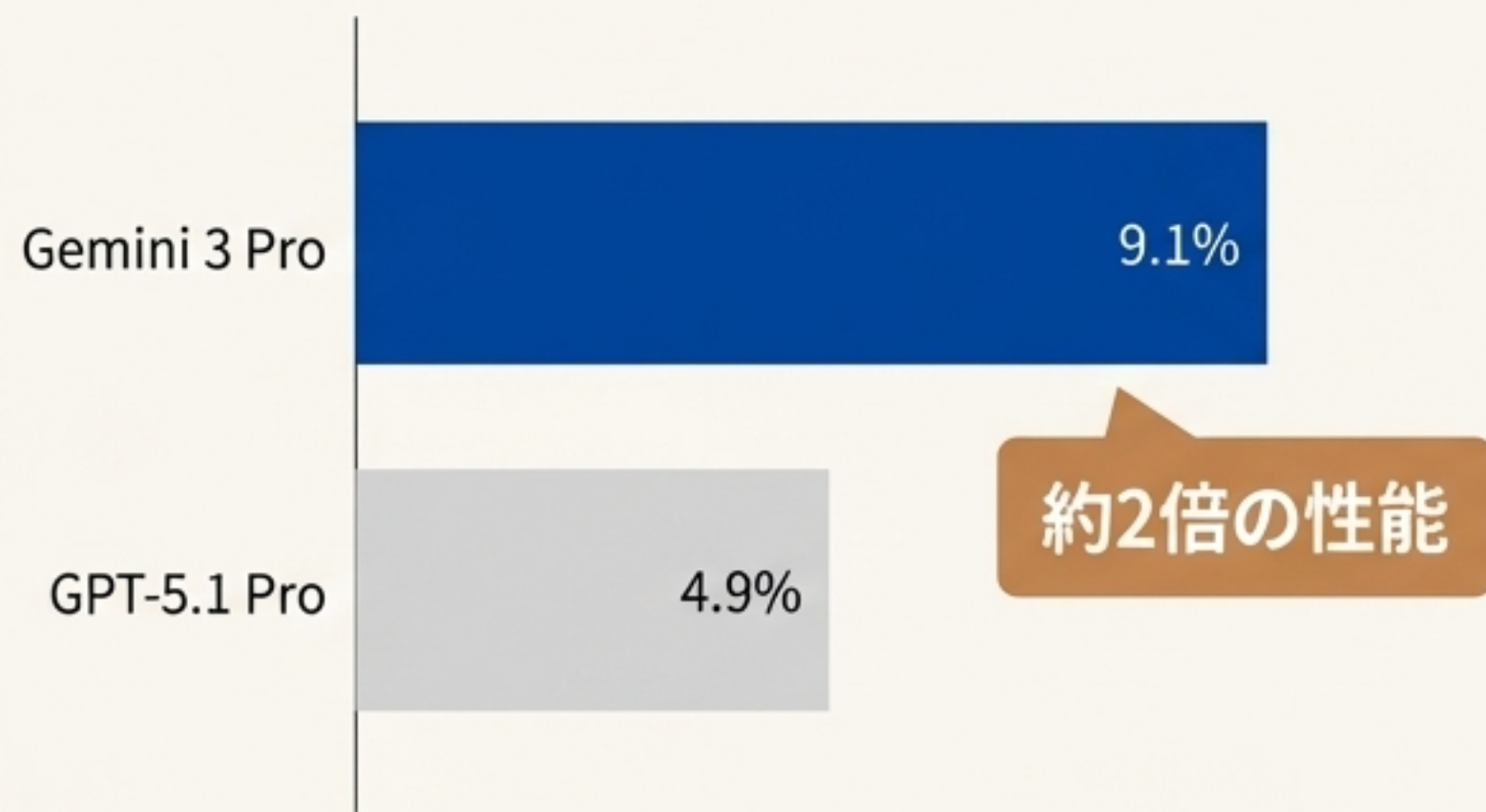


So What?

この差は、単なる計算能力ではなく、未知の数学的構造を発見し、新しい証明を創造する能力の差を示す。

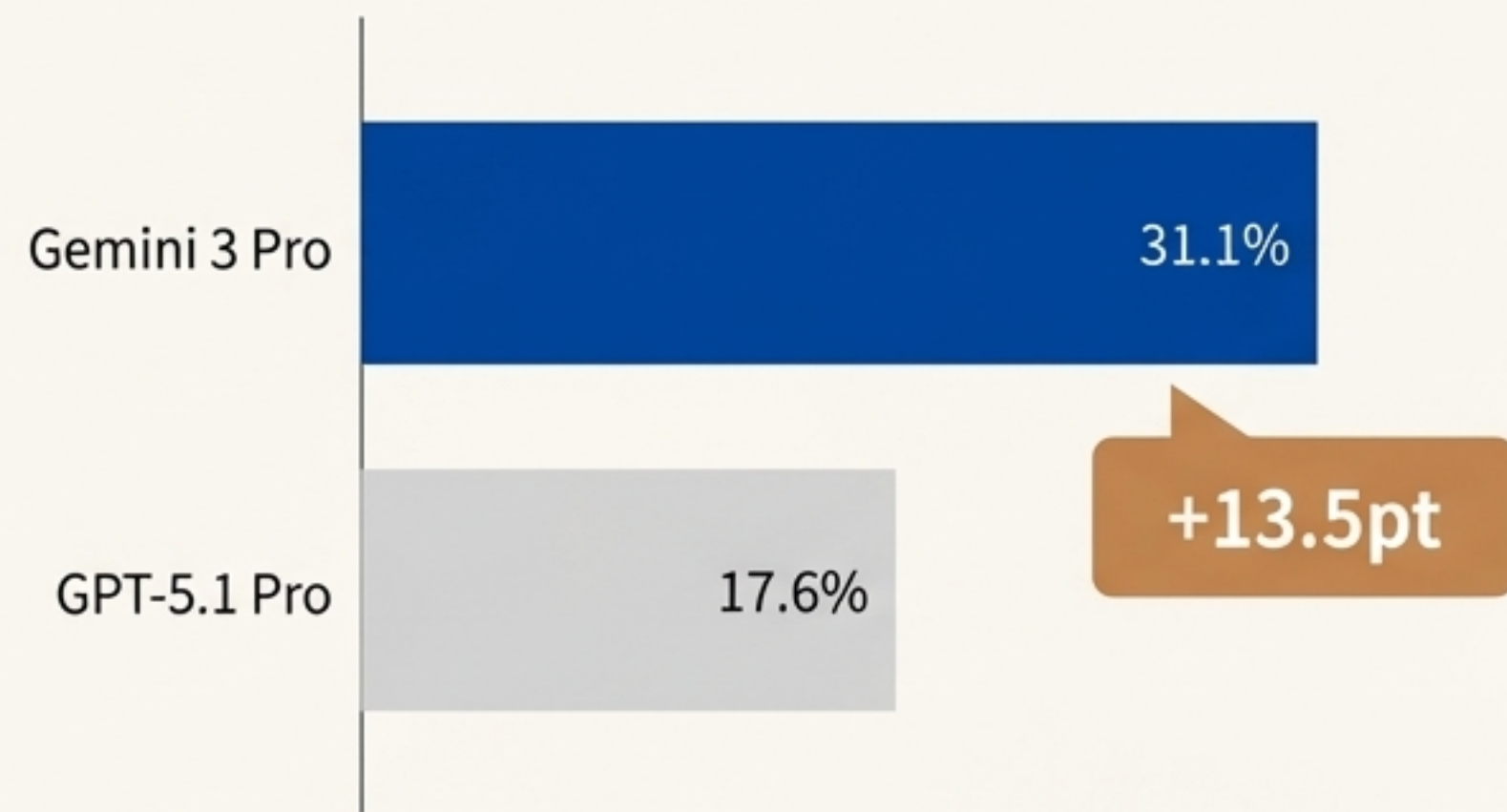
物理的直感と抽象的思考：AIの「ひらめき」を測る最重要指標

CritPt: 物理学研究における推論能力



真の物理的理解と推測困難な問題解決が要求される領域での明確な優位性。

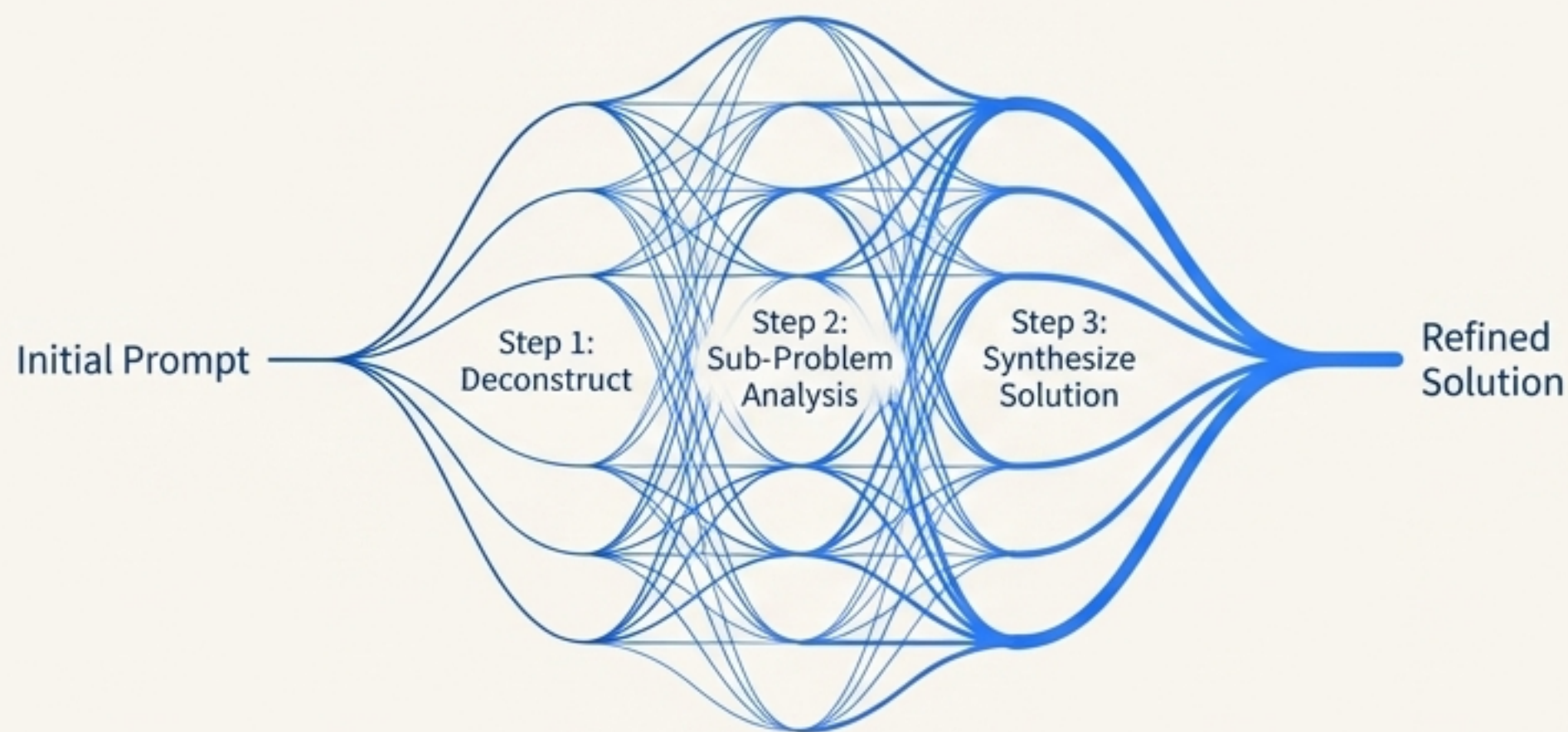
ARC-AGI-2: 抽象的推論と新規問題への適応



訓練データにないパターンを発見する能力。これこそが「0から1」を生み出す発明の核である。

「深く考える」能力：Deep Thinkモードが拓く、科学的推論の新次元

Deep Thinkモードは、複雑な問題に対し計算リソースを動的に配分し、多段階の思考プロセスを経て解答を導き出す。



Deep Thinkの性能高

ARC-AGI-2:

31.1% → **45.1%** (+14.0pt)

GPQA Diamond (大学院レベル科学):

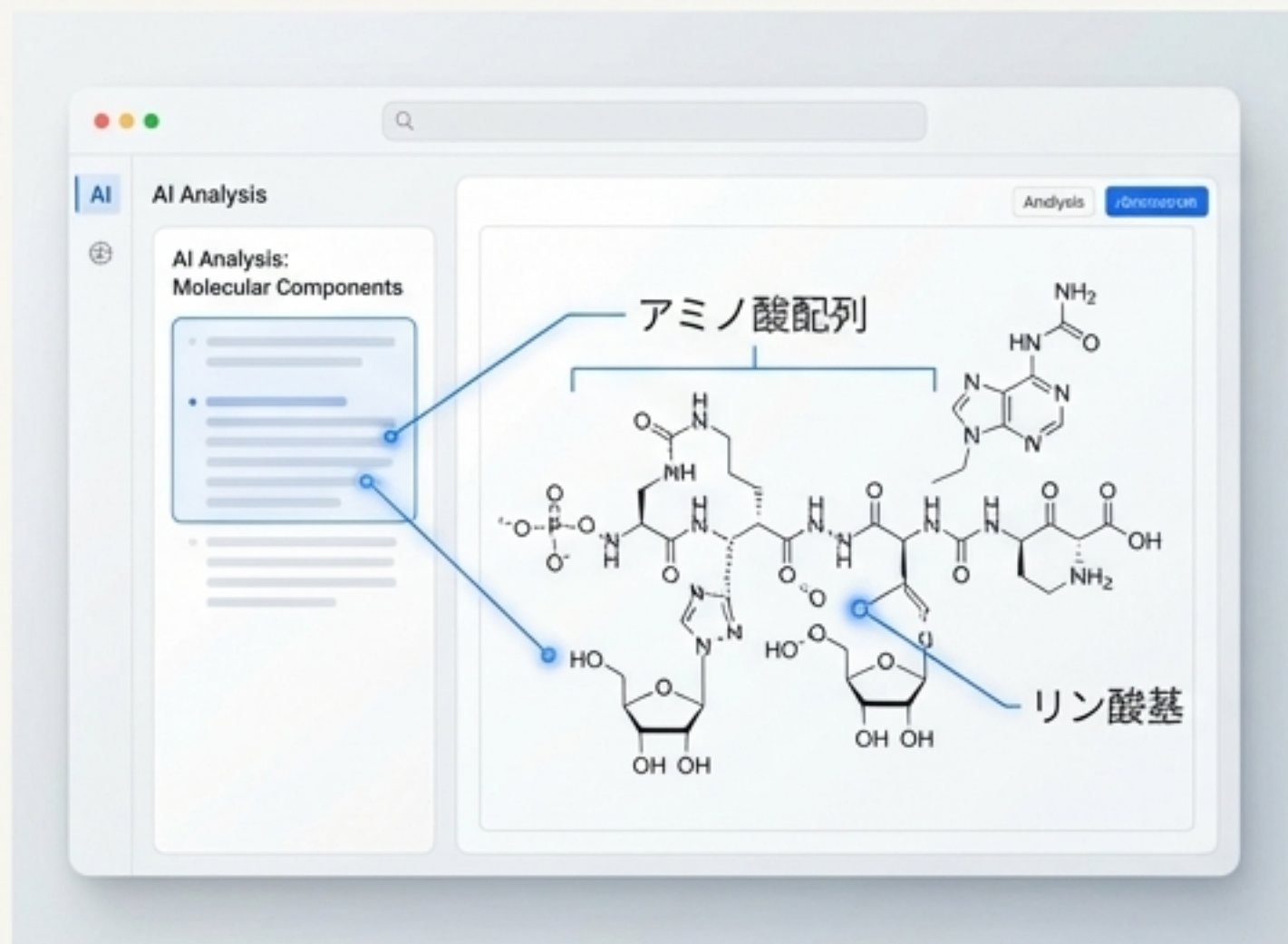
91.9% → **93.8%**

Humanity's Last Exam:

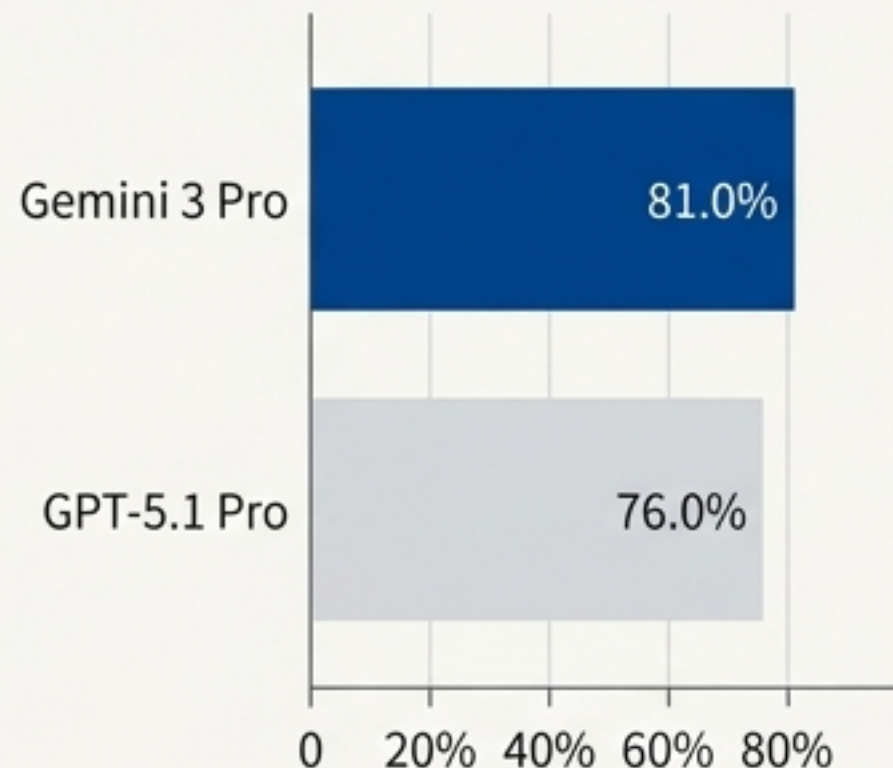
37.5% → **41.0%**

これは単なる性能向上ではない。最も困難な科学的問題に取り組むために設計された、質的に異なる推論能力である。

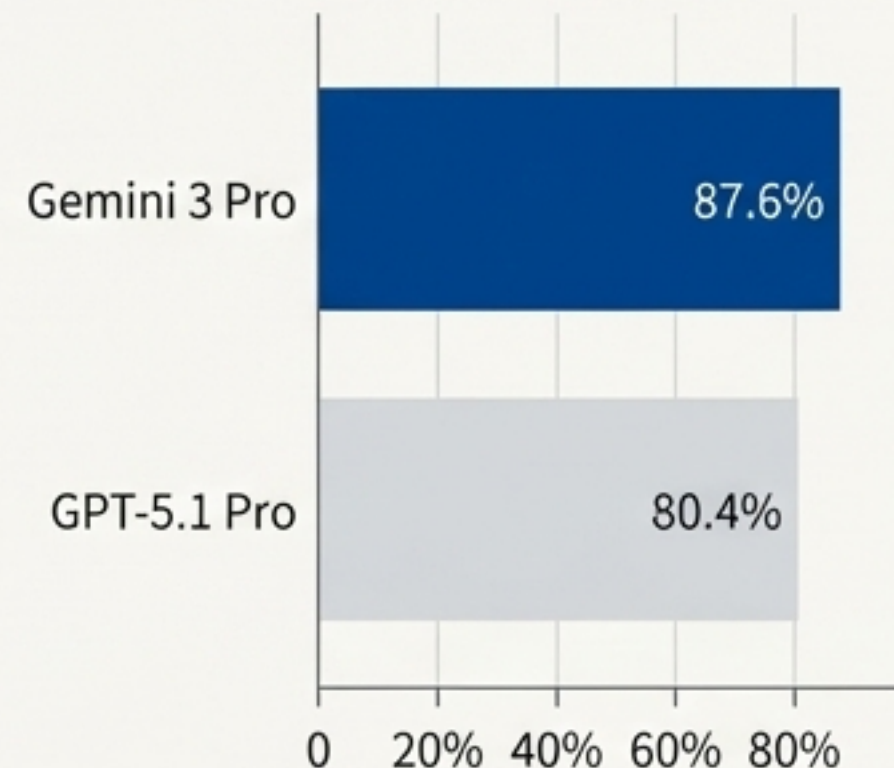
図面、データ、現象を理解する「眼」：マルチモーダル能力の決定的な差



MMMU-Pro: 専門的画像理解



Video-MMMU: 動的現象理解



Qualitative Insight

TrueFoundryの分析によれば、Gemini 3 Proは画像の細部に注意を払う「視覚的な丁寧さ」を持ち、ハルシネーションが少ない。これは、技術図面や実験結果の正確な解釈に不可欠な特性である。

100万トークンの知の地平線：膨大な先行技術を一度に読み解く力

GPT-5.1 Pro
128k Tokens



GPT-5.1 Pro
128k Tokens

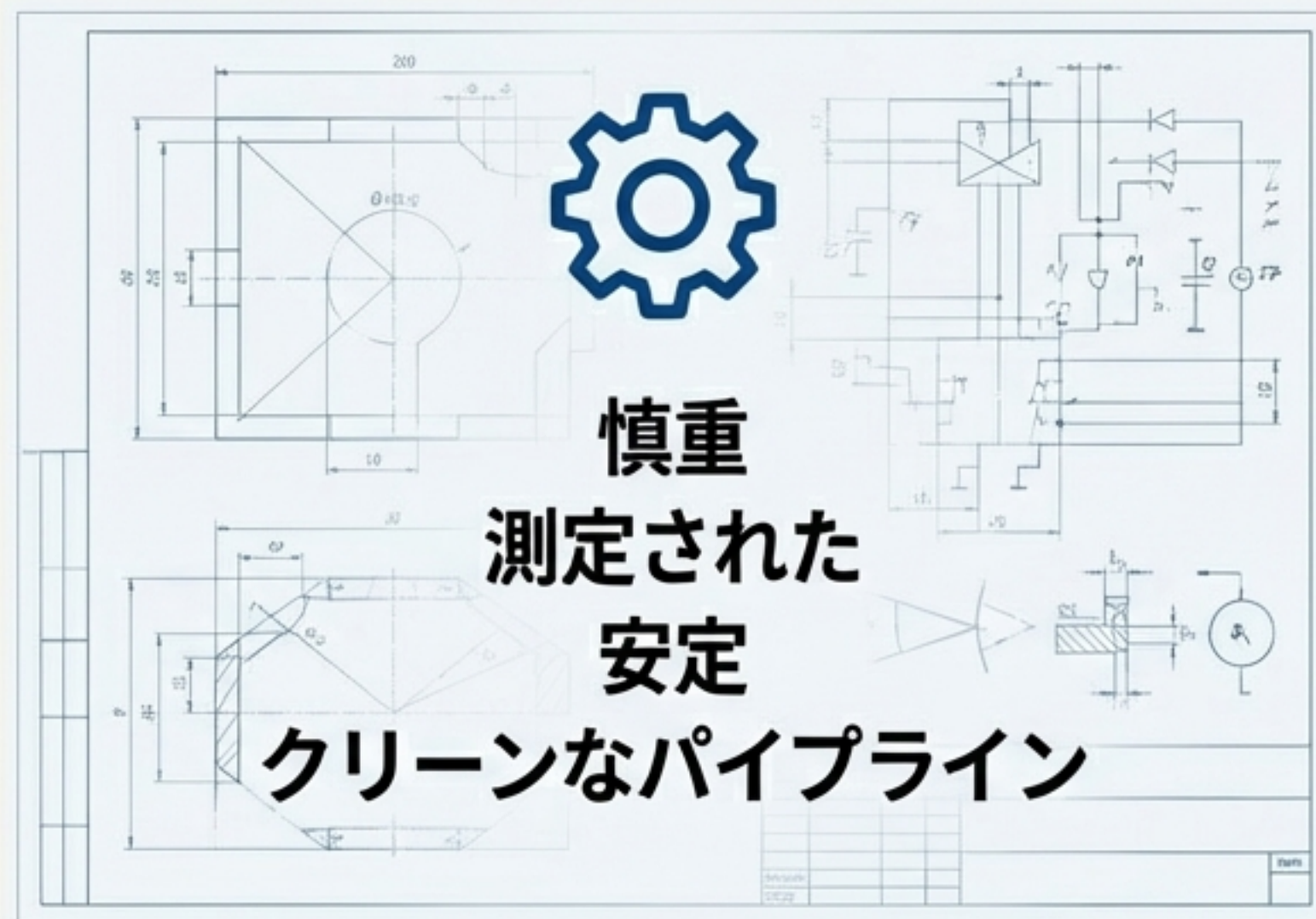
Gemini 3 Pro / 1,048,576 Tokens

- **What it means:** GPT-5.1 Proの約8倍の情報を一度に処理可能。数千ページの特許文献や科学論文をまとめて分析できる。
- **Why it matters:** CometAPIの分析で示された通り、長文脈での精度低下（Context Rot）を最小限に抑制。これにより、異なる文献間の新規な関連性発見や、包括的な技術動向の把握が可能になる。

「天才的研究パートナー」 vs. 「究極の熟練労働者」



「直感的な答えを素早くスケッチし、それを正当化しようとする。正しいときは素晴らしく見える。」 - Source: TrueFoundry

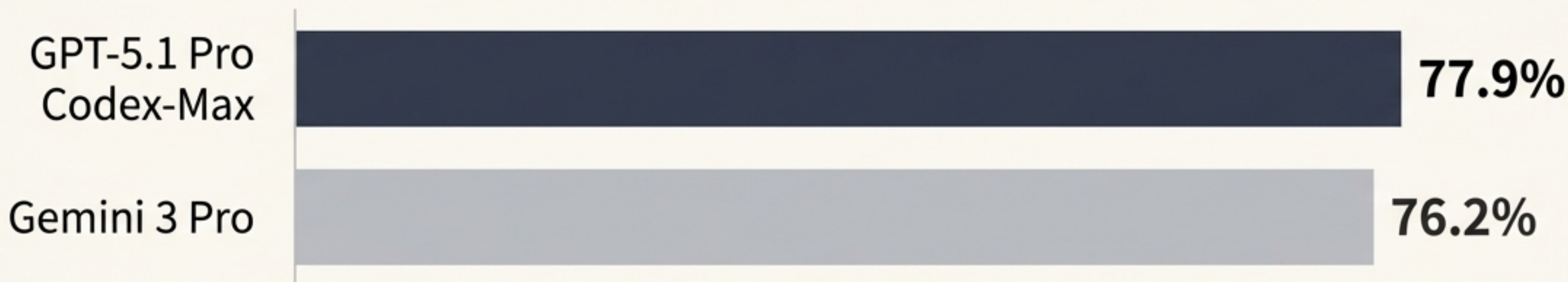


「注意深いTAのように変数を追跡し、最もクリーンな『読む→要約する→推論する』パイプラインを持つ。」 - Source: TrueFoundry

アイデアを形にする「実装」フェーズではGPT-5.1 Proが優位

確立された概念を、安定した高品質なコードや設計に落とし込むタスクでは、GPT-5.1 Proの慎重で測定されたアプローチが理想的である。

****SWE-bench Verified****（実世界のGitHub Issue解決能力）



「GPT-5.1は特にコーディング安定性、ツール使用の一貫性において競争力を持ち、予測可能性が重要な本番ワークロードにとって信頼できる選択肢である。」 - Source: Portkey

コスト効率がイノベーションを加速：Gemini 3 Proは最大9.6倍の費用対効果

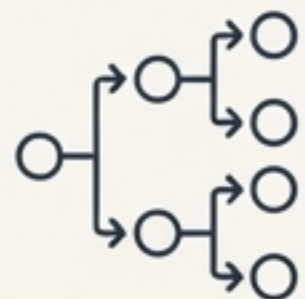
100kトークン入力 / 100kトークン出力のR&Dタスクコスト



発明は、多数の仮説検証と反復的なアイデア改善を必要とするプロセスである。この圧倒的なコスト差は、同じ予算内でより多くの探索を可能にし、最終的にブレークスルーの可能性を高める。

科学的発見の未来像：DeepAgentやAI Scientistが示す方向性

最先端の科学発見AIは、自律的な仮説生成、新規ツール発見、データからの理論導出といった能力を必要とする。



DeepAgentの自律的仮説生成



**CritPt/GPQAで証明された
Gemini 3 Proの科学的直感**

? $\rightarrow$ $\begin{cases} ? = \sum \frac{n_i}{v} x_i^2 \\ f(x)^2 = \pi^\infty \\ x = -\frac{a-b}{2} \end{cases}$

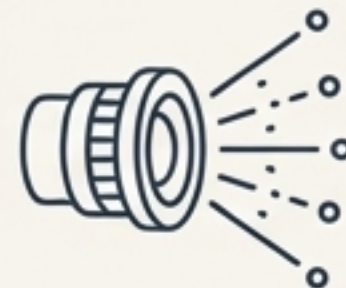
新規数学手法の発見



**FrontierMath/ARC-AGI-2で
証明された未知への適応能力**



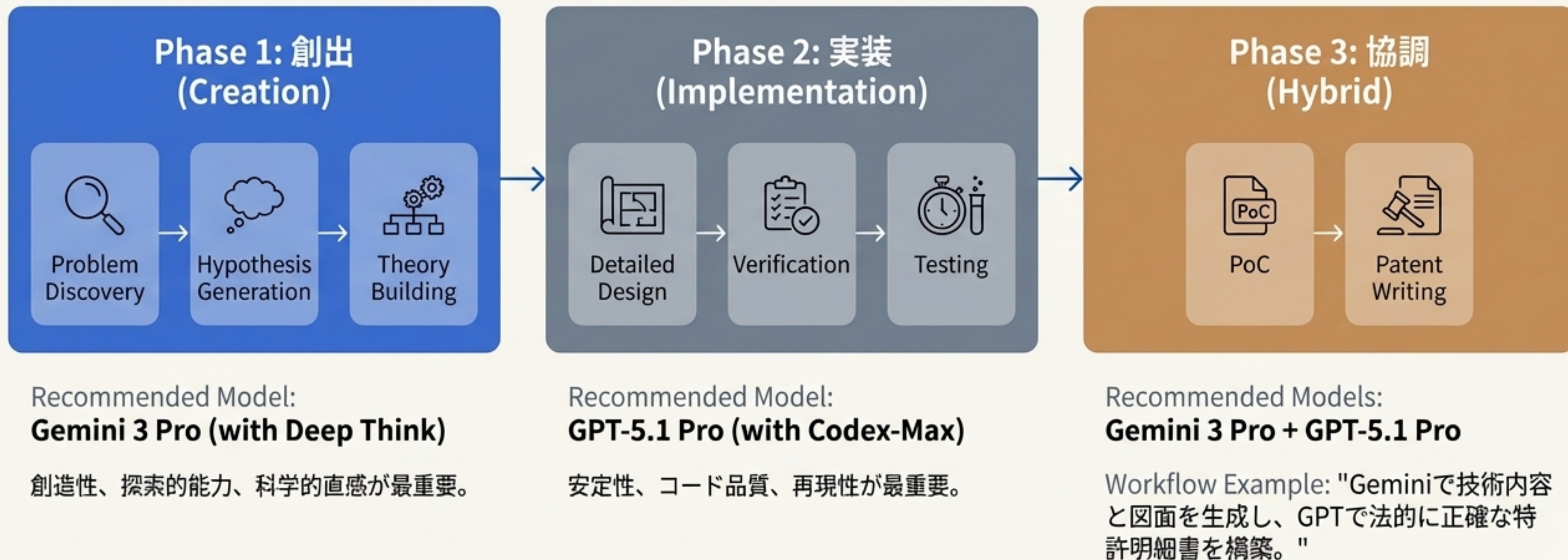
PhyE2Eのデータからの理論導出



**MMMU/1Mトークンによる大量
データからのパターン発見能力**

Gemini 3 Proの強みは、未来の科学研究基盤と完全に一致している。

発明プロセス全体を最適化するハイブリッド・ワークフロー



発明の最前線に立つなら、選ぶべきは「天才」である

- **科学的・創造的推論で圧倒:** FrontierMath, CritPt, ARC-AGI-2といった発明の核心能力を測るベンチマークでGPT-5.1 Proを凌駕。
- **現実世界の複雑さを理解:** 1Mトークンのコンテキストと最高クラスのマルチモーダル能力で、膨大な文献や複雑な図面を正確に処理。
- **発明の本質に合致:** 「大胆で探索的」な特性は、不確実性の中に答えを探す「0から1」のプロセスそのものに適合する。

ブレークスルーを追求する研究開発チームにとって、Gemini 3 Proは不可欠なパートナーである。発明プロセス全体を最適化するには、GPT-5.1 Proと組み合わせたハイブリッド戦略が最も効果的だ。

Appendix: 主要ベンチマークとデータソース

主要ベンチマーク (Key Benchmarks)

FrontierMath (Epoch AI): 未公開の研究レベル数学問題

CritPt (Complex Research using Integrated Thinking - Physics Test): 研究レベル物理学問題

ARC-AGI-2 (Abstraction and Reasoning Corpus - AGI v2): 抽象的推論と汎化能力

MMMU-Pro / Video-MMMU: 専門的・動的マルチモーダル理解

SWE-bench Verified: 実世界のソフトウェアエンジニアリング能力

主な分析ソース (Key Analysis Sources)

Google AI / DeepMind (Official Announcements)

Epoch AI (Benchmark Provider)

TrueFoundry (Qualitative Comparison)

Portkey.ai (Performance & Stability Analysis)

AceCloud (Cost-Effectiveness Analysis)

CometAPI (Long-Context Performance)