

自律的発明の境界線

孫正義氏のAGIレベル4定義に基づく Gemini 3 Proの包括的評価と技術的特異点の検証



2025年11月、AIは「知識労働者」から「発明者」へと質的な転換を遂げたのか？

本稿では、業界の巨人が提示した「定規」を用い、その真偽を検証する。

評価の「定規」：孫正義氏が描くAGIからASIへの進化論

孫氏は、単なるタスク処理能力ではなく、AIが経済と社会にと与えるインパクトを基軸とした5段階の進化モデルを提唱。真の目標は人間の知能を1万倍凌駕するASIであり、その鍵は「発明」にある。



Level 1: 会話能力 (Conversationalist)

- 人間と区別がつかない対話 (Achieved 2023-2024)



Level 2: 博士号レベル (Polymath)

- 複数の専門分野で人間を凌駕



Level 3: エージェント (Agent)

- 人間の代理としてタスクを遂行



Level 4: 発明者 (Inventor)

- AI自体による発明・発見 (Focus of this report)



Level 5: 組織的活動 (Society)

- AI同士の協調と組織化

レベル4の定義的特異性：「0から1」を生み出す能力

孫氏の文脈における「発明」とは、既存知識の組み合わせではない。それは以下の3つの能力を要する、新たな論理構造の創出である。



1. アブダクション (仮説形成的推論)

観測データから、最も尤もらしい「新しい法則」を導き出す能力。



2. 未知への適応

学習データにないパターンに対し、新しい枠組みを構築する能力。



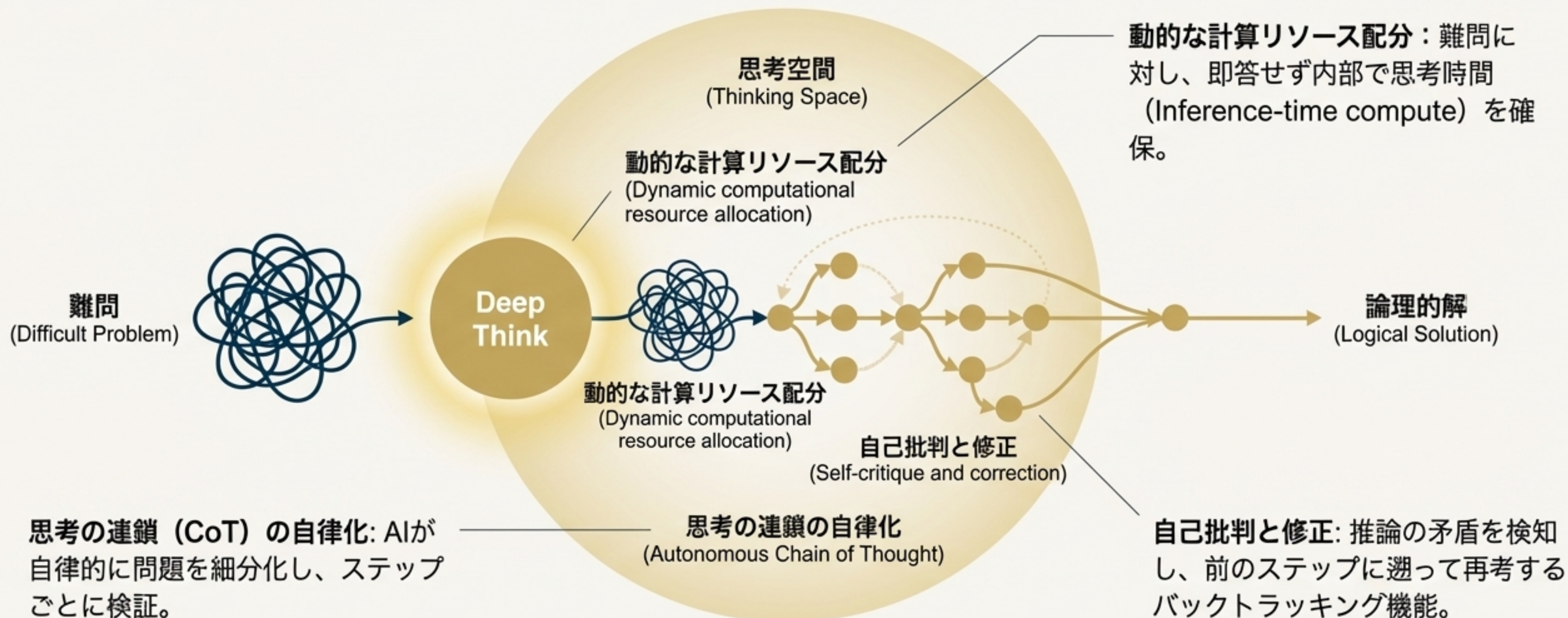
3. 検証の自律化

自ら立てた仮説を、論理的証明やシミュレーションで検証する能力。

なぜレベル4が重要なのか？ レベル3までのAIは「労働の代替（コスト削減）」だが、レベル4のAIは「知的財産（富の源泉）」を直接生み出すからである。

停滞からの飛躍：推論モデル「Deep Think」の登場

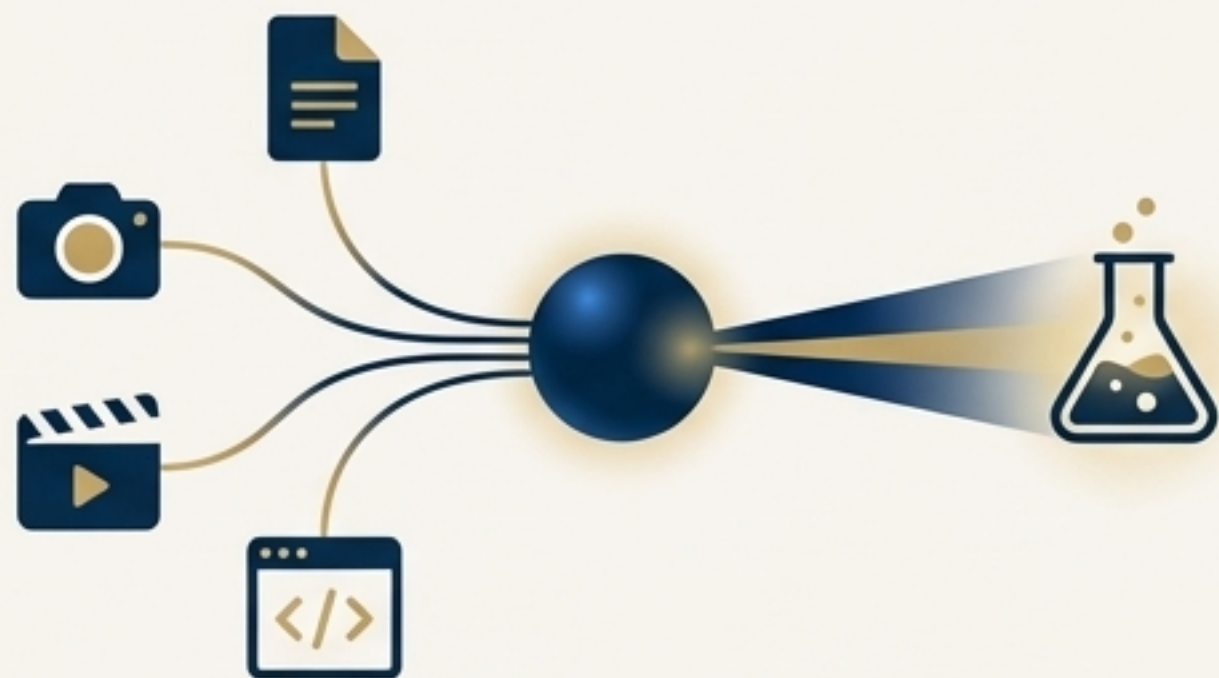
2025年初頭、業界は「スケーリング則の限界」に直面。性能向上は鈍化していた。Gemini 3 Proのブレイクスルーは、確率的な単語予測から、論理の積み重ねを可能にする「推論（Reasoning）」への根本的なシフトにある。



「発明」を支える2つの基盤技術

科学的発見はテキストだけでは完結しない。膨大な先行研究の理解が不可欠である。

マルチモーダリティの真の統合

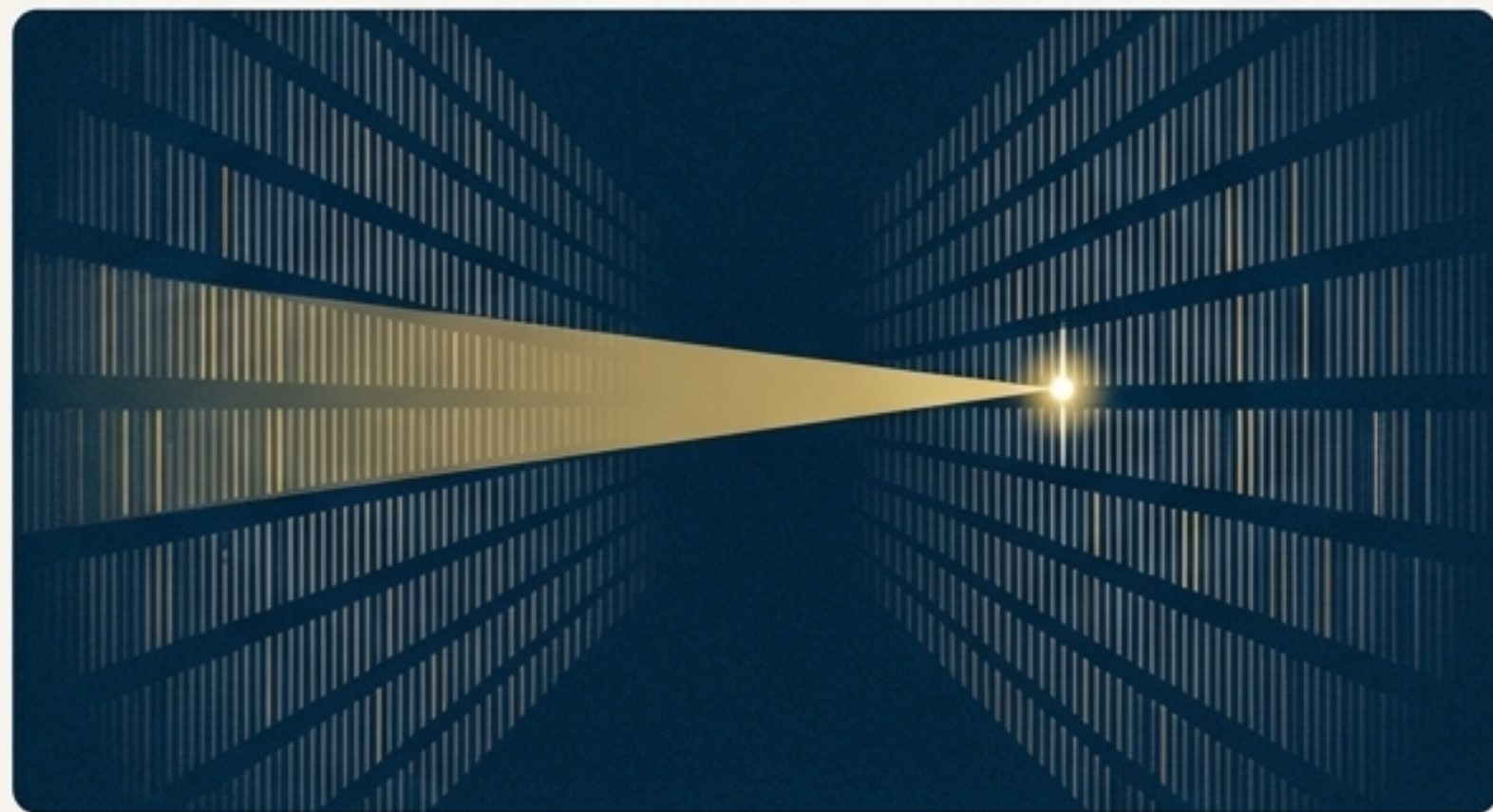


テキスト、画像、音声、動画、コードを単一モデルで処理。実験映像から化学反応を推論するなど、物理世界の事象理解に決定的な意味を持つ。

MMMU-Pro: **81.0%** (vs. GPT-5.1 Pro 76.0%)

Video-MMMU: **87.6%** (vs. GPT-5.1 Pro 80.4%)

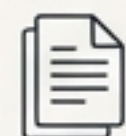
100万トークンの文脈と記憶



文庫本10冊分以上の情報を一度に読み込み、針の穴を通す精度で情報を抽出。数千の特許や論文を分析し、「まだ試されていないアプローチは何か？」という問いに答えることを可能にする。

「発明」のベンチマーク検証：未知の問題解決能力を測る

既存のテスト（大学入試レベル）は、知識の「再現能力」を測るに過ぎない。
レベル4の検証には、AIが学習データにない、全く新しい問題を解けるかを測る
「フロンティア・ベンチマーク」が不可欠である。



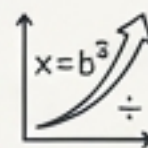
MMLU



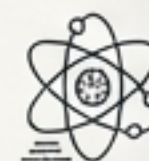
ARC-AGI

既存の知識
(Existing Knowledge)

FrontierMath:
未発表の研究レベル数学



CritPt: 物理学の
未発表研究課題



ARC-AGI-2:
抽象的推論と汎化能力

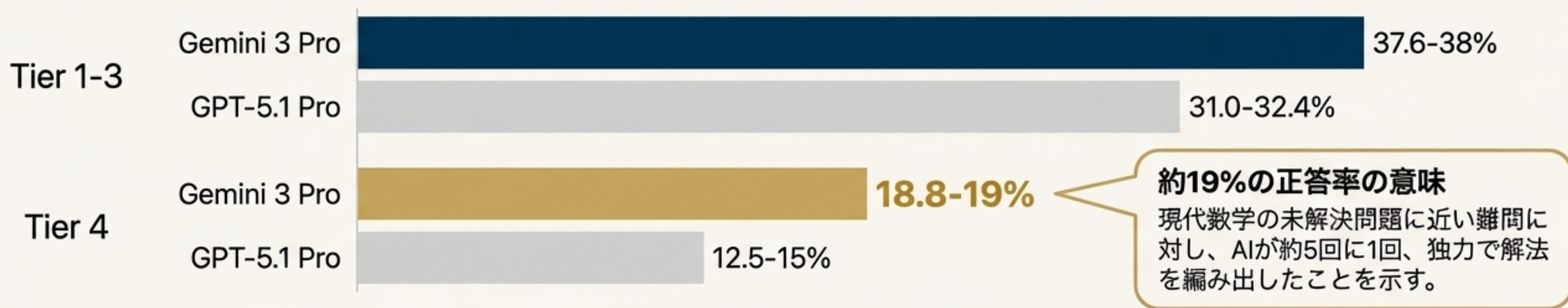


未知の領域
(Unknown Territory)

証拠①：未知の数学的・物理的構造の発見

FrontierMath - The Decisive Difference

Tier 1-3（大学院レベルまで）では差は限定的。しかし、Tier 4（研究レベル）での6.3ポイント差は、推論能力の質的な違いを示す。



CritPt - Physics Insight

物理学研究レベルの課題で競合の約2倍の性能。観測データから新しいモデルを構築する能力を示唆。

Gemini 3 Pro: 9.1%

GPT-5.1 Pro: ~5.0%

証拠②：「記憶」ではなく「思考」していることの証明

ARC-AGI-2: 学習データに全く含まれない、抽象的なパズルを解く能力を測る究極のテスト。AIの汎化能力と流動性知能を問う。



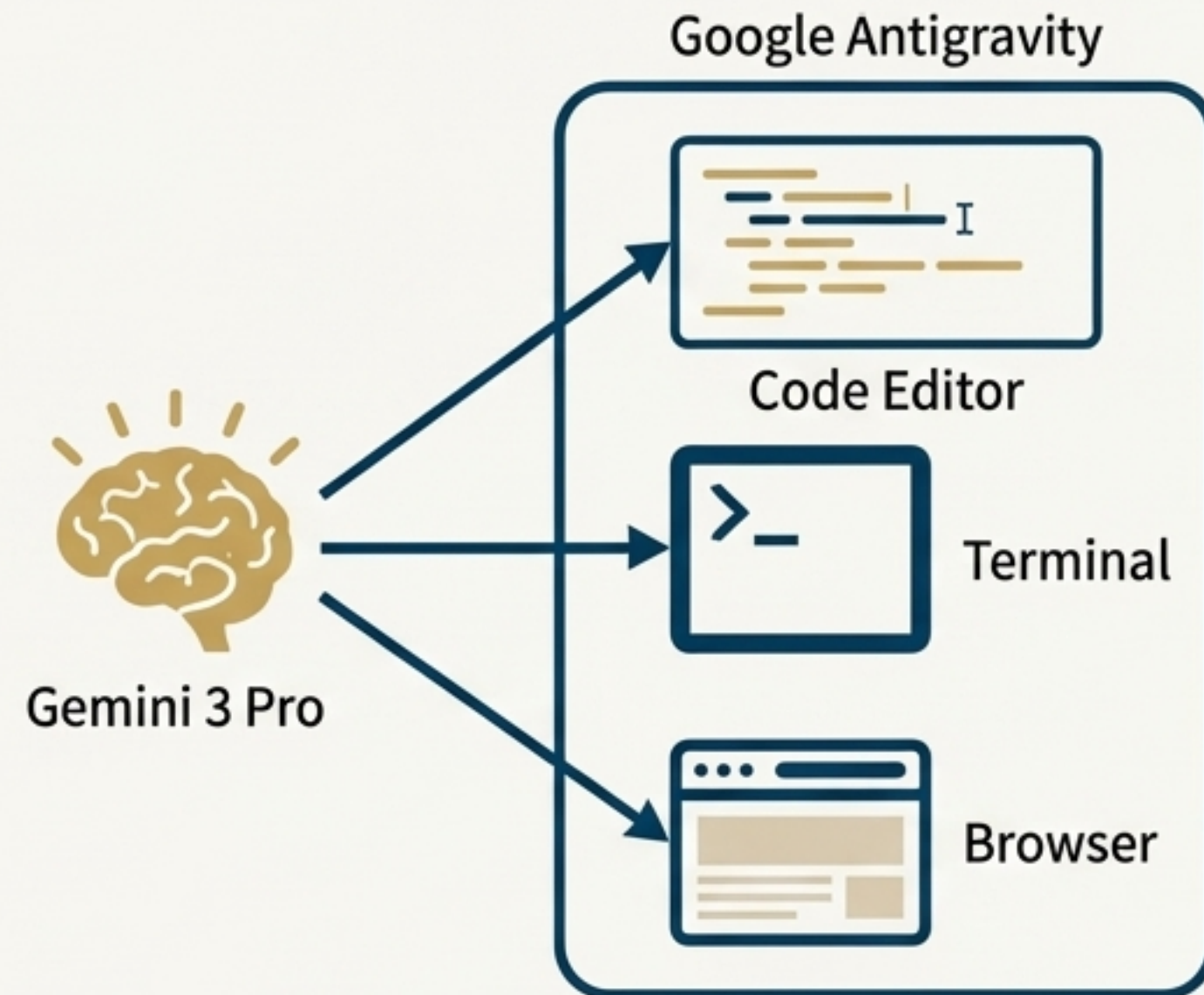
この圧倒的な差は、Gemini 3 Proが既知のパターンを適用する「熟練労働者」ではなく、未知のパターンを解読する「天才的なひらめき」を持つことを証明している。これは孫氏がレベル4に求める「創造性」の技術的担保である。

Humanity's Last Exam (HLE): 総合的超難問でも41.0%を記録し、GPT-5.1 Pro (26.5%)を凌駕。複合的な知識統合能力を証明。

自律性の検証：発明のための実験場「Google Antigravity」

能力（Capability）は証明された。
しかし、自律的に運用（Operation）できなければ「道具」に過ぎない。
孫氏の定義「AI自体による」を満たすには、Agency（自律的行動生成）が必須である。

AIエージェントがコードエディタ、ターミナル、ブラウザを直接操作できる統合開発環境。Gemini 3 Proの「頭脳」が仮説を立て、Antigravityという「手足」が実験を行う。



Key Antigravity Features



1. 自律的計画と実行: 「新しいアルゴリズムを考案し、実装し、テストせよ」といった高レベルな目標を達成。



2. 再帰的な改善ループ: コードを書き、テストし、エラーが出れば自ら修正する「試行錯誤」を自動化。



3. マルチエージェント協調: 複数のエージェント（コーダー役、テスター役など）が協調し、レベル5（組織的活動）の萌芽を示す。

戦略的分岐点：「究極の熟練労働者」 vs 「天才的な研究パートナー」

GPT-5.1 Pro 究極の熟練労働者



- 核心理念:
指示されたことを、既存の知識に基づき完璧にこなす。
- 重く点:
レベル3（エージェント）の完成度を追求。
- 強さ:
業務遂行の確実性、効率性。
- 制限:
未知のものを「0から1」で生み出す能力は限定的。

Gemini 3 Pro 天才的な研究パートナー



- 核心理念:
人間が思いつかないアプローチを提示し、未知の領域を探索する。
- 重く点:
レベル4（発明者）の扉をこじ開ける。
- 強さ:
新規の論理構造の創出、仮説生成。
- 制限:
現時点では研究テーマ設定など、人間の「指揮」を必要とする。

最終判定：Gemini 3 Proは「プロト・レベル4」に到達した

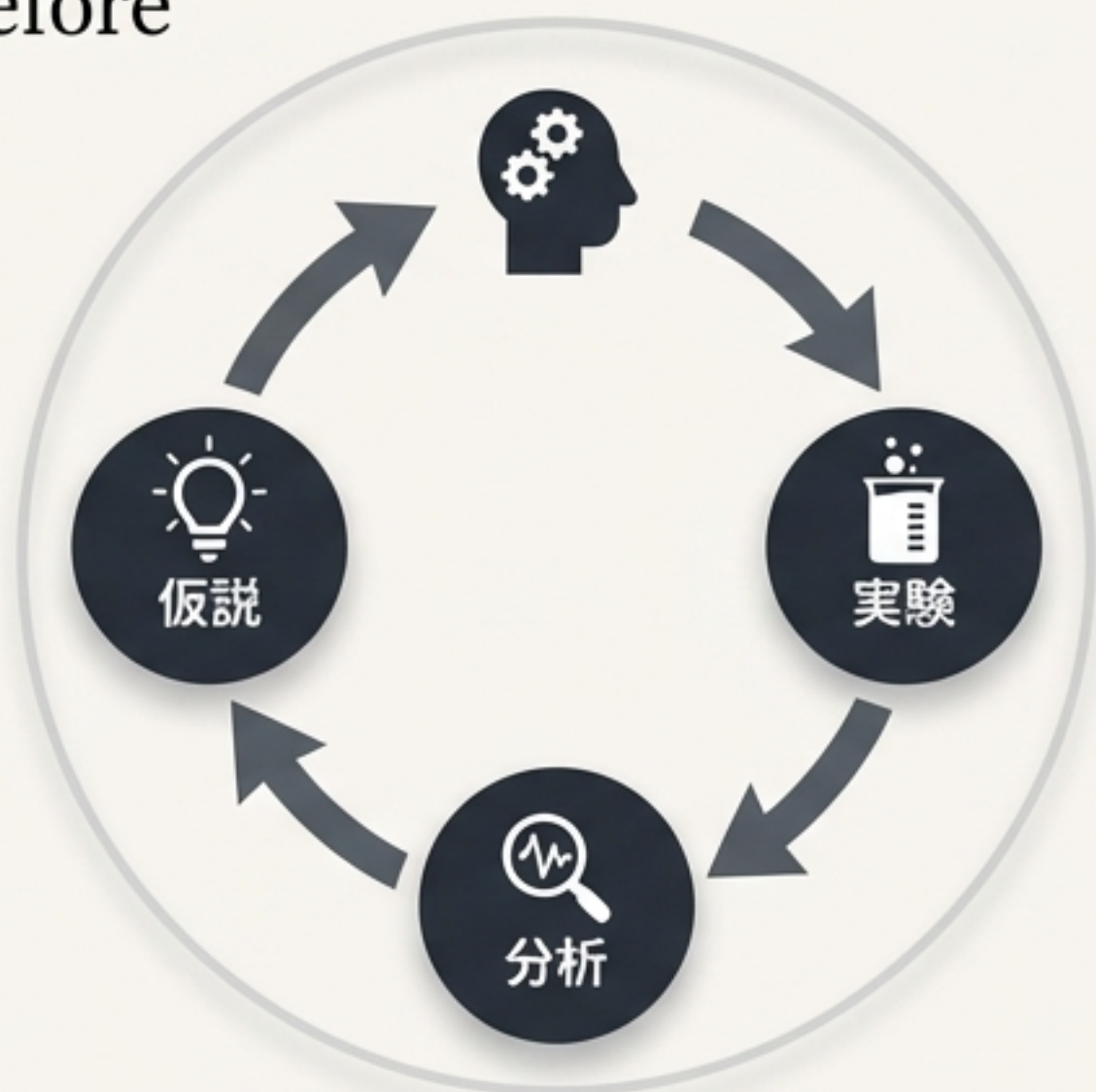
Gemini 3 Proは「レベル3」から「レベル4」への移行期にある存在と位置づけるのが最も正確である。



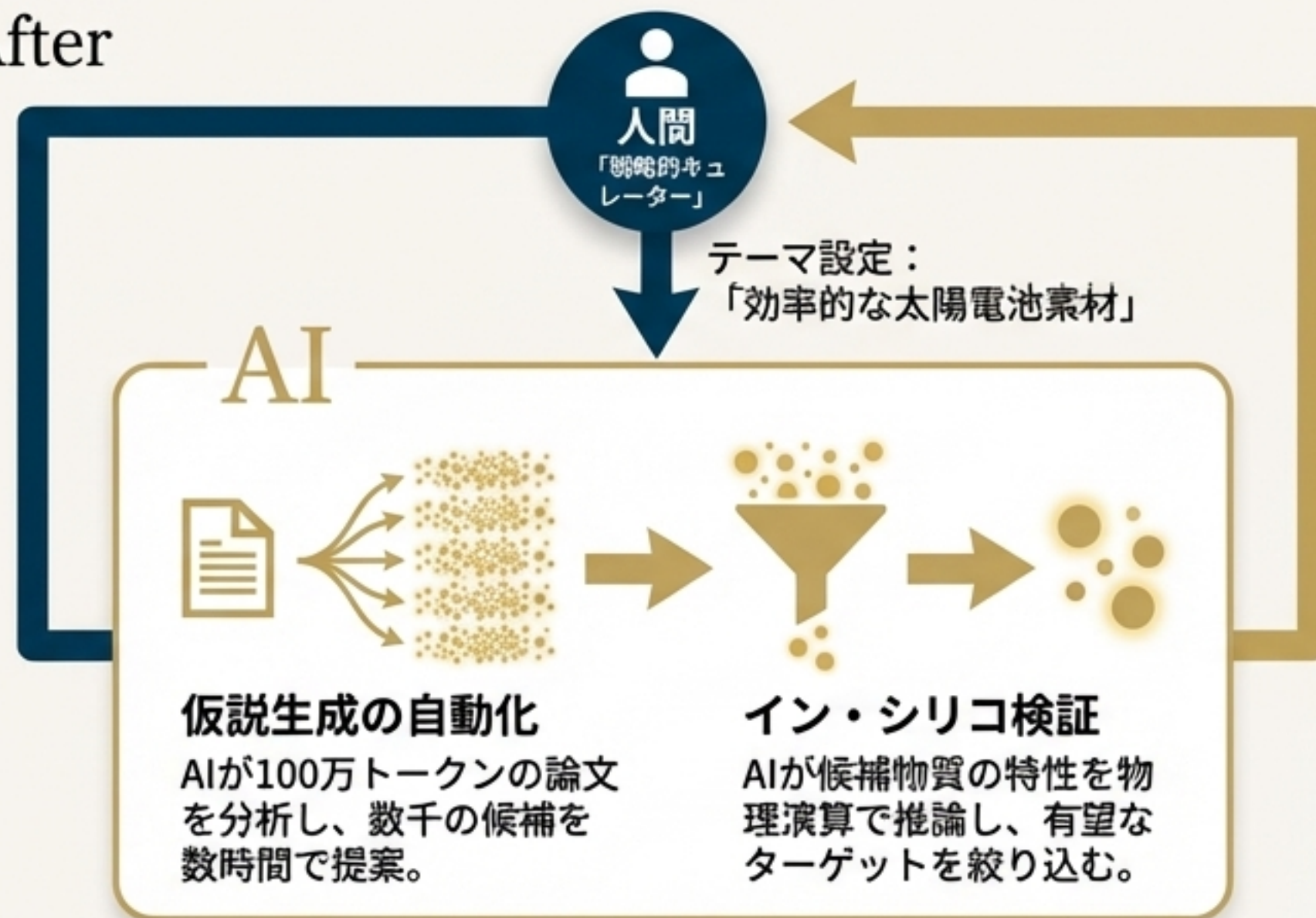
R&Dプロセスの不可逆的な変質

人間の研究者が主導してきた仮説→実験→分析のサイクルが、AIとの協業モデルへと根本的に変化する。

Before



After



知的財産（IP）の「ゴールドラッシュ」と次なる覇権争い
レベル4 AIの登場は、富の源泉を「労働」から「自動生成される知的財産」へとシフトさせる。AIが生み出す新薬の分子構造や新アルゴリズムが、次の時代の基幹産業となる。

ソフトバンクの戦略的布石

ソフトバンクグループがArm（チップ設計）、Stargate Project（計算資源）、そしてOpenAIとの連携を深める理由は、この「AIによるIP生成」の基盤インフラを支配するためである。

究極の賞（The Ultimate Prize）

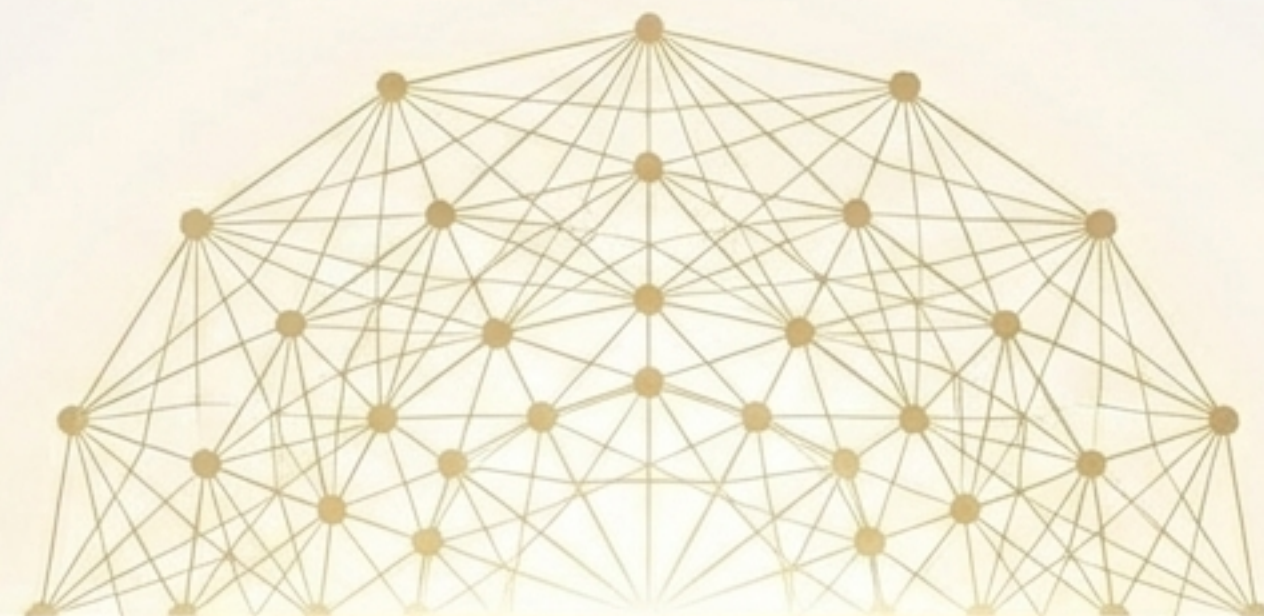
発明を自動化する時代において、その「知能（Gemini/GPT）」と「計算資源（Chips/Cloud）」を制する者が、次の経済覇権を握る。

知性の歴史における、不可逆的な一歩

**2025年11月、我々は「発明する機械」のプロトタイプを初めて手にした。
これは単なるソフトウェアの更新ではなく、知性の進化における特異点である。**

From Level 4 to Level 5:

Gemini 3 Proの登場は、孫氏のロードマップの確からしさを証明した。次のステップは、この「発明する知能」を持つエージェント同士が組織化され、人間が介在が介在せずとも次々とイノベーションを生み出し続ける「レベル5：AI社会」の形成である。



我々はその入り口に立っている。