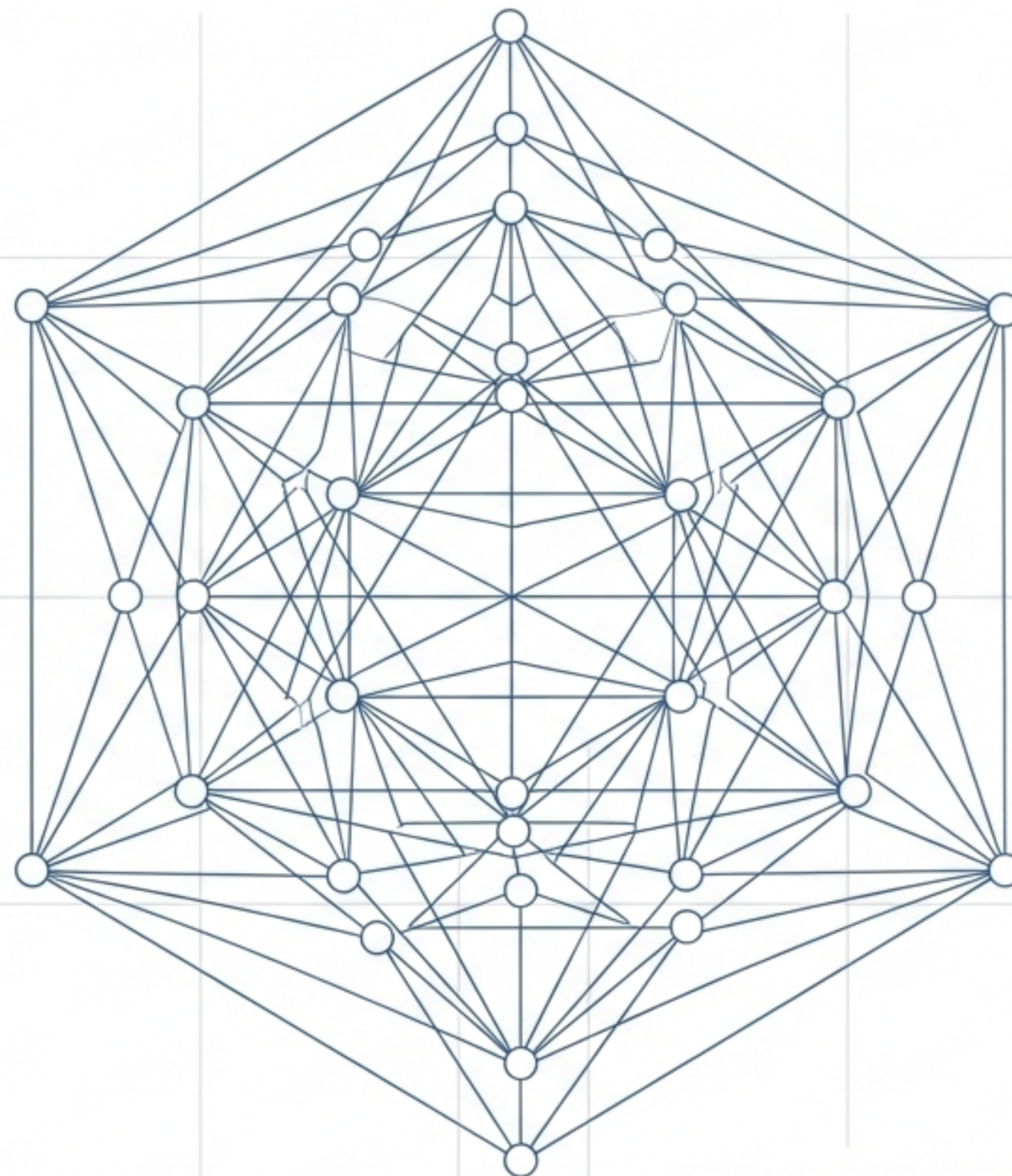
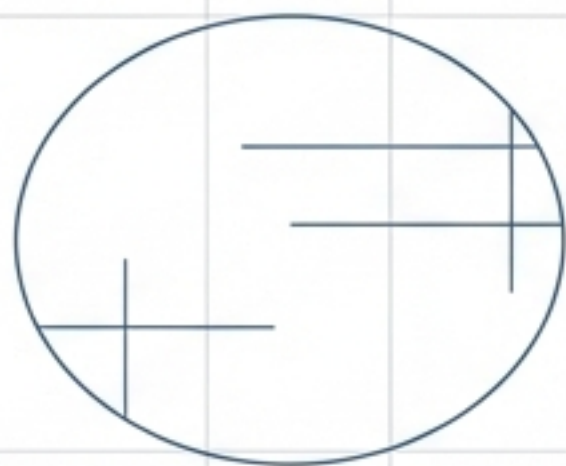


OpenAI Deep Research & GPT-5.2 Impact Report

「対話」から「エージェント」へ：
プロフェッショナル独式エンタート
プロフェッショナル知識労働の戦略的転換



要点 (Key Takeaways)

70.9%

GDPval (実務遂行能力)
Inter Regular

400k

Context Window (Tokens)
Inter Regular

エグゼクティブサマリー：自律型調査エージェントの産業化

基盤モデルの刷新

試験的なo3/o4から、プロダクションレディな「GPT-5.2」へ完全移行。
40万トークンの入力と12万トークンの出力が可能。

Deep Research 2.0

単なるWeb検索から、信頼性を制御可能な「ターゲット検索」と、プロセスを監視する「リアルタイムビューア」へ進化。

市場戦略の分断

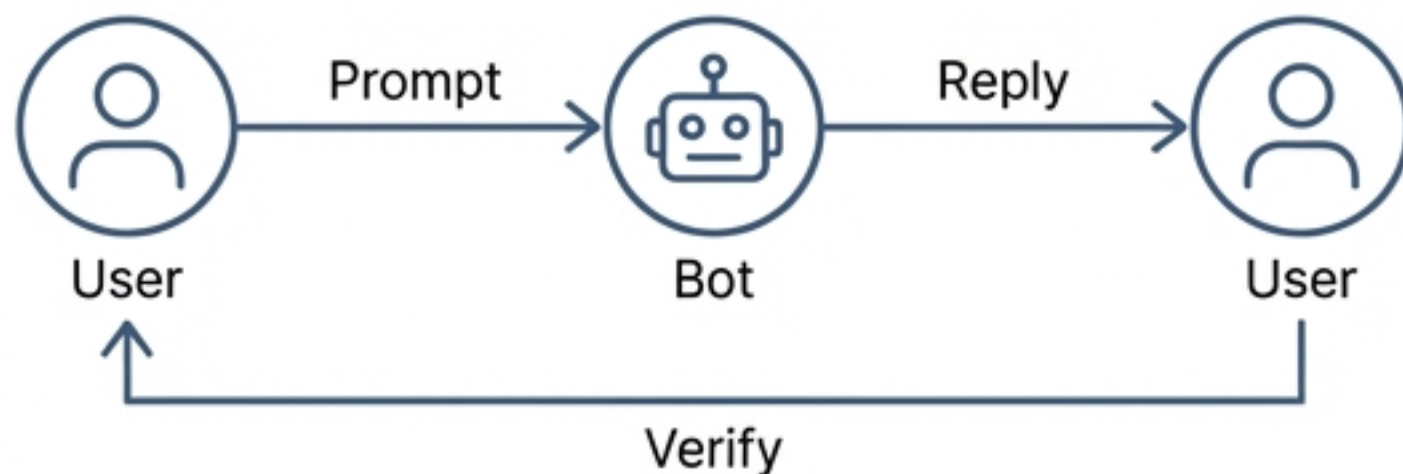
低価格な「Go (\$8/Ads)」、標準の「Plus (\$20)」、そしてハイエンドな「Pro (\$200)」への明確な階層化。

ビジネスインパクト

GDPval (実務遂行能力) で70.9%を記録し、人間の専門家と同等以上の成果を実証。

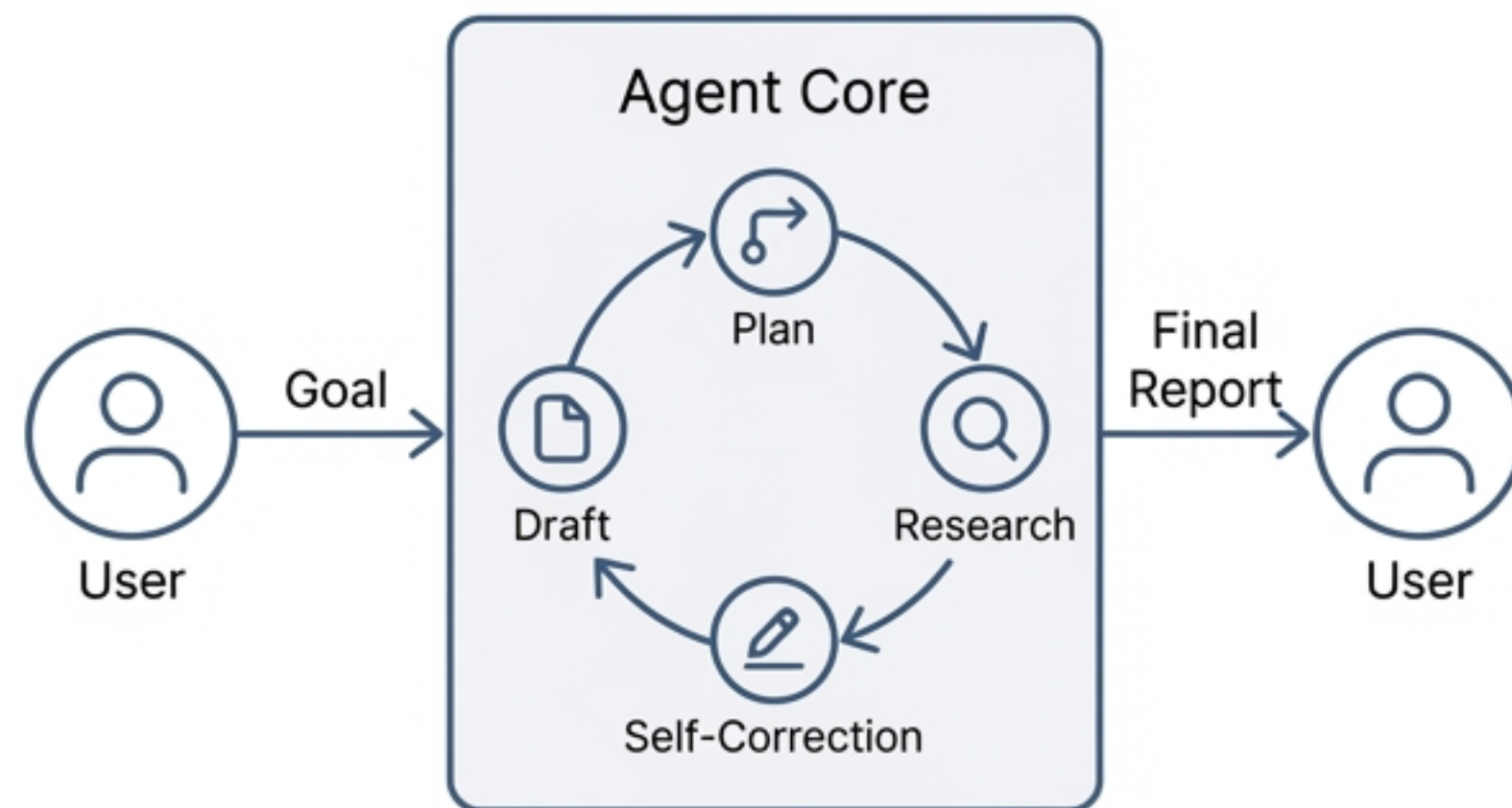
パラダイムシフト：受動的な「チャット」から能動的な「エージェント」へ

Legacy Chat Loop



Human-Heavy Loop

Agentic Loop 2.0



Outcomes-Oriented

Deep Researchは検索エンジンのラッパーではなく、数千ページを読み込み、統合的なレポートを作成する「リサーチアシスタント」として機能する。

GPT-5.2 モデルファミリー：タスクに応じた3つの「頭脳」

Instant (インスタント)



- Role: 高速・低遅延・バックグラウンド処理
- Use Case: 初期プロンプト解釈、スクリーニング、日常的な即答タスク

Thinking (シンキング)



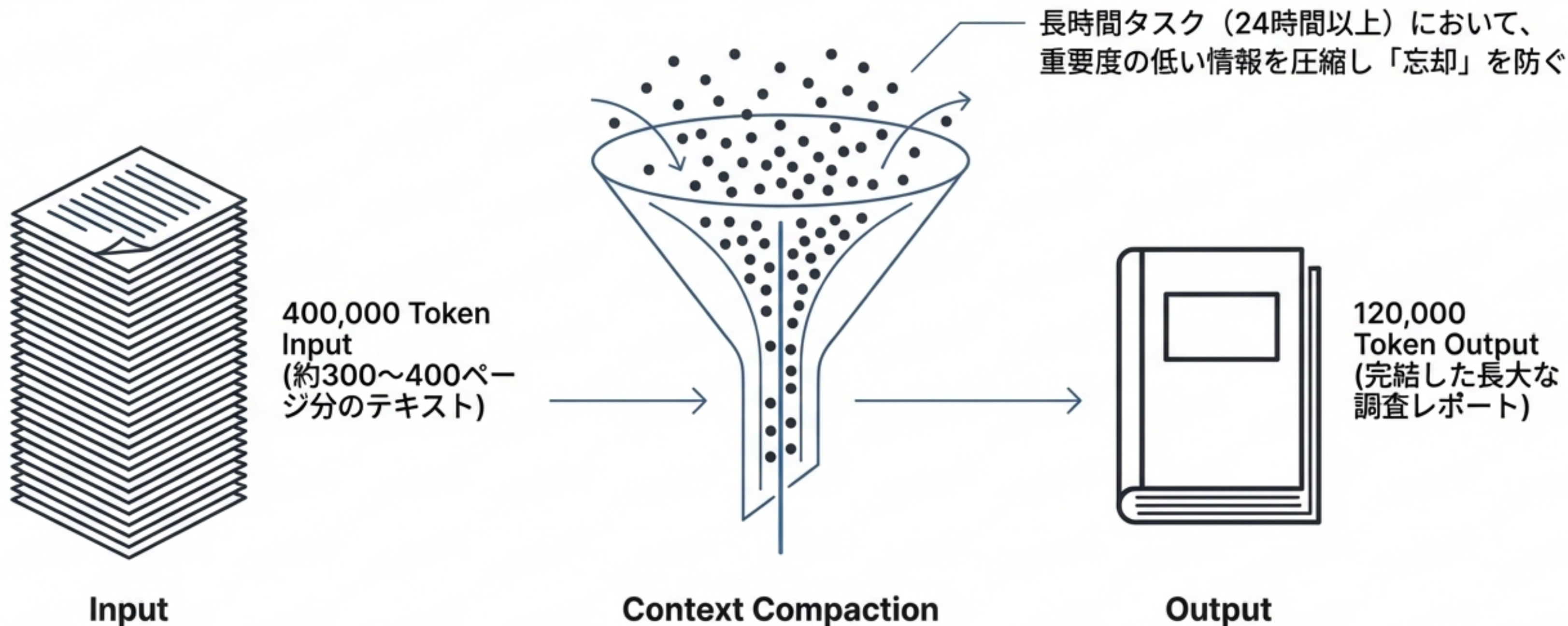
- Role: 中核モデル・論理推論
- Metric: ハルシネーション（幻覚）発生率を30-38%低減
- Use Case: 情報の統合、矛盾の解消、実務的なレポート作成

Pro (プロ)



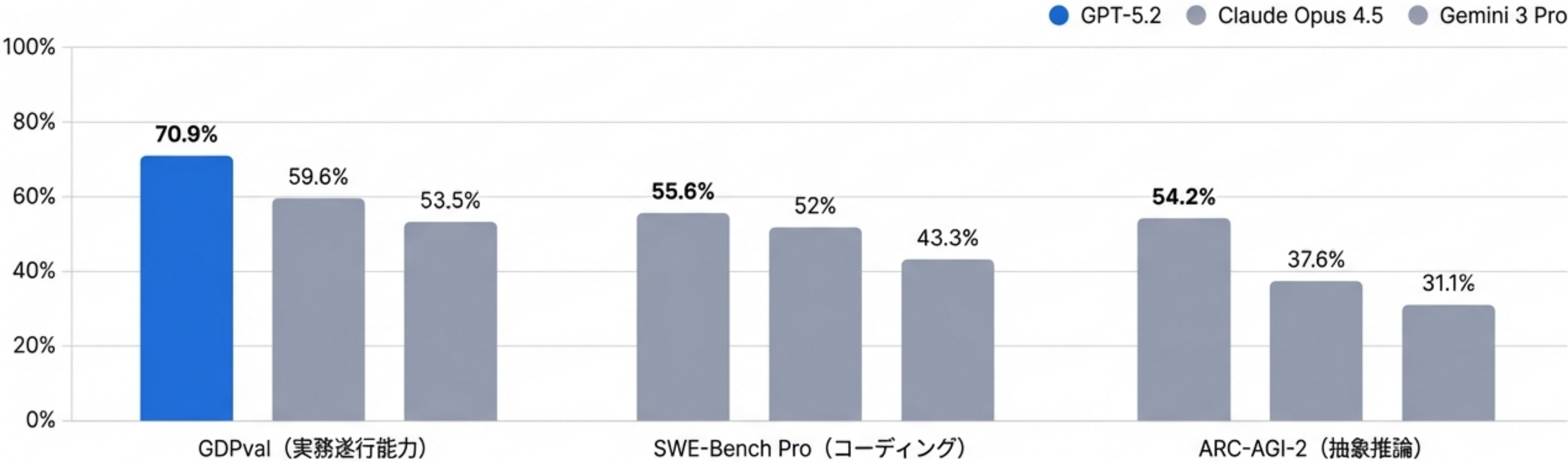
- Role: 最高レベルの推論 (xhigh)
 - 深考
- Use Case: 数学的証明、法的文書のクロスリファレンス、数千ページの資料分析
- Note: 月額\$200のProプランまたは高難度タスクで起動

Swiss International Scientific コンテキストと記憶の技術的ブレイクスルー



Swiss International Scientific

ベンチマーク評価：実務能力における圧倒的なリード



GDPval（実務遂行能力）
GPT-5.2 Thinkingは70.9%を記録。AIが補助ツールを超え、実務タスクにおいて人間の専門家と「引き分け以上」の結果を出せることを証明。

ARC-AGI-2（抽象推論）
54.2%を記録し、競合（Claude: 37.6%, Gemini: 31.1%）を大きく引き離す。未知の課題への適応力が極めて高い。

SWE-Bench（コーディング）
55.6%でSOTA（State-of-the-Art）を樹立。

Deep Research 2.0：制御可能な「ターゲット検索」

Problem: Open Web Noise



課題: 従来のAIはWeb全体を検索し、信頼性の低いブログやノイズを含んでしまった。

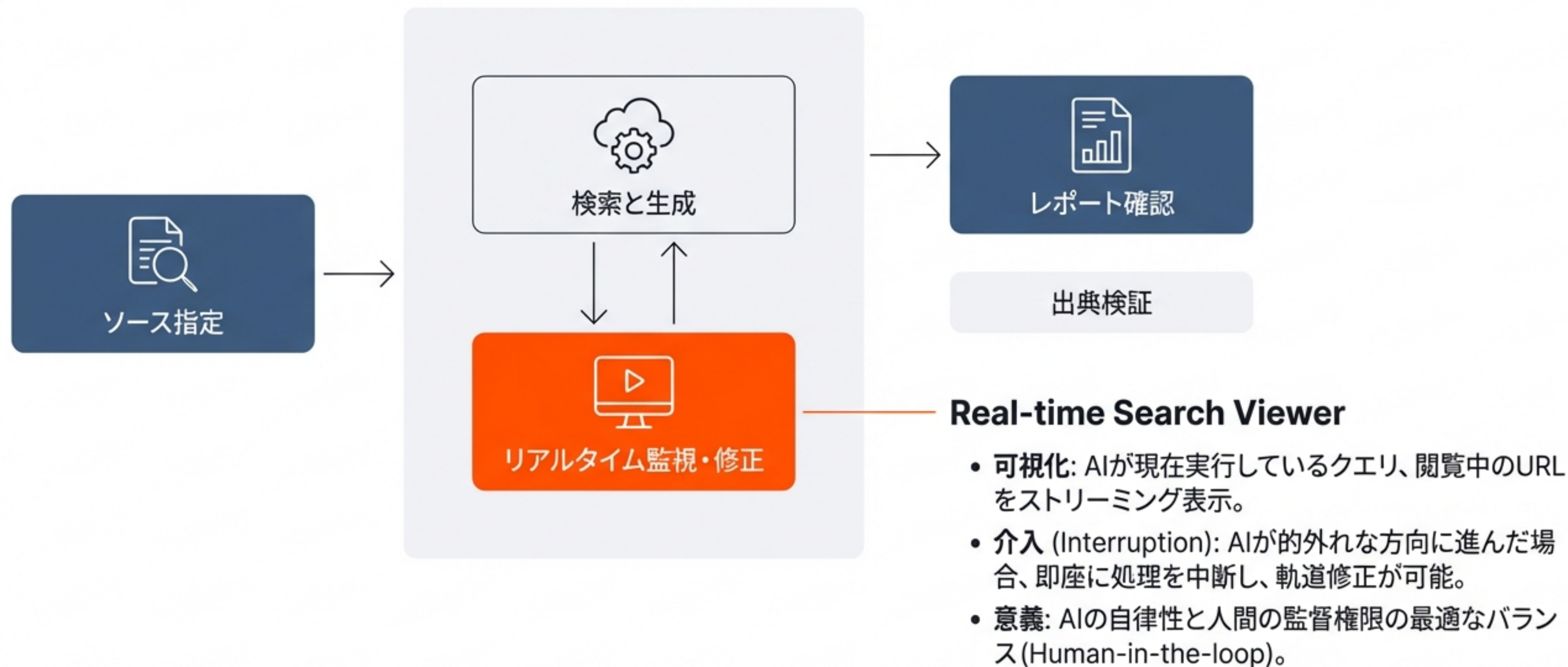
Solution: Targeted Scope



解決: ユーザーによる明確な境界線設定。

- ・ドメイン指定: 信頼できる公式サイトや特定のIRページのみを調査対象に限定。
- ・アプリ接続: 外部Webだけでなく、Google Drive、GitHub、HubSpotなどのプライベートデータ資産も対象に。

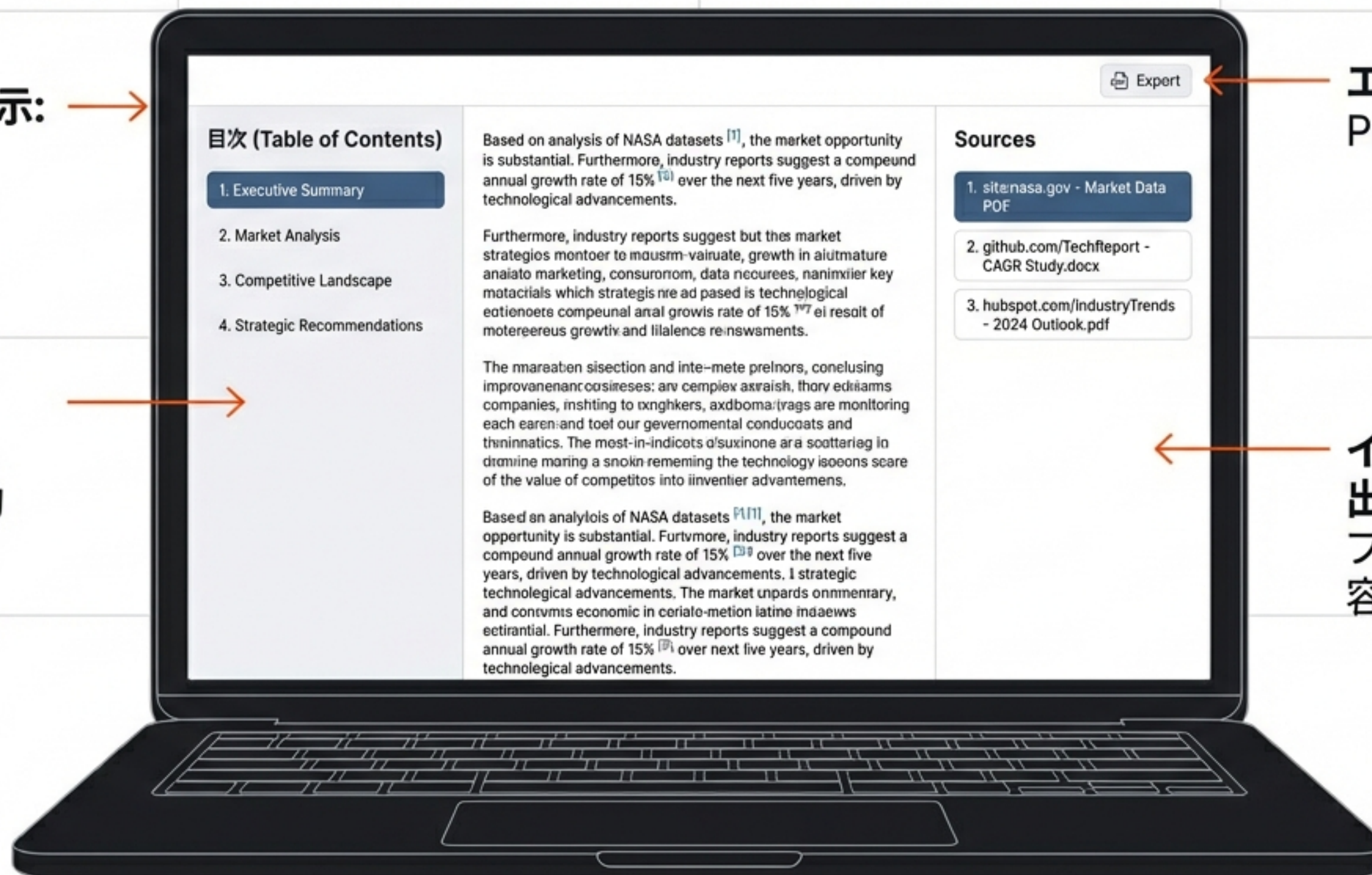
透明性と介入：ブラックボックスの解消



成果物：ビジネス品質のフルスクリーンレポート

フルスクリーン表示:
閲覧専用のUI

構造化された
ナビゲーション:
目次で瞬時に移動

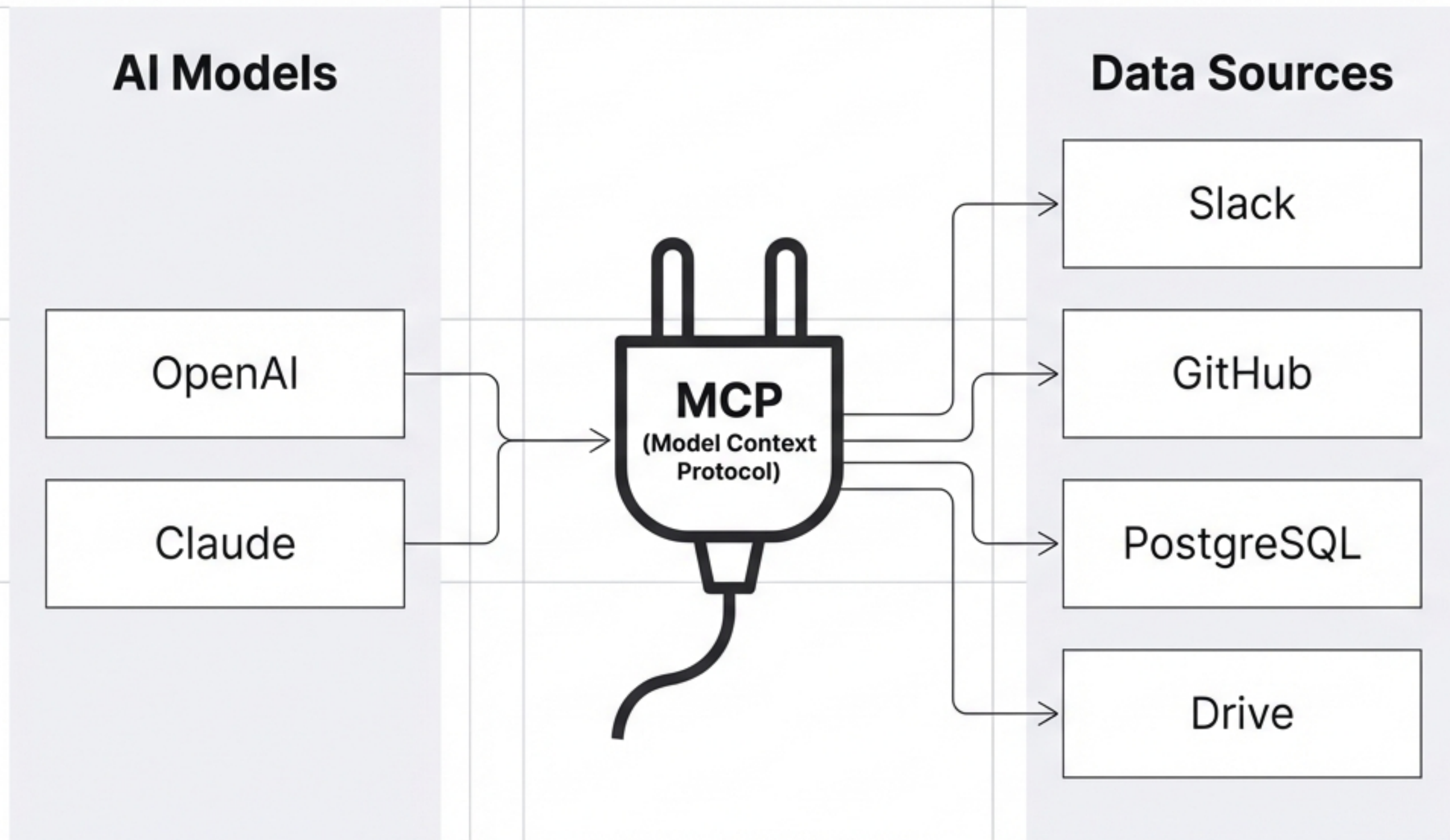


エクスポート:
PDF / DOCX

インタラクティブな
出典管理:
ファクトチェックが
容易

相互運用性：標準規格「MCP」の採用

Swiss International Scientific



Analogy: 「AIのためのUSB-C」。

Implication:

- Slack, GitHub, PostgreSQLなどに対し、モデルごとに**個別のコネクタを作る必要がない**。
- MCPサーバー側で「**Search**」と「**Fetch**」を定義するだけで、Deep Researchが自動的に探索方法を理解する。

新しい経済圏：サブスクリプションの階層化

Swiss International Scientific



エージェント戦争：競合3社のポジショニング

Google (Gemini 3 Pro)

Quantity & Multimodal

強み: 「量」と「マルチモーダル」。200万トークンの圧倒的コンテキスト, Google Workspaceとのネイティブ統合。

Anthropic (Claude Opus 4.6)

Inter Regular Use

強み: 「操作 (Computer Use)」。ブラウザやPC画面を直接操作する能力、自律的なコーディング開発。

OpenAI (GPT-5.2)

Inter Regular & Reasoning

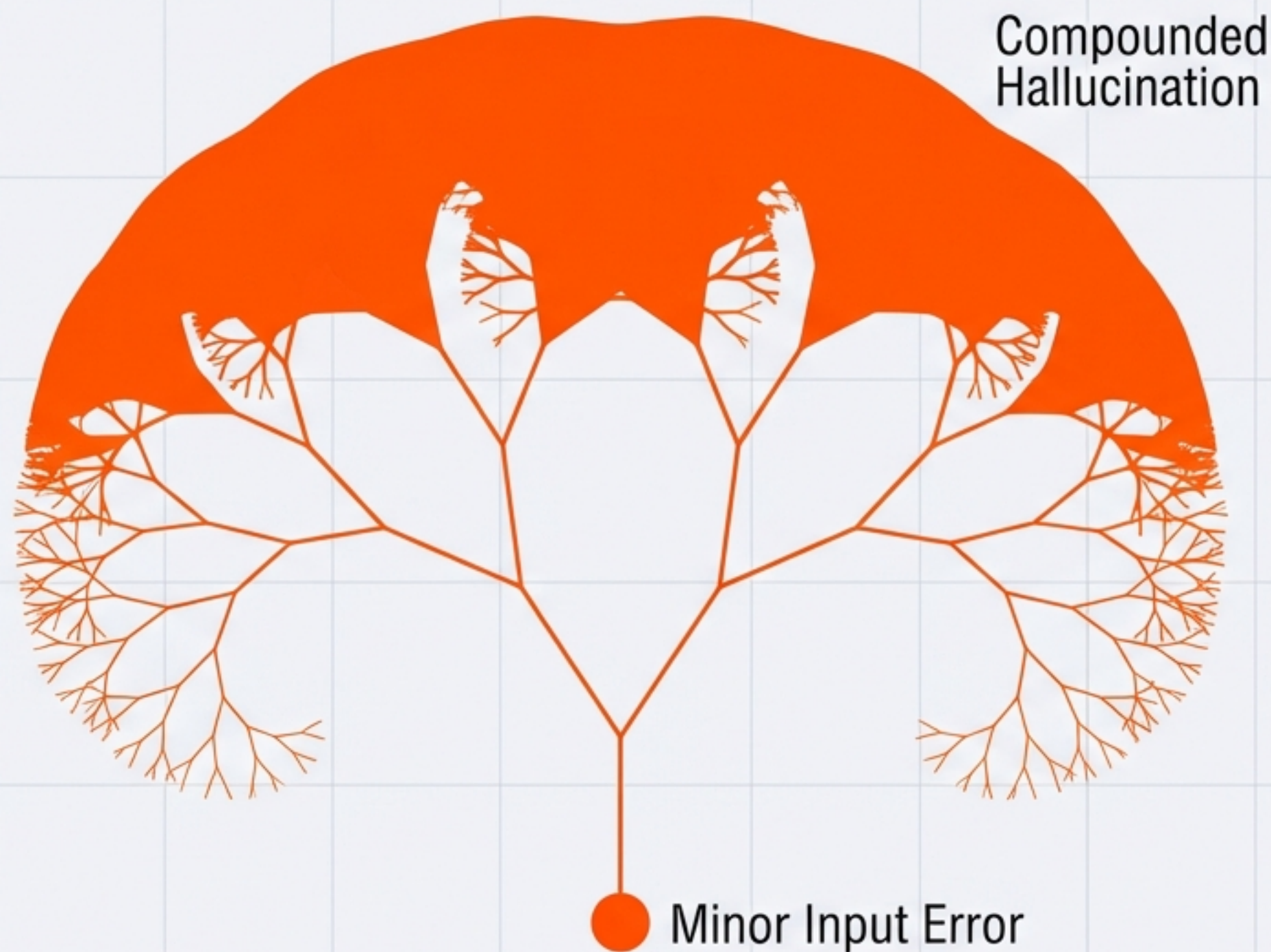
強み: 「調査ワークフロー」と「推論の質」。Deep Researchによる特定業務への最適化、MCPによるコネクティビティ。

プロフェッショナルワークへの影響： 「収集」から「判断」へ



- Workflow Shift: 市場調査や技術サーベイなど、数日かかっていた「下準備」が数十分で完了。人間の役割は検証と決定へ。
- Cultural Shift: 「企業的 (Corporate)」なトーン。ハルシネーション抑制のため、柔軟性や愛嬌が抑制され、より厳格で冷徹なアウトプットへ変化。

リスクと課題：エラーの複合化



Error Compounding

初期段階の小さな誤認が、多段階の推論プロセスを経て増幅され、最終的にもっともらしいが完全に誤った結論を導くリスク。

The AI Divide

月額\$200のProプランと\$20のPlusプランの間で、利用できる知能の格差（AI Divide）が拡大する懸念。

結論：AIを「使う」から「監督する」時代へ

**Success now depends on
Supervision Skills.**

**成功の鍵は、プロンプトエンジニアリングから
「監督する技術」への移行である。**

- GPT-5.2とDeep Researchは、AIをツールから「専門家エージェント」へと進化させた。
- リアルタイムビューアでの監視、ソースの裏取り、そしてAIの出力を盲信せず判断を下す能力が、これからのプロフェッショナルの核心的スキルとなる。