Architectural Transparency

ChatGPT「deep research」アップデート詳細分析報告書（2026.02）

GPT-5.2への移行と「透明なボックス（Glass Box）」化するAIエージェント

対象アップデート：2026年2月10日（米国時間）
基盤モデル：GPT-5.2 / 5.2-chat-latest
主な変更点：信頼サイト指定、MCPアプリ接続、リアルタイム介入

エグゼクティブサマリー：制御可能な自律型リサーチへの転換

GPT-5.2 駆動

2026年2月10日より基盤モデルを更新。推論能力とコンテキスト管理が強化。

4つの主要変更

1. **ソース統制:** ユーザーによる「信頼サイト（Trusted Sites）」の指定とキュレーション。
2. **アプリ連携:** MCP経由での社内データ/SaaSへの「読み取り専用」接続。
3. **プロセス介入:** 計画編集および実行中のリアルタイム追跡・方向修正。
4. **成果物ビュー:** 全文ビューア（目次・ソース・履歴）とドキュメント形式での出力。

Strategic Impact

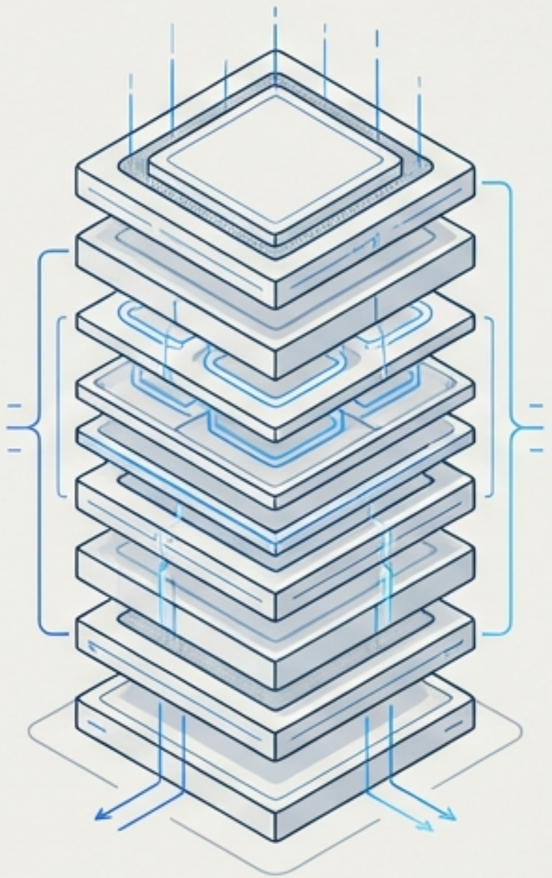
信頼性の向上: 参照元のブラックボックス化を解消し、業務監査への適合性が向上。
セキュリティ要件: 外部アプリ連携に伴うデータガバナンス（RBAC、共有範囲）の再定義が必須。

Rollout Status

Plus/Proユーザーは即日利用可能。Free/Goユーザーは数日以内に展開予定。

技術基盤：GPT-5.2 とコンテキスト管理の進化

CONFIRMED SPECS



- **推論能力:** APIパラメータ `reasoning.effort: xhigh` の導入により、複雑な論理構築能力が向上。
- **コンテキスト処理:** 'Server-side compaction'（サーバー側圧縮）の実装。大量の検索結果と長文脈の保持効率が改善。
- **開発者ガイド:** コード生成（特にフロントエンドUI）およびスプレッドシート理解の精度向上。
- **安全性:** GPT-5由来の「Safe completion」方針を継承。拒否ではなく、安全な範囲での有用な回答を優先。

UNSPECIFIED / UNKNOWNS

- **内部アーキテクチャ:** 層構造、学習データの詳細は一次資料にて「未指定」。

- **Deep Research固有ロジック:** 検索アルゴリズムやランキング検証の具体的なパイプライン仕様は非公開。

入力制御①：Webソースのホワイトリスト化「Manage Sites」

Feature Overview:

- **機能:** ユーザーが「承認済みURL（Approved URLs）」を指定、または検索範囲を特定ドメインに限定可能。
- **UI:** 設定メニュー「Manage sites」からドメイン単位で管理。

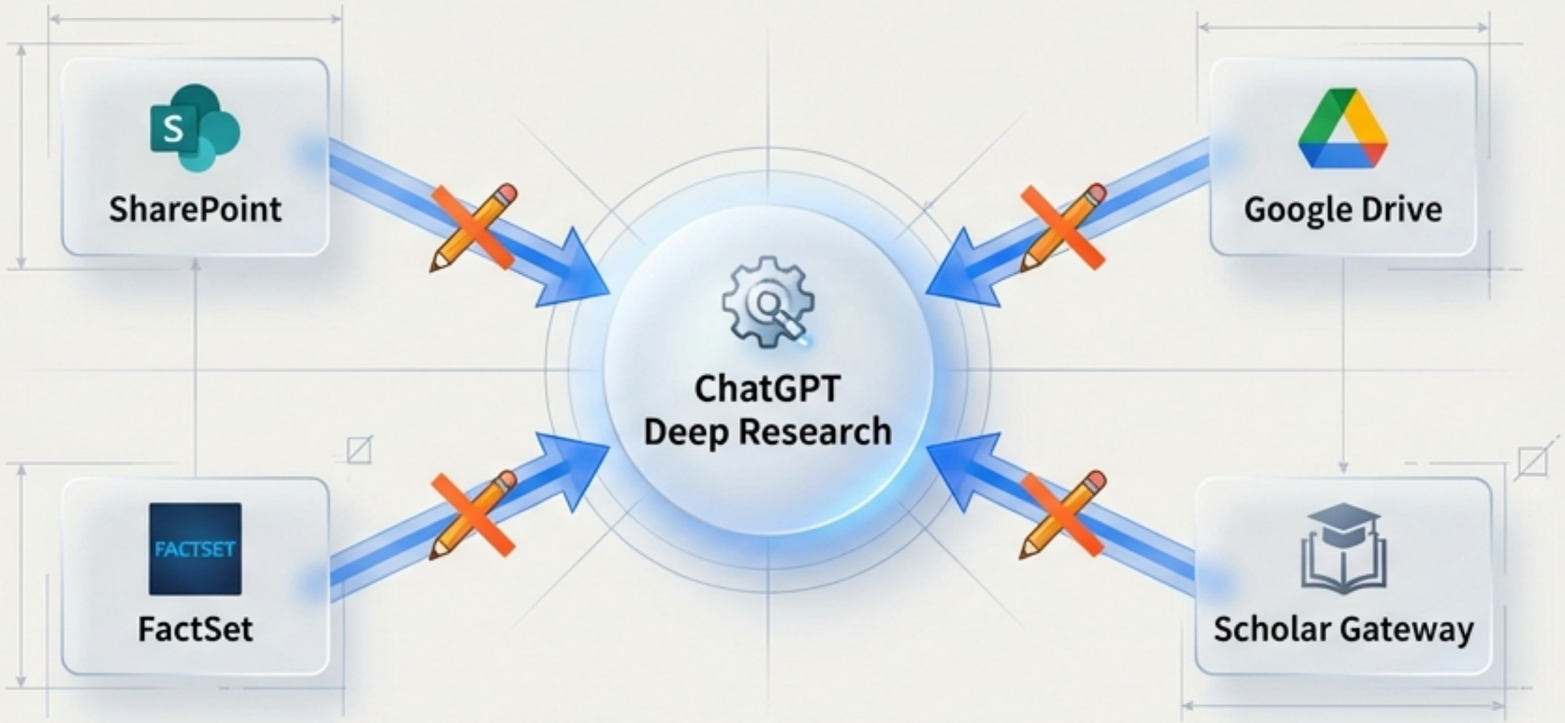
ハルシネーション抑制: 不確かな一般Web情報を排除し、業界標準や一次情報のみに基づいた回答を生成。

ノイズ低減: 検索空間を狭めることで、リサーチ結果の再現性と精度を物理的に向上させる。

閉じたループ（Closed Loop Errors）: 指定した信頼ソース自体に誤りがある場合、AIはそれを「正」として出力するリスク。

バイアス: ユーザーの選定バイアスがそのままレポートの結論に反映される可能性。

入力制御②：MCPによる接続アプリと「読み取り専用」原則



Integration Mechanics

Model Context Protocol (MCP): 社内データや専門DBをリサーチ対象として統合。

接続フロー: `Settings` -> `Apps` -> `Connect` -> `OAuth認証`。

Security Protocol: Read-Only

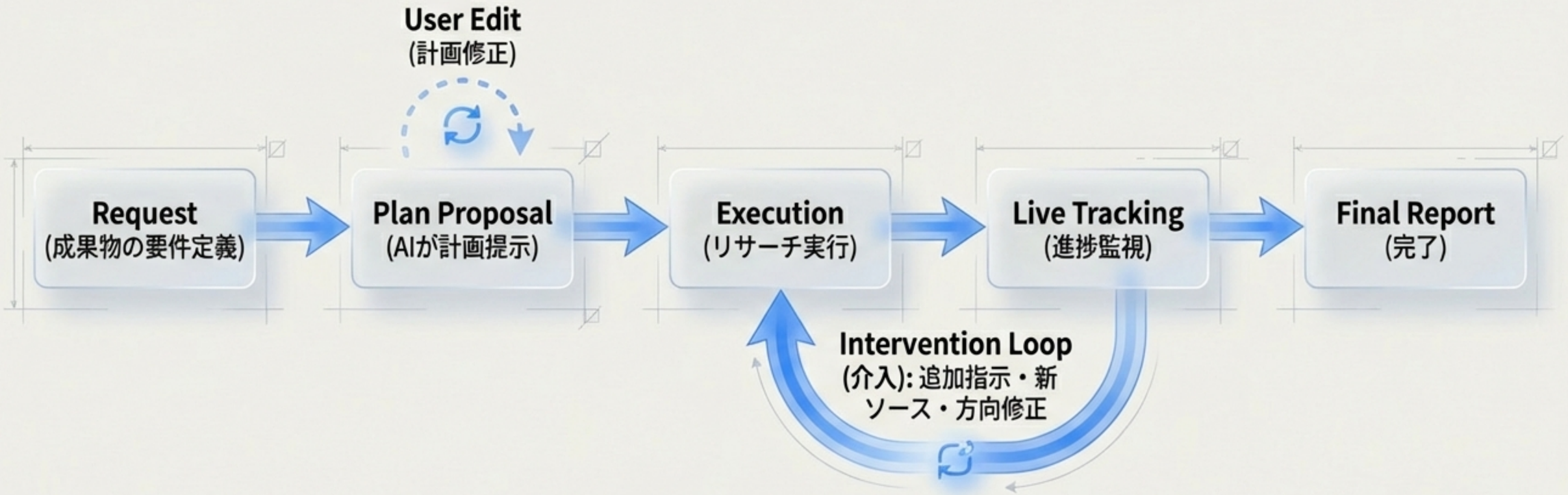
仕様: Deep Research実行時、接続アプリに対しては「Read Actions（読み取り）」のみが許可される。
重要性: 外部データベースへの書き込みや変更（Write Actions）は発生しない。誤作動によるデータ破損を防ぐ。

Deployment

任意のMCPサーバーへの接続が可能（Web版のみ）。

Business/Enterprise/Eduプランでは管理者が利用可否を統制。

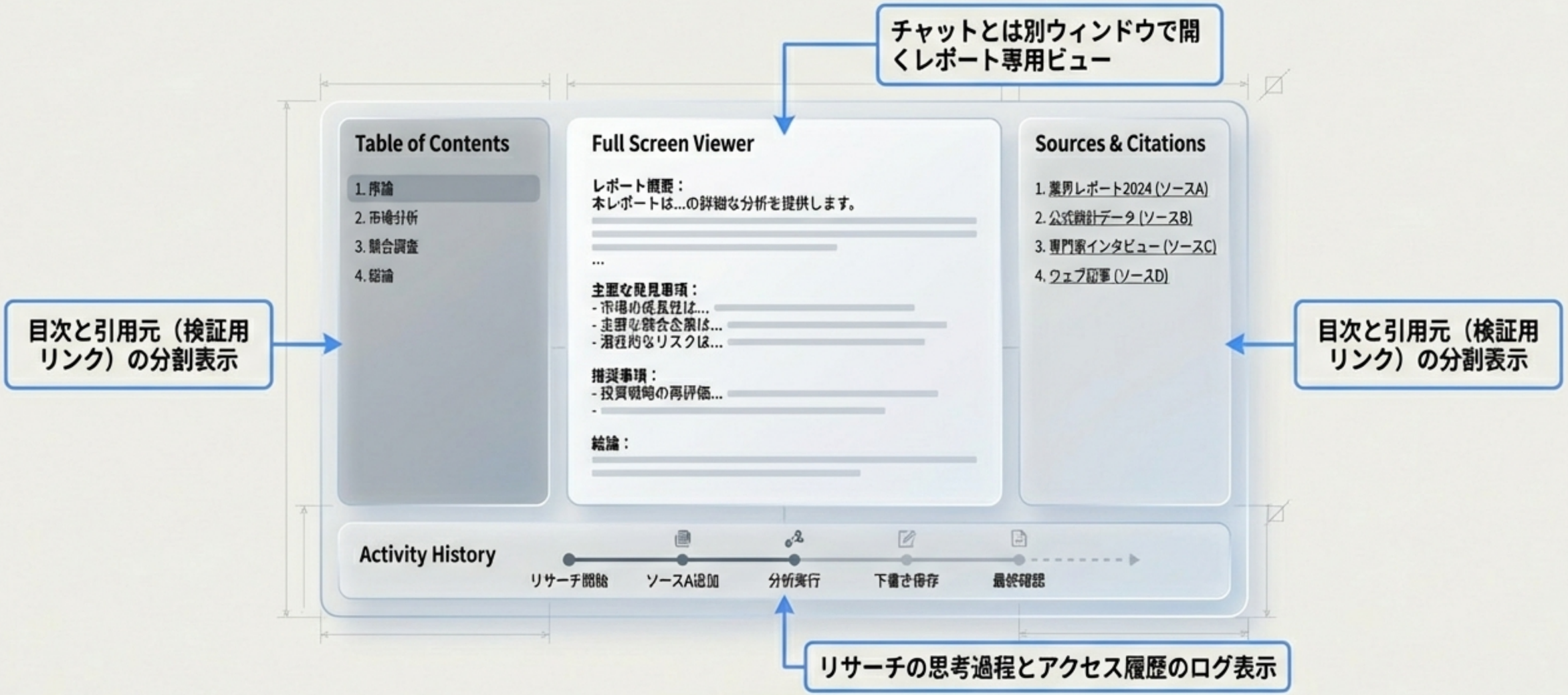
実行プロセス：「計画・監視・介入」の循環ワークフロー



Key Insight:

以前: 入力後は結果が出るまでブラックボックス。
今回: プロセスが可視化（Observability）され、進行中の軌道修正が可能に。これにより「期待外れ」の成果物が生成される時間を削減。

ユーザー体験：チャットから「ドキュメント」へ



Artifacts: Download: PDF, Word (.docx), Markdown 形式でのエクスポートに対応。
会話ログとしてではなく、独立した「成果物」として共有可能。

ガバナンスとセキュリティ：データ境界の再考

Data Flow & Privacy

第三者提供: アプリ接続時はOpenAIに加え、アプリ提供元（Third Party）ともデータが共有される（各アプリのPrivacy Policy準拠）。

学習利用 (Training):

* **Business/Enterprise**: デフォルトで学習利用なし。
* **Personal**: デフォルトで学習利用あり（Data Controlsでオプトアウト可能）。

Retention Policy

保持期間: 原則として削除後30日以内にシステムから抹消（法的例外を除く）。

NYT訴訟関連の一時的保持義務は終了済み。

Risk Mitigation

Prompt Injection: 信頼できないMCPサーバー経由の攻撃リスクあり。接続先は厳選が必要。

Compliance

GDPR/APPI: 越境移転および第三者提供の同意形成プロセスの確認（特に個人情報を含む社内データ利用時）。

Logs: Activity HistoryによるProvenance（来歴）追跡。

実務へのインパクト：ペルソナ別分析

Researcher / Analyst

- **メリット:** 「信頼サイト」限定による情報の質担保。ソースの明示によるファクトチェック工数の削減。
- **変化:** ドラフト作成ツールから、検証可能なリサーチパートナーへ。

Developer

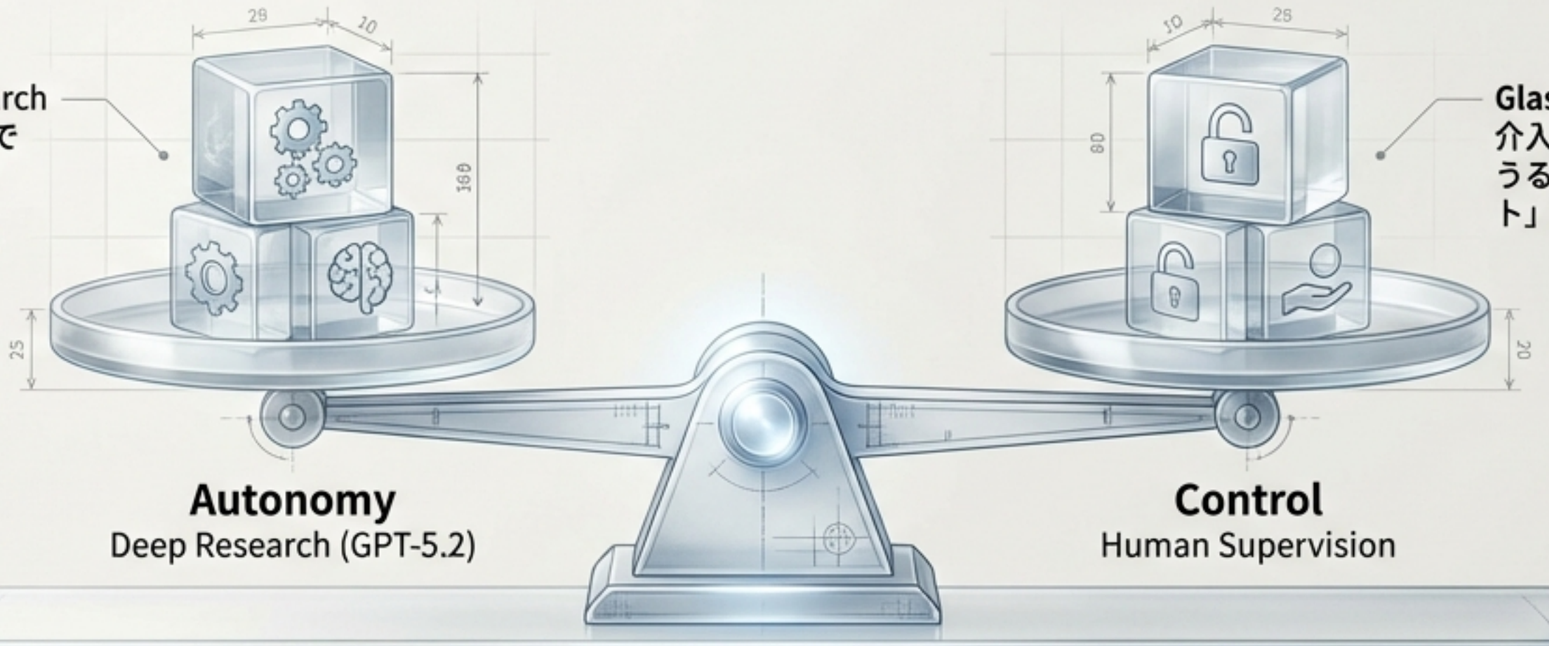
- **メリット:** GPT-5.2のコンテキスト圧縮とコード生成能力を活用したエージェント開発。
- **注意点:** 信頼できないMCPサーバー接続時のセキュリティ実装。

Enterprise Admin

- **メリット:** 社内ナレッジと外部Web情報の統合検索。
- **課題:** アプリ接続権限（RBAC）の設計と、データ共有ポリシーの策定が急務。

結論：自律性と統制のバランス

バラダイムシフト: Deep Research (GPT-5.2) は、単なる検索強化ではなく「人間が監督する自律リサーチ」への移行を示す。



Glass Box化: プロセスの透明化と介入機能により、業務利用に耐えうる「納得感のあるアウトプット」が可能に。

Call to Action:

1. **データポリシー確認:** 接続アプリ（Apps）の許可範囲と学習設定の見直し。
2. **ソースリスト整備:** 自社・業界における「信頼できるサイト（Trusted Sites)」リストの定義。
3. **プロセス設計:** 「放置」ではなく「介入」を前提とした業務フローへの変更。

16:9

付録：参照一次資料一覧

OpenAI Blog: Introducing deep research (2026/02/10 Updated)

Release Notes: ChatGPT — Release Notes (2026/02/10)

Help Center: Deep research in ChatGPT / Apps in ChatGPT / Developer mode and MCP apps

Technical Docs: OpenAI API Changelog / Using GPT-5.2 Guide

Policies: Privacy Policy plicy / Data Controls FAQ / Enterprise privacy

Media: The Verge, The Decoder (機能仕様の確認)