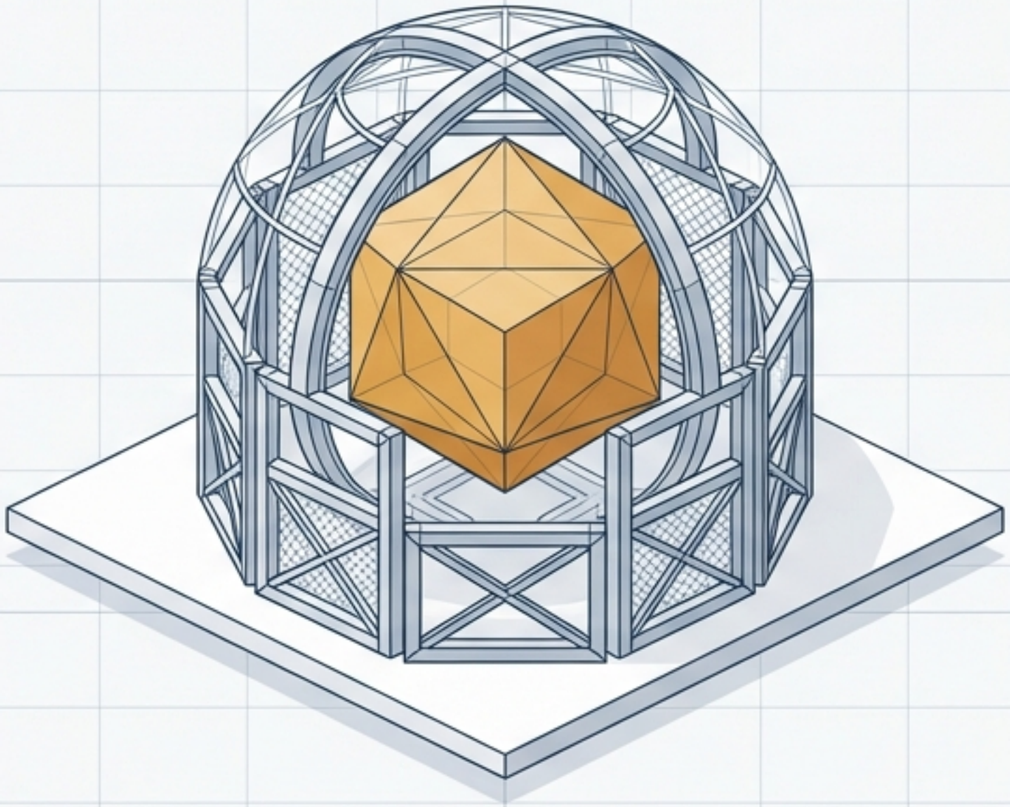
自律型AIの衝撃とエンタープライズ実装の全貌

Microsoft Copilot Cowork と Anthropic 提携がもたらす「Wave 3」の戦略的分析と導入プレイブック



Strategic Analyst Briefing

Executive Summary

Implementation Guide

パラダイムシフト：「回答するAI」から「自律実行するAI」へ

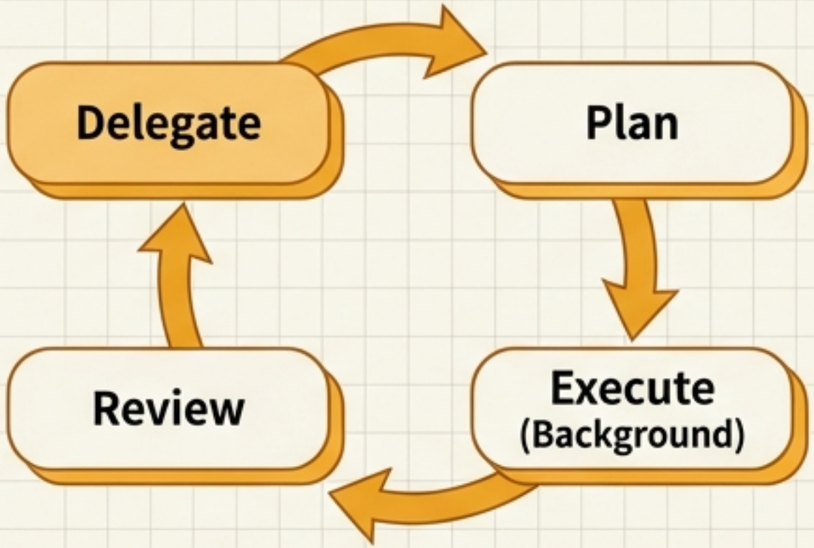
Copilot Coworkは、単発のチャット応答ではなく、Microsoft 365環境内で業務を計画し、継続実行し、監査可能な形で成果物を提示する「実行レイヤー」へと進化した。

Wave 2: 対話型AI

ユーザーが都度指示を与え、テキストやコードを生成。

単一ターン・単一アプリでの作業補助。

Wave 3: 自律型エージェント（Copilot Cowork）



自然言語で「求める成果」を委任。

複数アプリ（Outlook→Teams→Excel等）を跨ぎ、数分〜数時間にわたりバックグラウンドで継続実行。

提携から実装までの「加速カーブ」

2025年11月18日

インフラ・モデル提携

Microsoft・NVIDIA・Anthropicの戦略提携。
AnthropicがAzure計算資源を30Bドル分購入 (最大1GWまで拡張可能)。
ClaudeをAzure上でスケールさせるコミットメント。

2026年1月12日

ローカル自律実行のプレビュー

Anthropicがデスクトップ向け「Claude Cowork」を発表。
隔離VM上でローカル実行し、ファイルやMCP統合に直接アクセス可能に。

2026年3月9日

エンタープライズWave 3の統合

Microsoftが「Copilot Cowork」を発表。M365ガバナンス境界内でのクラウド実行。Agent 365 ($15/月)とM365 E7 ($99/月)の価格体系を提示。

パートナーシップの現実：確定領域と要検証領域

🟢 公開情報で確定している領域

技術統合: Copilot環境でのClaudeモデル（Sonnet 4.5 / Opus 4.1等）への継続アクセスとマルチモデル戦略。

インフラ: Azure上でのサーバーレス提供とEntra認証の統合。

実行環境: Copilot Coworkはローカルではなく「保護されたサンドボックス化クラウド環境」で実行。

🟡 企業契約において確認すべき未指定領域

商用条件: 収益分配構造、上限超過時の従量課金単価、モデル利用量の精算方式。

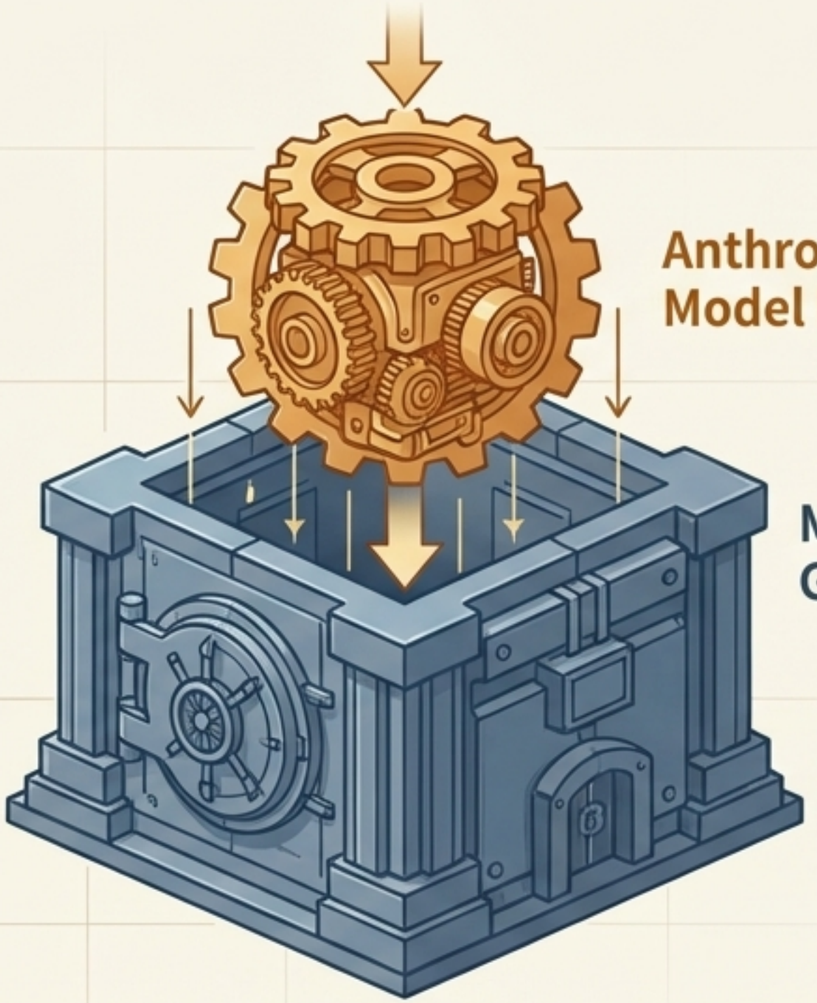
データ境界: Cowork固有のクロスリージョン推論や、ツール連携時のデータ越境・ログ保管場所。

運用SLA: 障害時の一次責任分界点、およびモデル更新の排他・優先順位。

AIの主戦場は「モデルの知能」から「ガバナンスと統合」へ

自律型エージェントの価値は、モデル単体の性能（IQ）だけでは決まらない。

エンタープライズ環境において実運用を可能にするのは、**既存の業務基盤への「統合深度」と、セキュリティ境界という「防具」**の存在である。



Anthropicの頭脳（エージェント技術）

高度な計画能力、マルチステップ実行、コンテキスト理解。
単体では「ローカル隔離VM」での運用となり、企業統制と摩擦が生じる。

＋

Microsoftの鎧（配布網と統制）

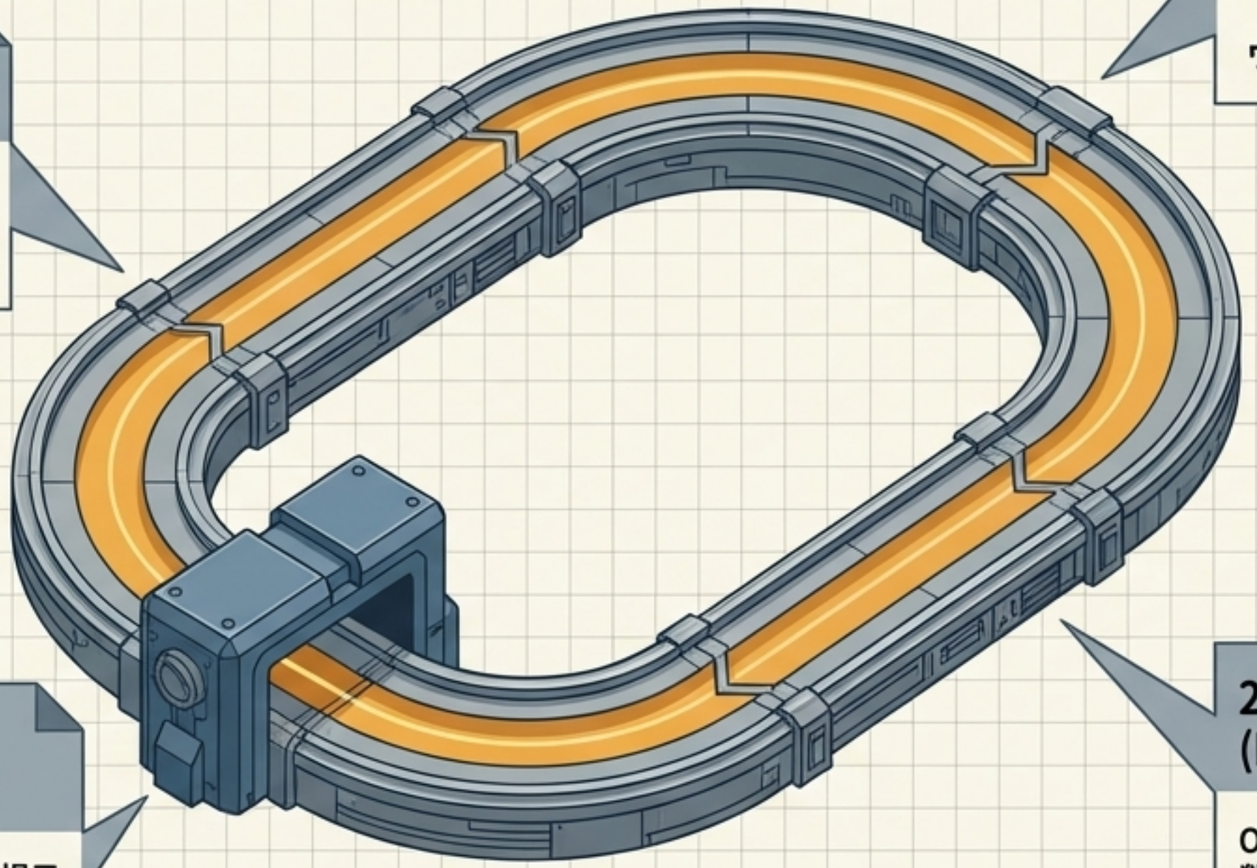
Entra IDによる認証、PurviewによるDLP・監査、M365データ境界。
この「鎧」を着せることで、企業が許容できるクラウド監査モデルが成立する。

Copilot Coworkの「継続実行ループ」

プロンプトではなく「委任と可視化」によるタスク進行

1. 計画 (Plan)

ユーザーが望むアウトカム（成果）を自然言語で指示。AIが業務データ（メール、会議、ファイル）にグラウンディングし、手順を立案。



2. バックグラウンド実行 (Execute)

Outlook、Teams、Excel等を跨ぎ、数分から数時間にわたり自律的にタスクを継続。

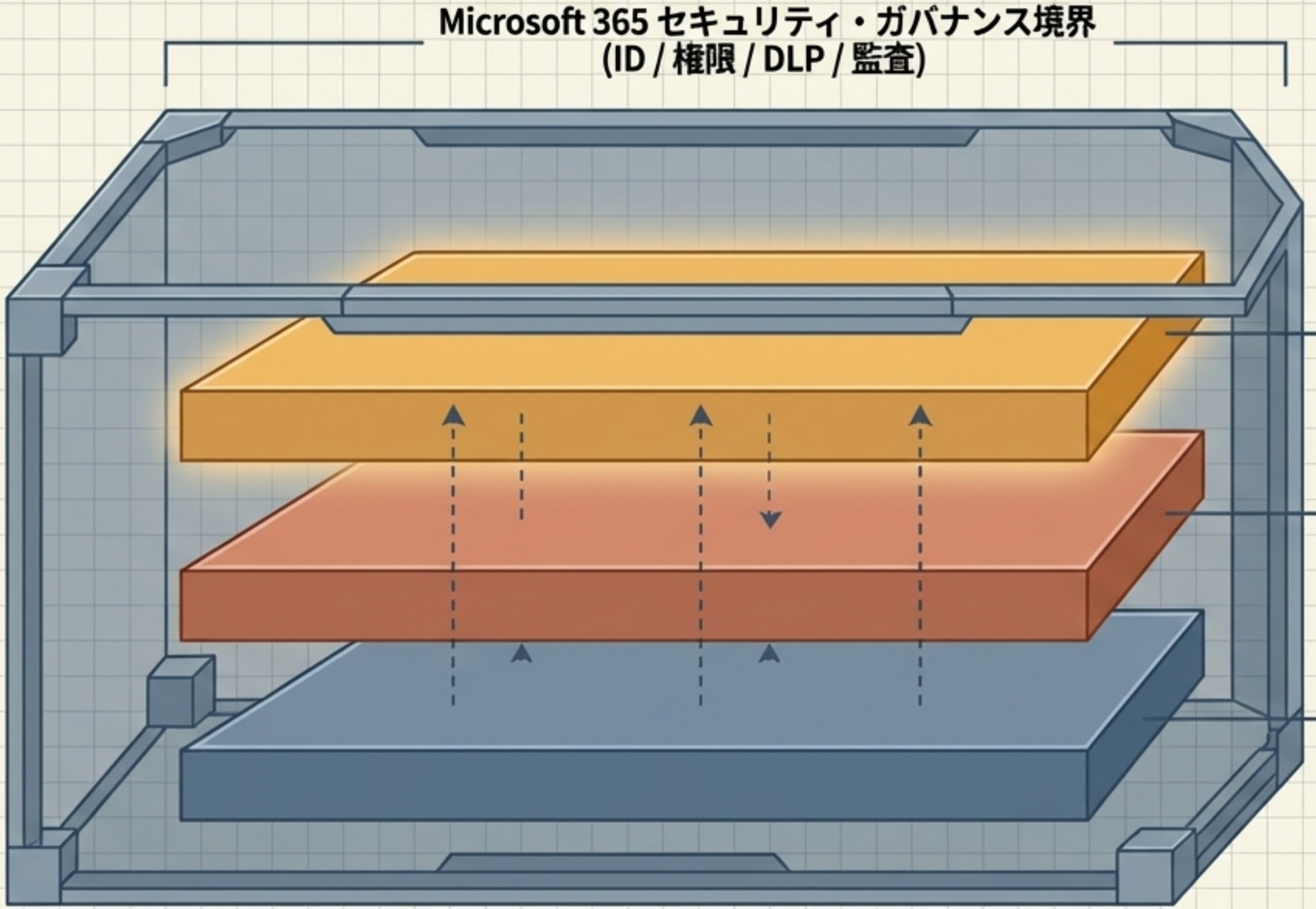
3. チェックポイント (Checkpoint/Human Steer)

進捗状況と推奨アクションをユーザーに提示。ユーザーの介入：承認・修正・停止が可能。

4. 監査可能な適用 (Apply & Audit)

承認された変更のみを適用。すべてのアクションと出力はM365内で監査ログとして記録される。

アーキテクチャ解析：Work IQとガバナンス境界



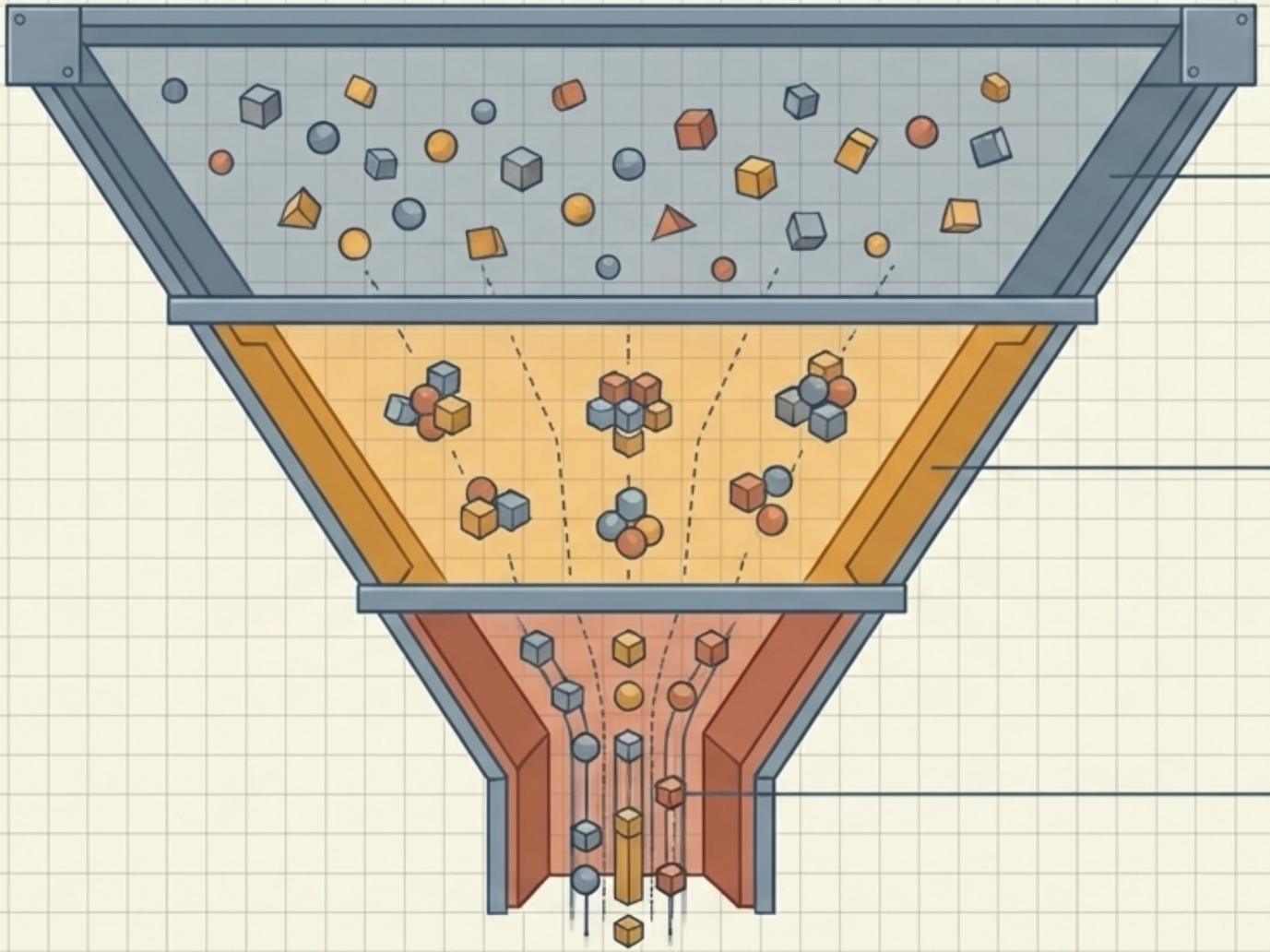
最上位層：Skills & Tools (スキルとツール)
意図を解釈するスキルと、外部に作用するツール群。MCP (Model Context Protocol) サーバー、API、プラグイン連携。

中間層：Context (コンテキスト・知能)
業務の文脈を理解するレイヤー。マルチモデル推論（タスクに応じOpenAI/Anthropicモデルを自動選択・適用）。

最下位層：Data (データ基盤)
M365テナントデータ（SharePoint、Exchange等）。コネクタ経由で取り込まれた外部データ。

コンテキストを再構成する「3層の記憶構造」

無制限のデータ保持を避け、業務文脈を動的に構築するCopilotメモリ設計



明示メモリ (Explicit / Persistent)

ユーザーが意図的に与える情報。カスタム指示や「保存してほしい」と明示された記憶。永続的要素として機能。

暗黙メモリ (Implicit / Behavioral)

活動から推論される洞察。チャット履歴、Teams/Outlook等での過去の活動パターンから抽出される文脈。

グラウンディング (Semantic Search)

都度取得される意味ベースの検索。既存の権限・ラベル・テナント境界を維持したまま、タスク実行時に必要な情報をセマンティックインデックスから瞬時に引き出す。

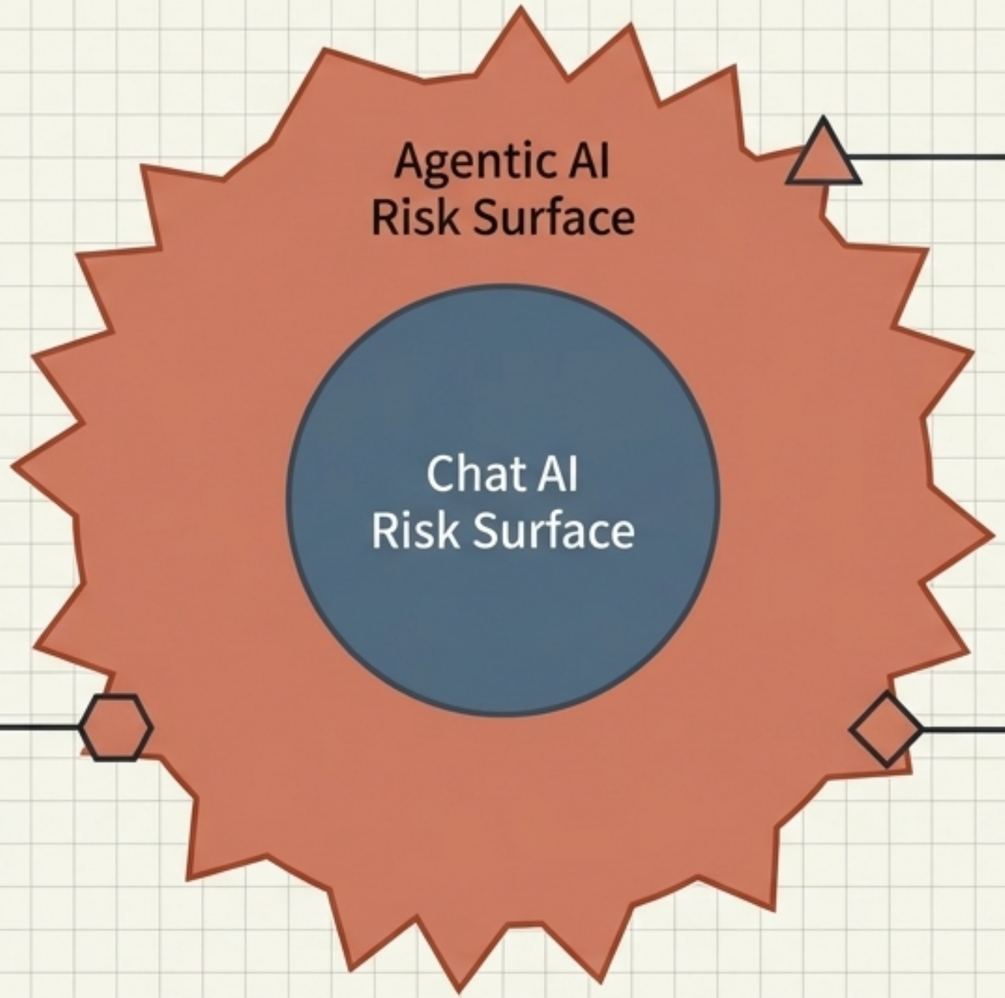
自律型AIランドスケープ・マトリクス

	Copilot Cowork (Microsoft)	Claude Cowork (Anthropic)	M365 Copilot (従来版)	ChatGPT Ent. / Agent (OpenAI)	Workspace Studio (Google)
自律性レベル	高 [🟢]	高 [🟢]	中 [🟡]	高 [🟢]	中〜高 [🟡]
エンタープライズ統制	高 (M365継承) [🟢]	低 (ローカルVM) [⚪]	高 [🟢]	中 [🟡]	高 [🟢]
デプロイ形態	クラウド(保護サンドボックス)	ローカル (隔離VM)	クラウド	クラウド	クラウド

最大の差別化要因：「長時間の自律実行」と「既存のM365権限・監査・DLPの完全適用」の両立。導入可否（CISOの承認）に直結する設計。

自律化に伴うリスクの「面積拡大」

タスクが長時間化し、外部接続が増えることで顕在化する3つの主要リスク



過剰共有の顕在化 (Data Over-sharing)

課題: M365内の権限設定が甘い (共有範囲が広い、機密ラベル不備) 場合、エージェントが「正しく」機密データにアクセスし、結果に含めてしまう本質的リスク。

悪用と外部接点のリスク

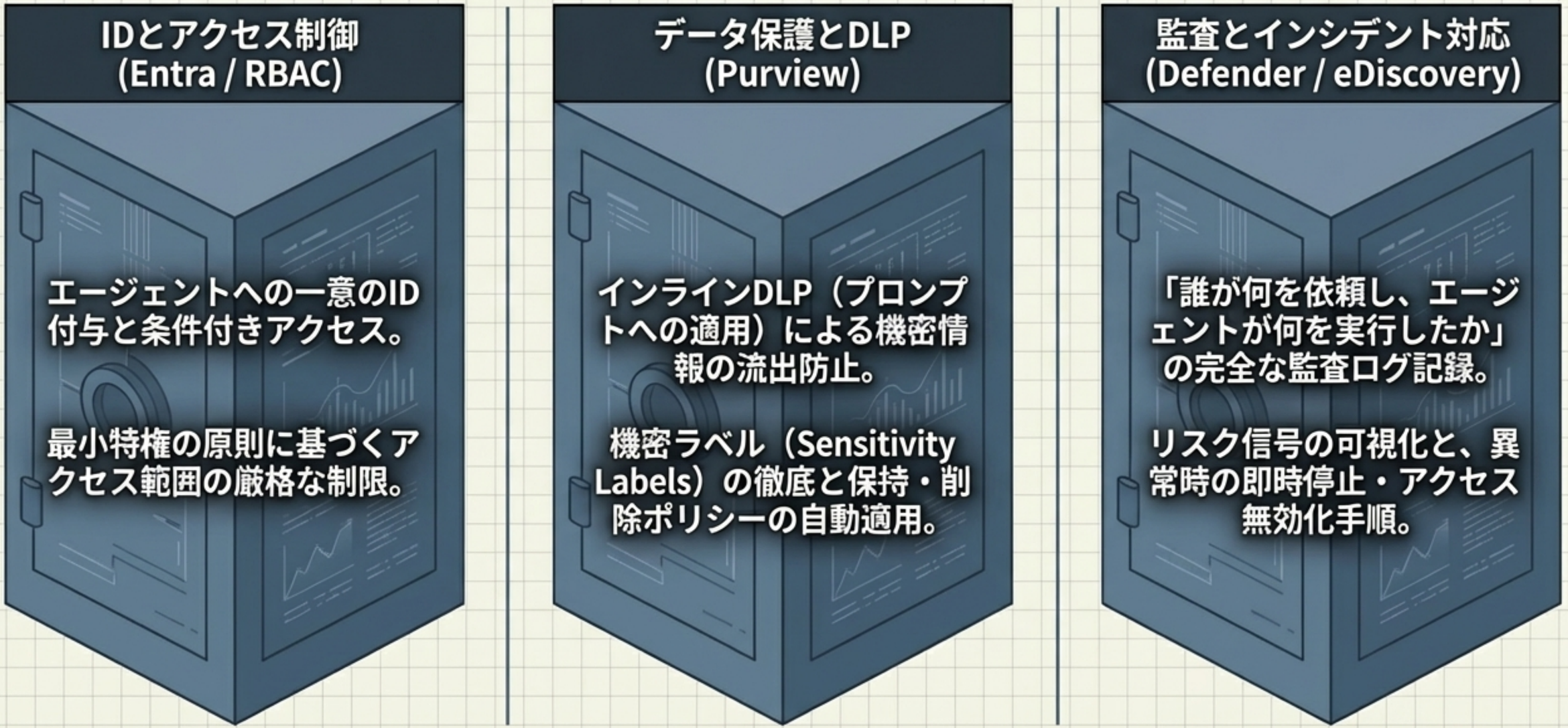
課題: Web検索や外部ツール連携 (MCP等) を通じたプロンプトインジェクション。計画の改変や、エージェトを経由した攻撃連鎖の可能性。

ハルシネーションの業務混入

課題: 自律実行された調査・分析・意思決定支援のアウトプットが、人間による精査を経ずにそのまま業務プロセス（外部送信など）に組み込まれるリスク。

The Fortress：Agent 365 による防衛と統制ライン

エージェント増殖による可視性低下を防ぐ「コントロールプレーン」



権限とリスクに応じた「3段階のロールアウト戦略」

Step 1: 低リスク・内部完結型 (Short-term)

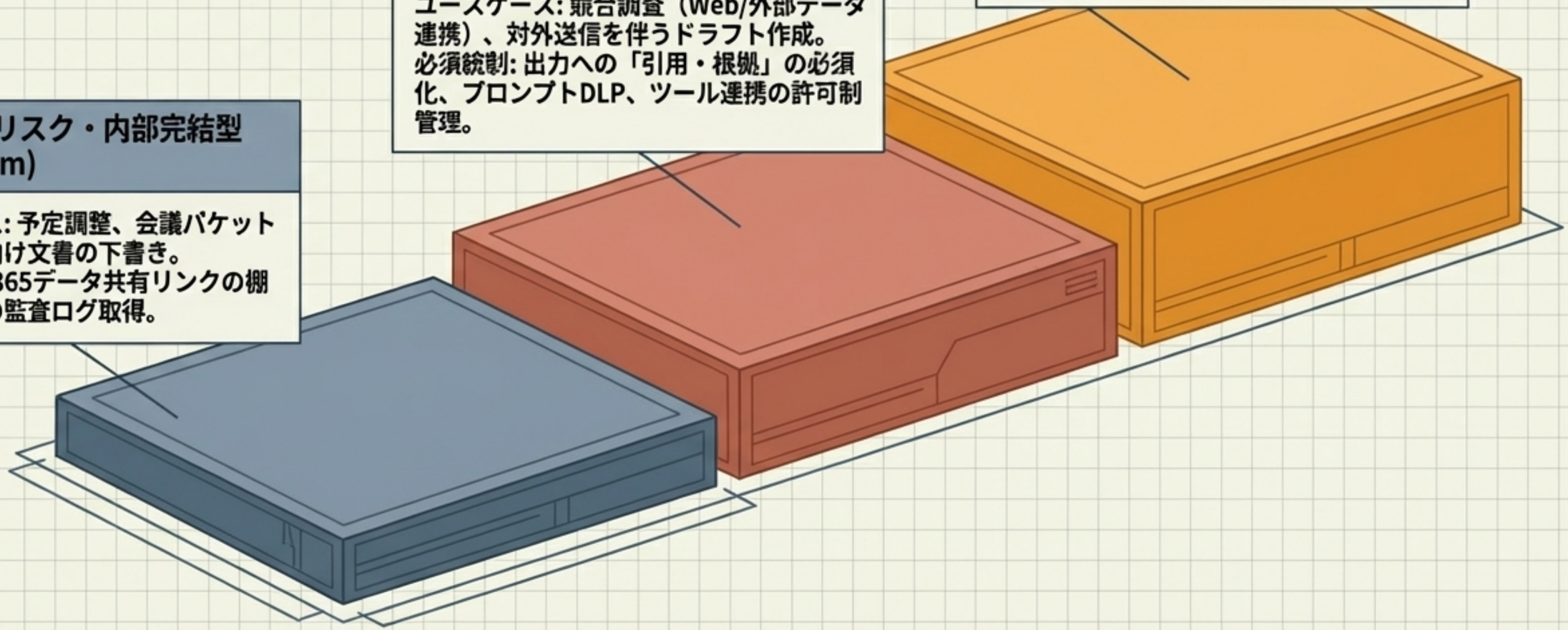
ユースケース: 予定調整、会議バケット作成、社内向け文書の下書き。
必須統制: M365データ共有リンクの棚卸し、基本の監査ログ取得。

Step 2: 中リスク・外部接点型 (Mid-term)

ユースケース: 競合調査（Web/外部データ連携）、対外送信を伴うドラフト作成。
必須統制: 出力への「引用・根拠」の必須化、プロンプトDLP、ツール連携の許可制管理。

Step 3: 高リスク・意思決定型 (Long-term)

ユースケース：ローンチ計画、財務分析、サプライチェーン連携タスク。
必須統制：重要数値の二重検証（別ツール再計算）、人間によるレビュー（Human-in-the-loop）通過の強制。



導入に向けた実務・セキュリティ・チェックリスト

データ・ガバナンス（Data Readiness）

- ☐ SharePoint / OneDrive内の過剰な共有リンクの棚卸しと無効化が完了しているか。
- ☐ Purviewによる機密ラベル（Sensitivity Labels）が全社データに正しく適用されているか。

セキュリティ・アクセス統制（Security & Access）

- ☐ Agent 365 または M365 E7 を活用したエージェントへのアクセス制御・ID管理方針が定義されているか。
- ☐ プロンプトおよび出力に対するDLP（情報漏洩防止）ポリシーが設定・テストされているか。

監査・運用管理（Audit & Operations）

- ☐ 外部ツール連携（MCPやプラグイン）の申請・許可プロセスの台帳管理が整備されているか。
- ☐ eDiscoveryを活用し、「誰が・どのエージェントに・何をさせ・何が変更されたか」を追跡できる状態か。
- ☐ ハルシネーションによる業務影響を防ぐための「引用必須」「人間による承認必須」のワークフローが定義されているか。