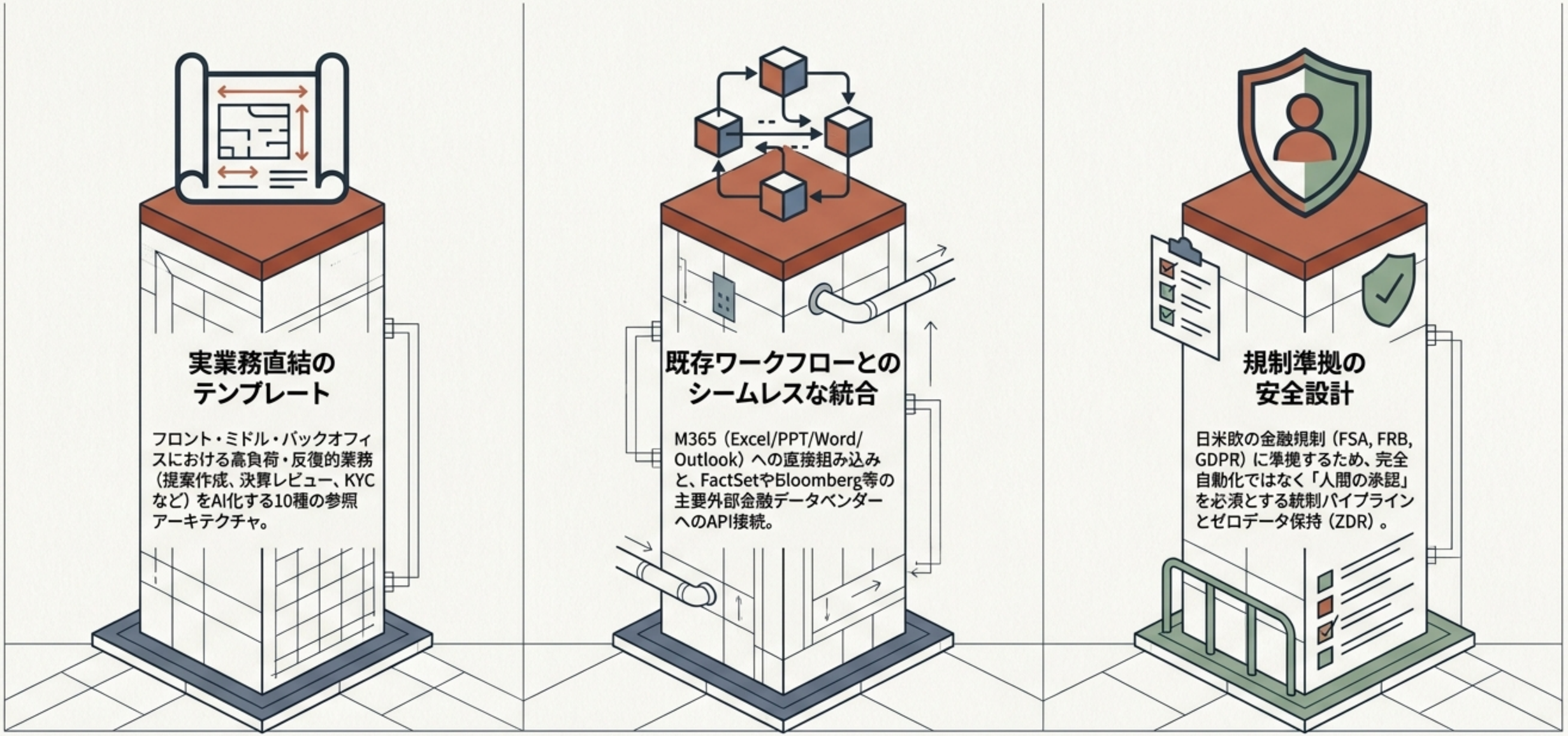
金融業務AIの実装ブループリント

Anthropicが提示する10のエージェントと「人間の介在（Human-in-the-loop）」を核とした次世代オペレーションOSの構築

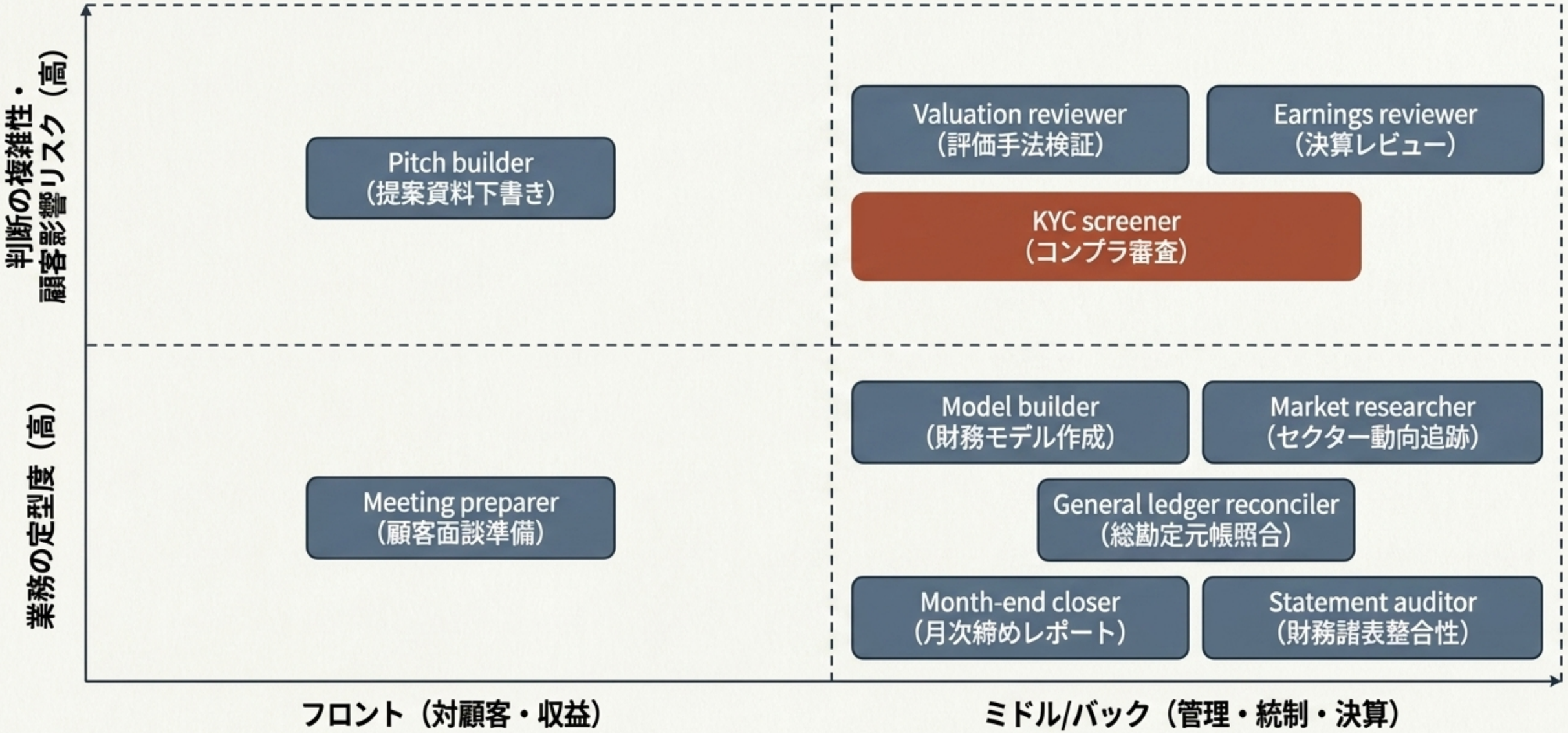
経営層・リスク管理・DX推進責任者向け戦略評価レポート

単なるAIツールから「金融業務OS」への進化

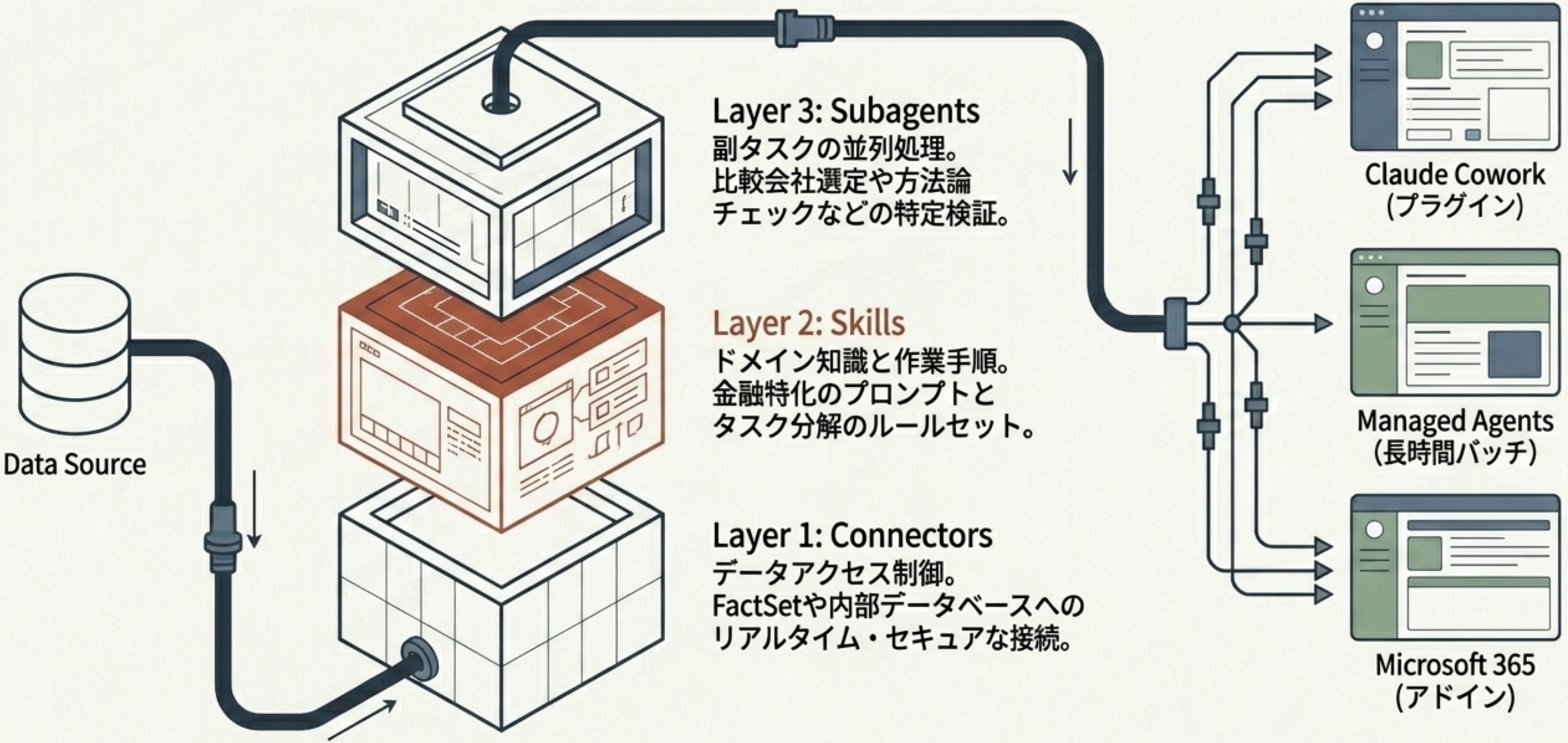


10の金融向けエージェント：リスクと機能の分類

PoCの留意点：KYCやValuationなど顧客影響リスクが高い領域は、初期導入において自律実行ではなく「例外抽出・下書き生成」に限定することが推奨される。

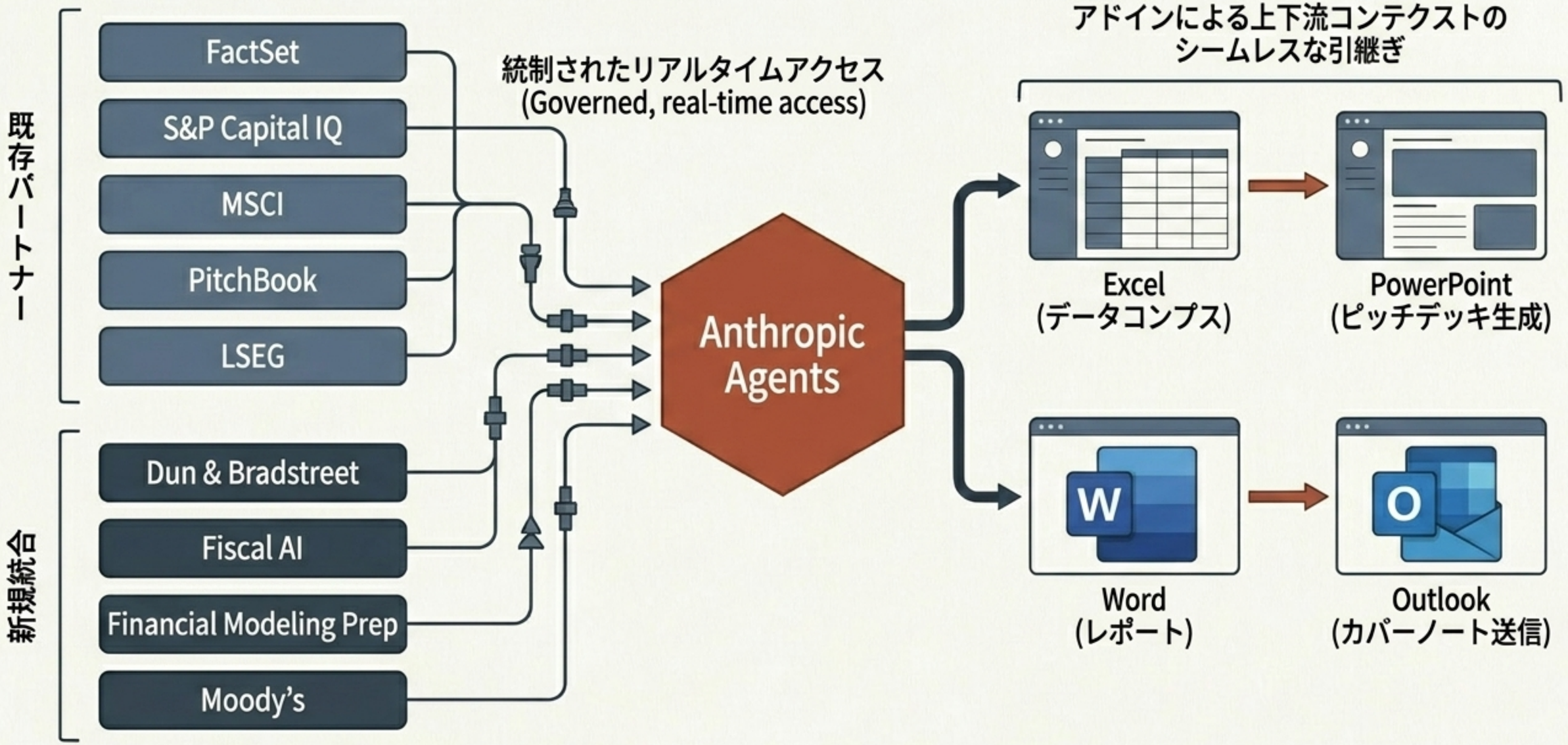


エージェントの技術的アーキテクチャ（3層レイヤー）



※推奨モデル: Claude Opus 4.7

データ・エコシステムとM365統合



グローバル金融規制とAI対応マトリクス

米国（FRB / OCC） SR 26-2（モデルリスク管理）， OCC 2023-17（第三者リスク管理）	実サービス単位の責任分界、文書化、ベンダー監督による対応。
欧州（EU AI Act / GDPR / DORA） 自動意思決定制限（GDPR Art 22）， 高リスクAI指定, ICTリスク管理	DPIA要件の履行、DORAに準拠した障害時復旧計画と第三国委託管理。
日本（FSA） 2025 AI Discussion Paper	顧客影響業務における「人間の介在（Human-in-the-loop）」とAgile governanceの実装。

現実解は「完全自律化」ではなく、既存のモデルリスク管理原則を拡張し、AIの出力に対する説明責任（Accountability）を担保する設計である。

導入における4つの主要リスクとエンジニアリング対策

誤情報・幻覚（Hallucination）

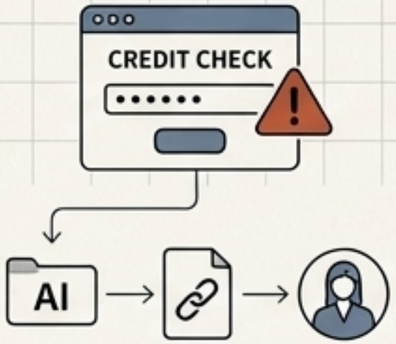
事実
金融特化ベンチマークでも最良AIの精度は48.9%であり専門家より低い。

対策
最終判断の自動化を避け、起案・照合・例外抽出の重労働に特化させる。

バイアスと説明責任（Accountability）

事実
KYCや信用審査における不透明な推論リスク。

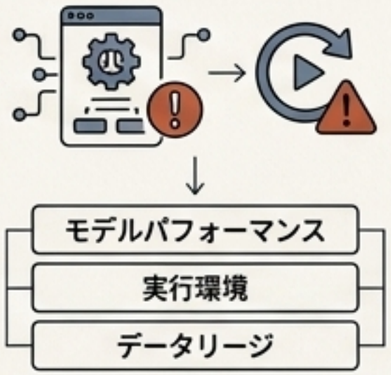
対策
全出力への原典（ソース文書）リンク付与と、人間による二次審査フローの実装。



運用リスク（Operational Risk）

事実
Managed Agentsの途中停止や再実行時の再現性不足。

対策
モデル外の実行リスクを包含したモデル台帳（Inventory）の再定義と構築。

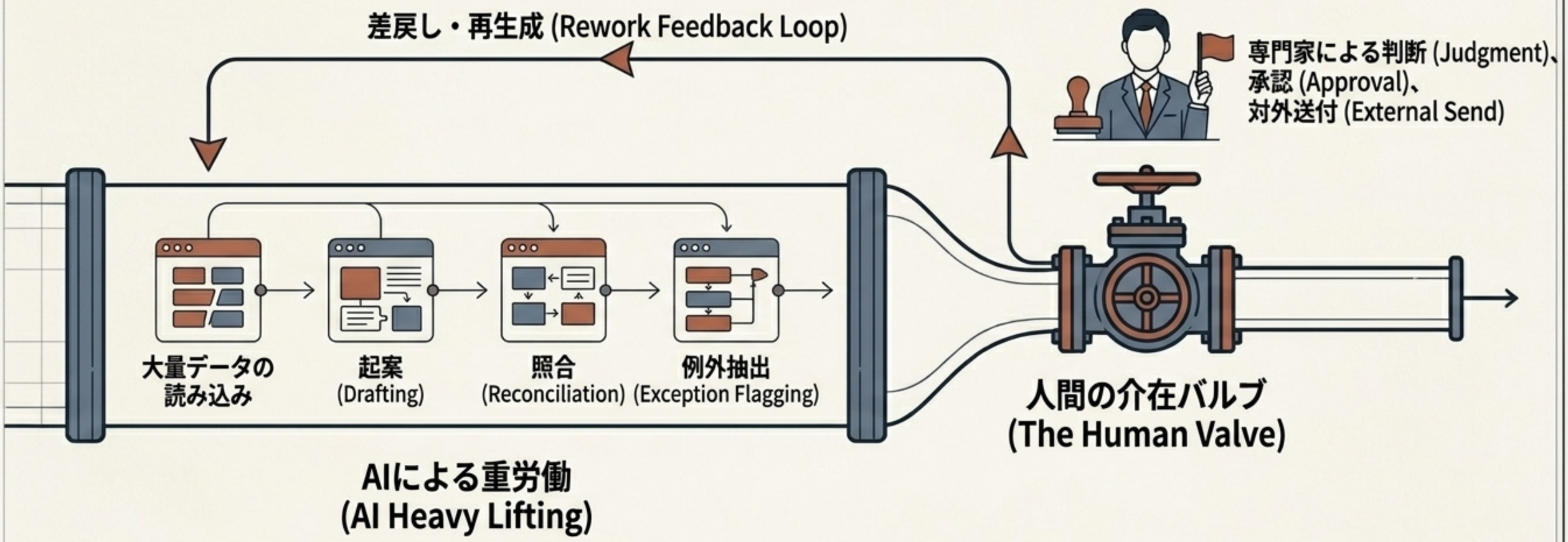


サイバー・データ保護（Data Protection）

事実
外部データコネクタ増加に伴う攻撃面（Prompt injection等）の拡大。

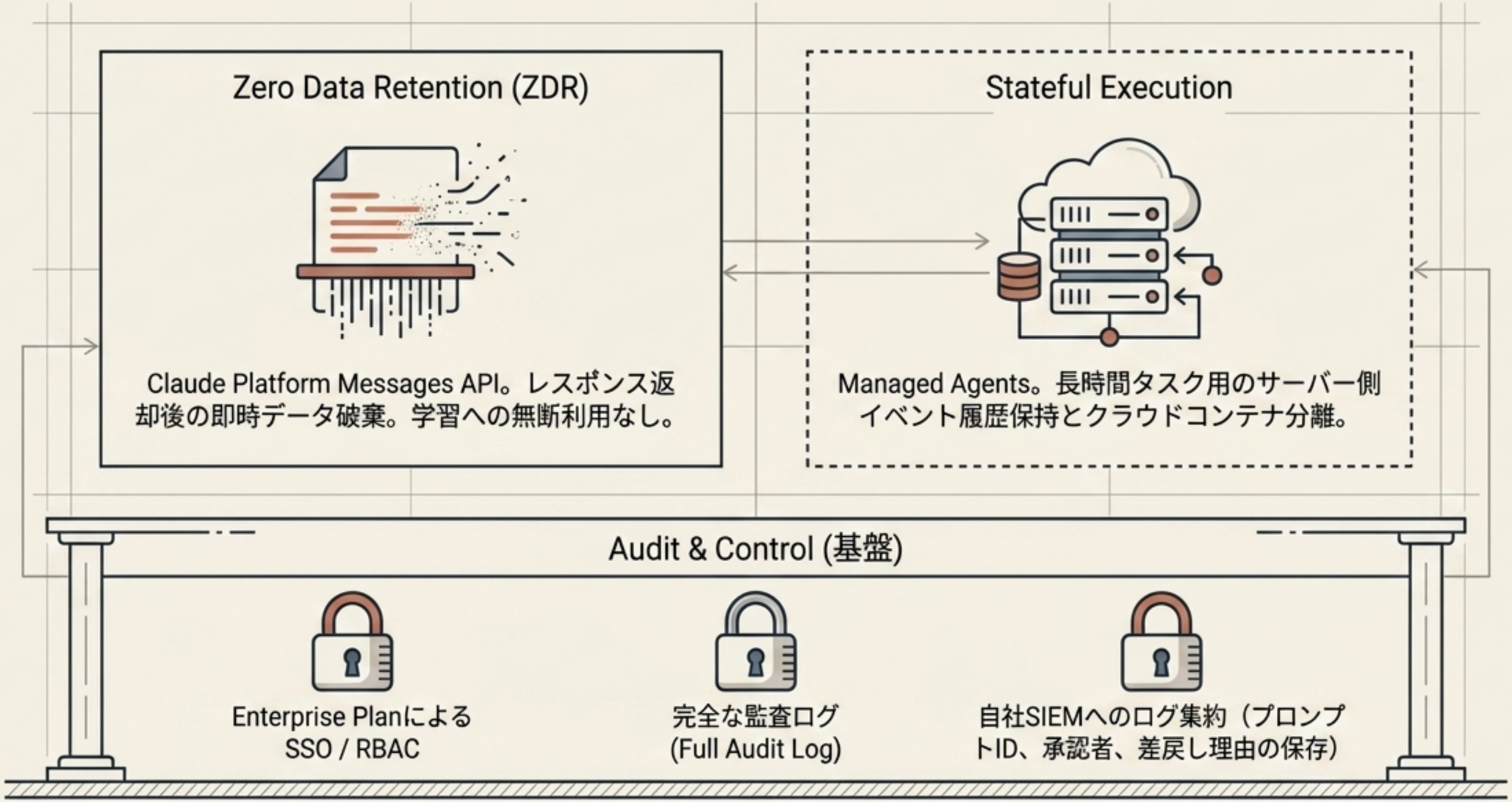
対策
接続先ごとの厳格な認可スコープ設計と、M365周辺の境界防御。

人間の介在（Human-in-the-Loop）バルブ・アーキテクチャ

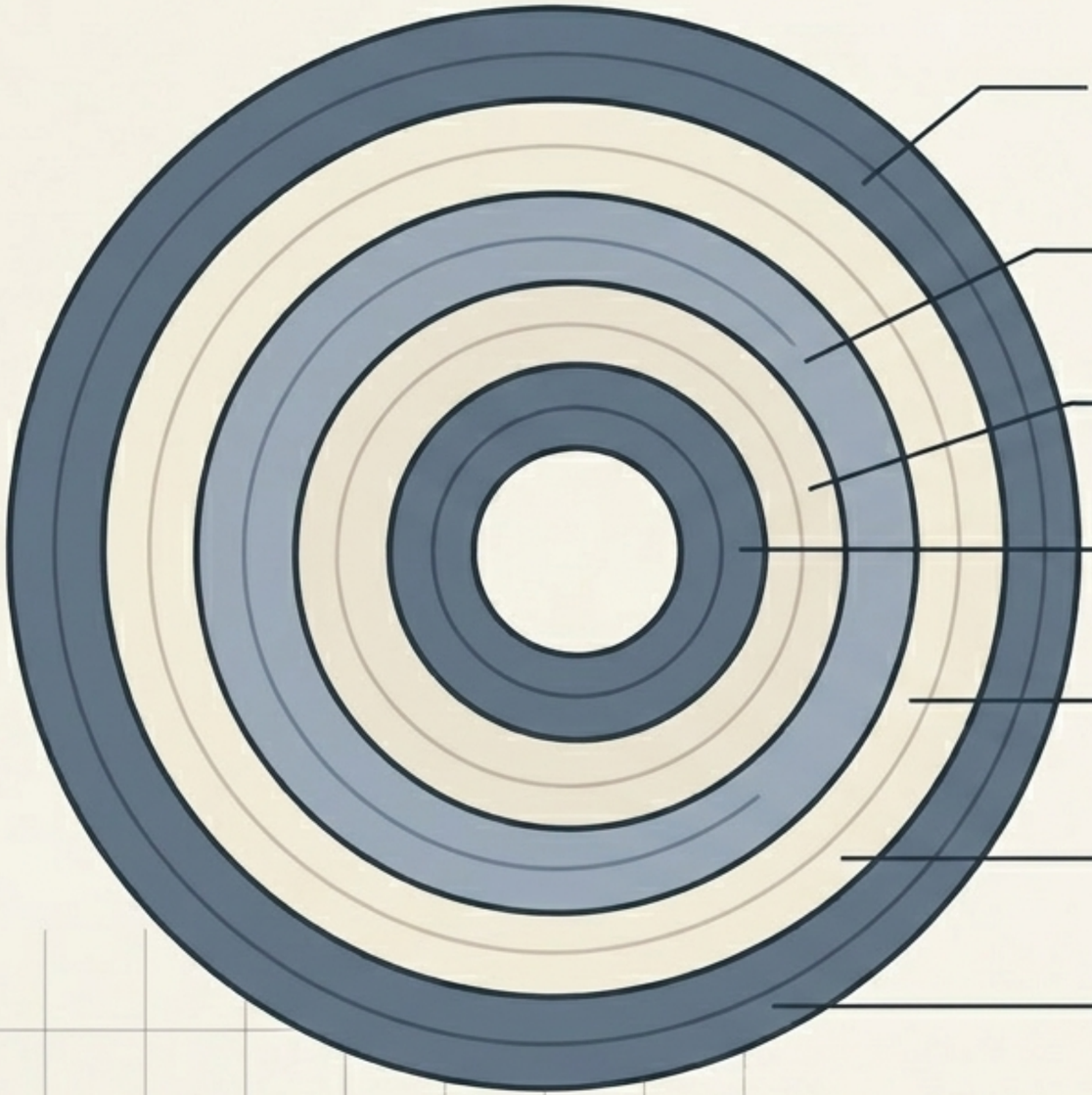


AIの真の価値は『自律性』ではなく、このバルブを通過するまでの『レビューコスト・工数を劇的に下げること』にある。

データ保護とセキュリティ・アーキテクチャ



AI実運用のコスト構造モデル（The Cost Onion）



運用・コンプライアンス監査・
ログ保存コスト

導入・PoC環境構築・
検証コスト

APIトークン従量課金
(Prompt caching/Batch
処理による最適化含む)

席課金（Pro $20/mo,
Max $100/mo 等）

外部データコネクタ・
金融ベンダーAPI接続料

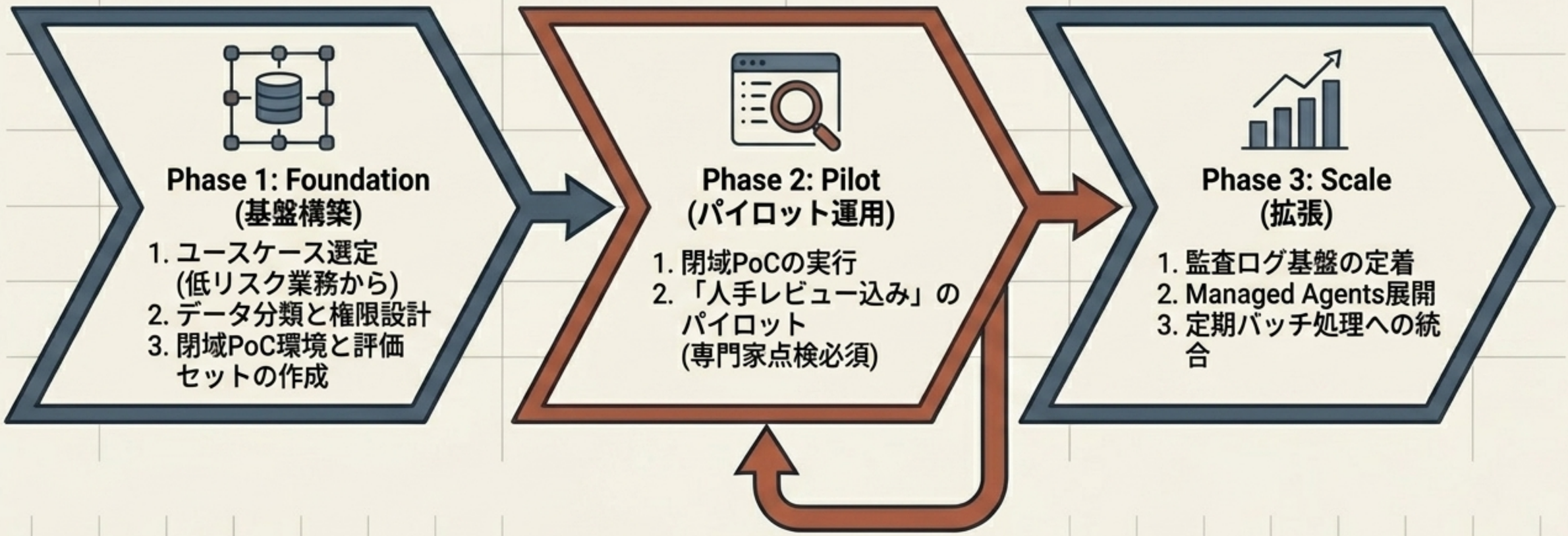
導入・PoC環境構築・
検証コスト

運用・コンプライアンス監査
・ログ保存コスト

財務上のインサイト

単純なSaaSの月額料金だけでは予算化は不十分。APIの再試行率や、金融文書の長大なコンテキストサイズがコストを牽引するため、Layer 2（トークン消費）とLayer 3（データ接続）の精緻なモニタリングがROI達成の鍵を握る。

安全で確実な3段階の導入ロードマップ



PoC評価ダッシュボード：4つの定量KPI

Metric 1: 精度と完全性

自社特有の例外パターンやフォーマットへの追従率。
誤検知・見逃しの割合を評価。

Metric 2: 再現性と根拠提示

複数回実行時のブレの無さ。出力に原典
（ソース文書）への正確なリンクが含まれるか。

Metric 3: 差戻し率 (Rework Rate)

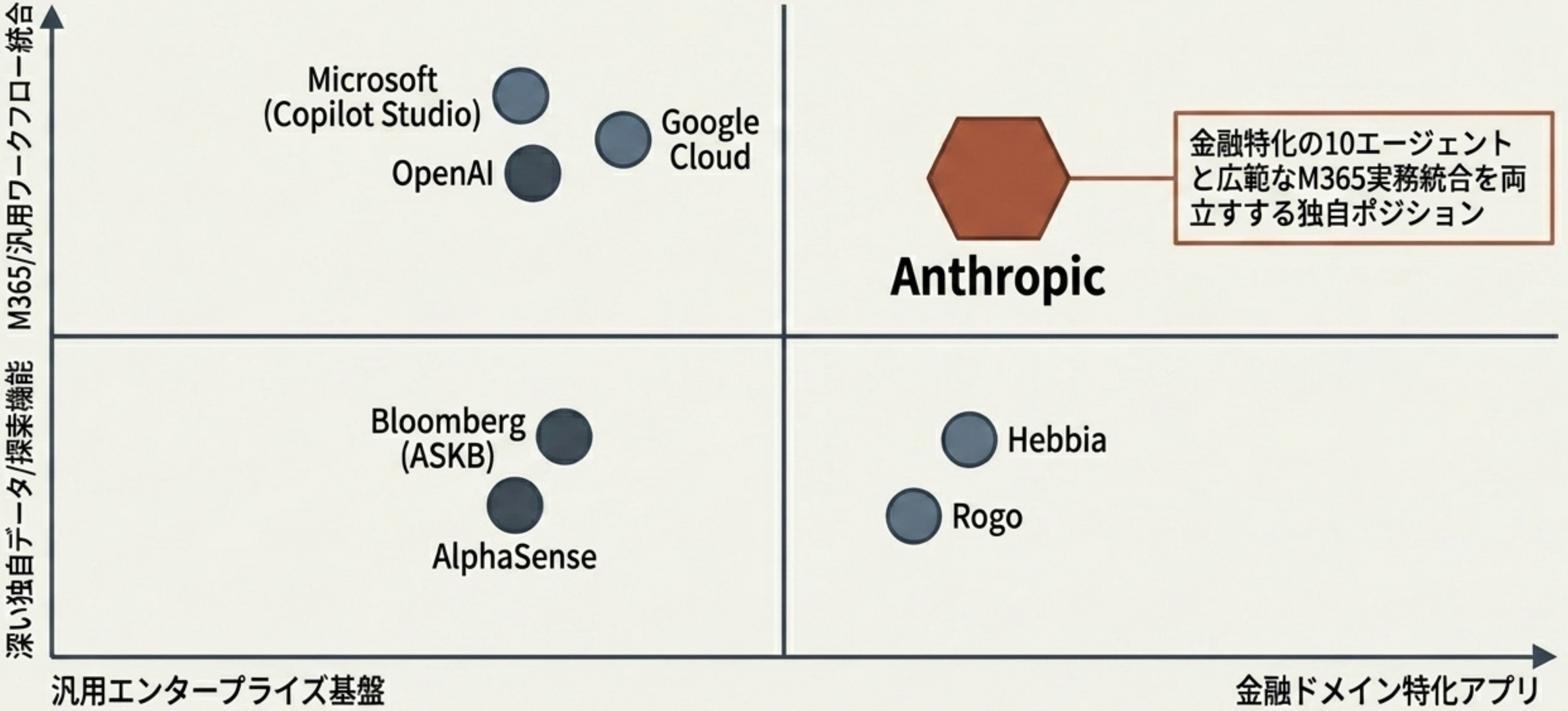
人間（専門家）がAIの出力に対して修正・再指示
を出した割合。低いほど優秀。

Metric 4: レビュー時間短縮率

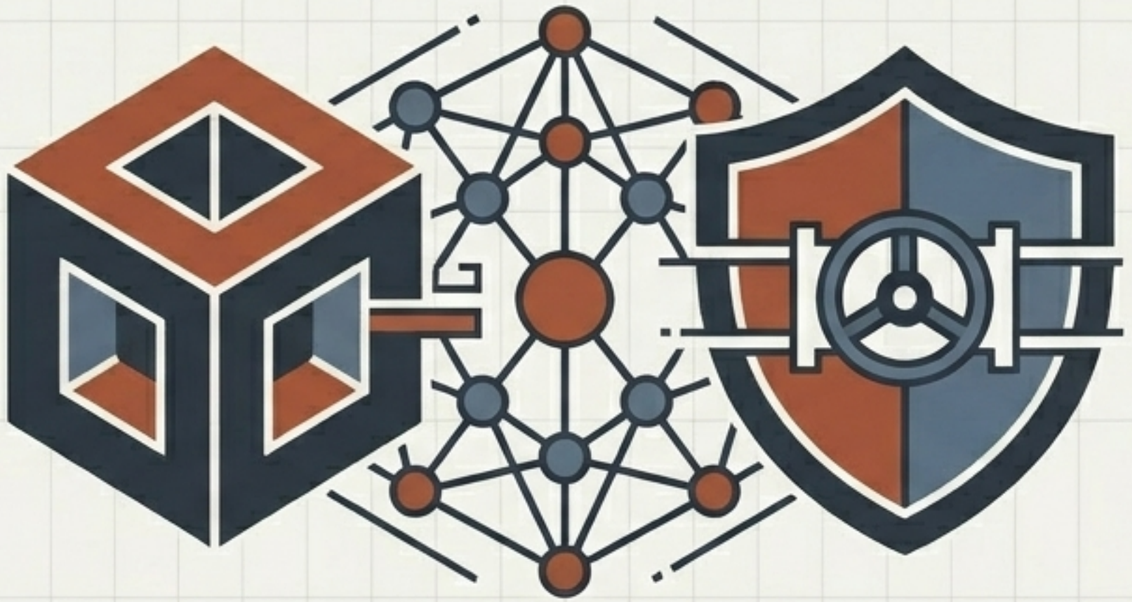
ゼロからの作成と比較し、AI起案物をレビューする
ワークフローによる純粋な工数削減効果。

※一般的なベンチマークではなく、「自社データ・自社テンプレート」での閉域評価が規制整合性の前提条件となる。

AIプロバイダー競合ランドスケープ



「最強モデルの導入」から「金融業務OSの構築」へ



Anthropicの真の価値は、モデル単体の性能ではありません。既存のデータ基盤・Office実務・そして人間の承認フローの間に、AIを安全かつシームレスに「サンドイッチ」する参照アーキテクチャを提供したことにあります。

Action: 完全自動化の幻想を捨て、専門家の「レビューコスト」を下げるための内部基盤設計から直ちに着手すべきです。

The New Standard for Financial Operations.