AI事業者ガイドライン（第1.2版）の深掘り：自律型AI時代のガバナンスと企業戦略

自律型AI時代におけるガバナンス、Human-in-the-Loopの義務化、企業戦略のアクションを視覚的に解説。

ガバナンス対象の劇的な拡張とリスクの変容

第1.1版（受動的AI）

従来のWeb上の生成AI。
受動的なツールとしての利用。

第1.2版（自律型AI）

AIエージェント（自律的実行）

採用審査や金融判断を代行。ハルシネーションによる誤送金、バイアスによる不当判断の自動実行リスク。

フィジカルAI（物理世界で動作）

自動配送ロボット、自律走行車。
誤作動による身体的危害、財産的損害のリスク。

システム設計の必須要件「Human-in-the-Loop」

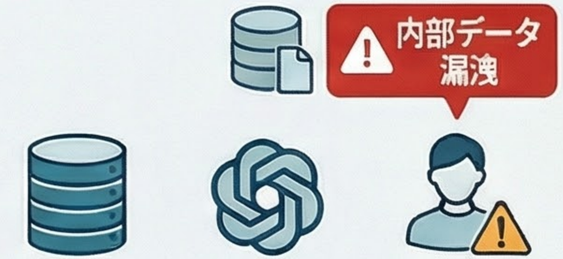
Human-in-the-Loop（HitL）の義務化

AIが不可逆的な外部操作を行う前に、必ず人間が確認・承認するプロセス。企業としての説明責任（アカウンタビリティ）を担保。

承認ゲートキーピングの境界線

文書ドラフトの「作成」はAIの自律性を許容するが、「送信」直前には必ず人間の介在を求める。

RAG（検索拡張生成）における新たな脅威



企業の61%がガイドラインを認知、35%が会社活用。RBAC不備による機密情報の漏洩リスク。トレーサビリティ（いつ、誰が、どのデータで判断したか）の確保が不可欠。

各主体に求められる責務マトリクス

開発者（Developers）	提供者（Providers）	利用者（Users）
セーフティ/プライバシー・バイ・デザインの導入、学習データの公平性確保、技術特性の文書化	モデルドリフトの継続的な監視、施話性対策、利用規約やオプトアウトの明示	人間による合理的な最終判断(HitL)、機密情報の入力防止、ステークホルダーへの説明窓口設置

今すぐ実行すべき3つの戦略的アクション

1. AI利用状況の網羅的な棚卸し

シャドーAIを特定・排除し、利用台帳を作成して公式な管理下に置く。

2. 責任範囲と「自律の境界線」の策定

AI単独で完結させてよい業務と、必ず人間の確認が必要な業務（顧客送信、契約締結等）の線引き。

3. 企業独自の社内ガイドライン更新

入力禁止情報の定義、インシデント発生時の連絡体制を含む、実施的なポリシーを全社に周知徹底する。