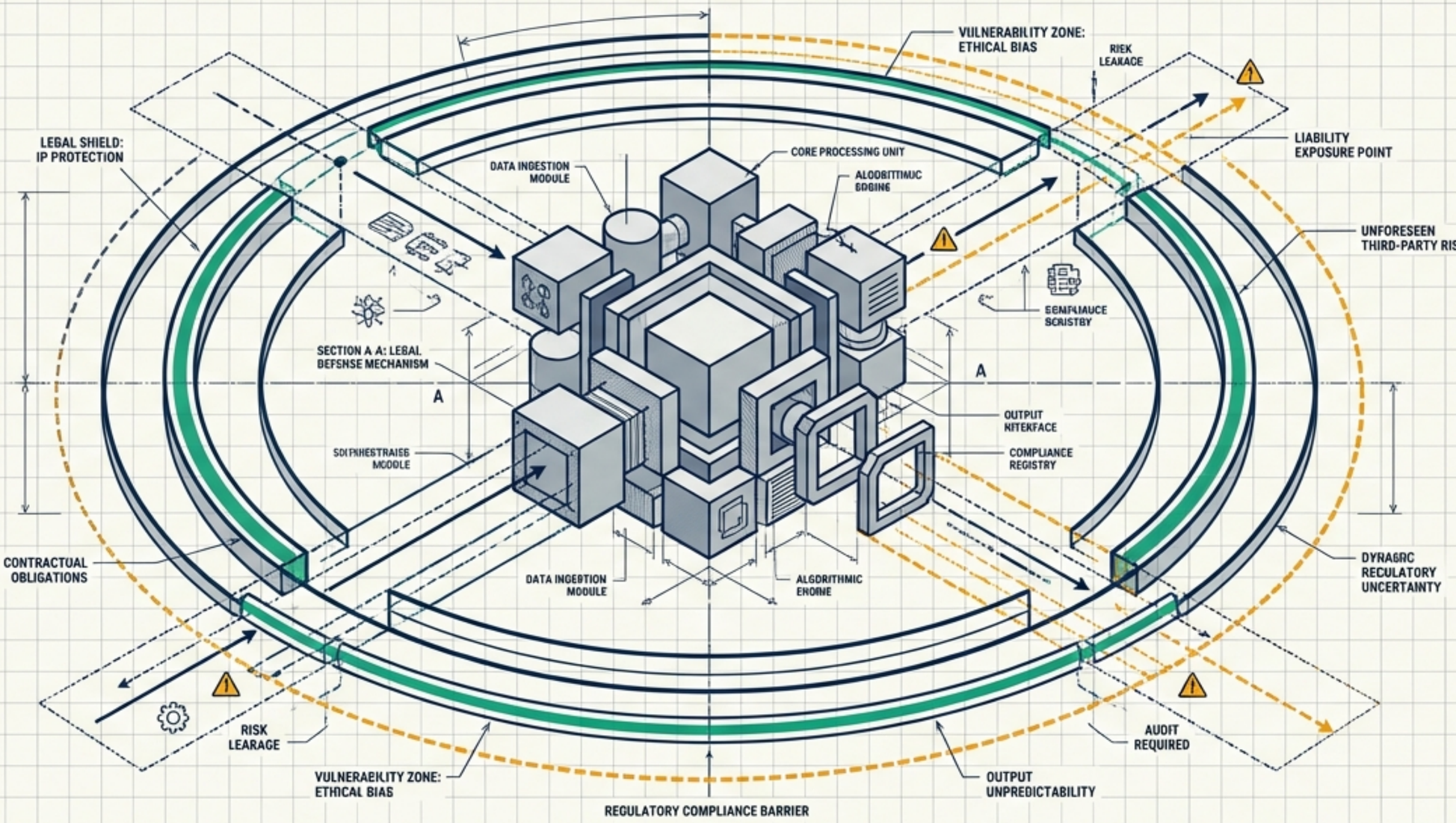
# AI利活用における法的リスクのアーキテクチャ

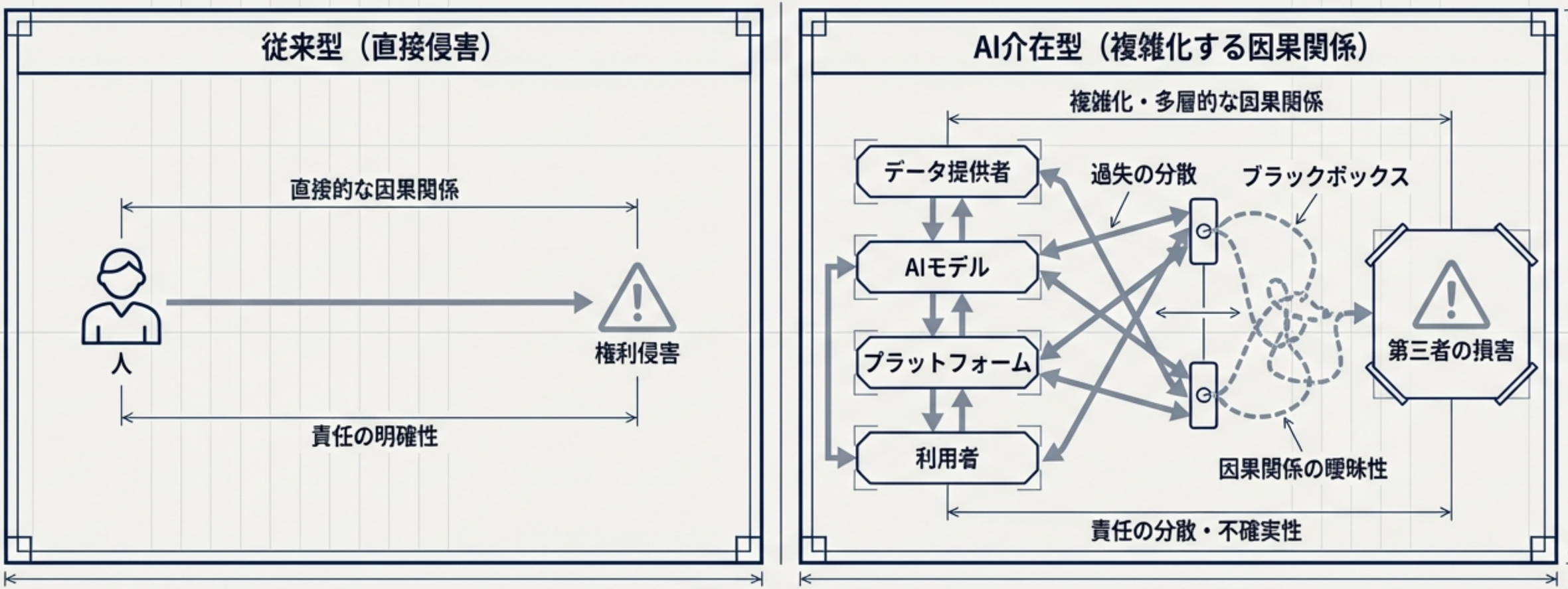
## 経済産業省「AIガイドライン」を実装する知財・コンプライアンスの設計図




株式会社ストラテジック・コンプライアンス / 2026年5月

# AI時代のデフォルト・ルール：第三者損害への対応

知財侵害紛争の典型は「明確な契約ルールが及ばない第三者との不法行為（民法709条）」




## 起点

契約関係のない第三者の権利侵害リスク（著作権、特許、営業秘密等）。

## 争点

AIの介在により「誰の過失か（注意義務の所在）」と「因果関係」の特定が難化。

## 解決策

法律論を「AIの運用設計（アーキテクチャ）」の問題として再定義する。

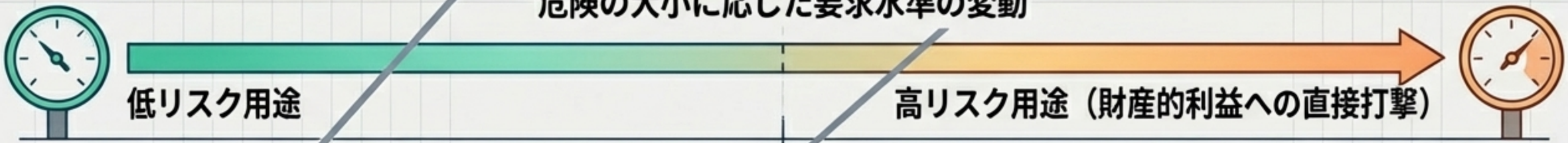
# 運用パラダイムの転換：2つのAI類型

| | 補助／支援型AI | 依拠／代替型AI |
|---|---|---|
| 前提 | 最終的に「人の判断・行動」を介在させる | 人の判断を代替し、AI出力に直接依拠する |
| 注意義務の重心 | 個々の出力の「正誤・権利侵害チェック」 | システム全体の「体制構築・運用・監視」 |
| 典型的な用途 | 商標候補生成、広告画像生成、RAGでの社内要約 | 自動公衆送信（自動出品）、広告の完全自動配信 |
| デフォルトの姿勢 | 知財実務における推奨ベースライン（ヒト介在必須） | リスク増大。プロバイダ側の説明・ガードレール義務が強行化 |

# AI不法行為（民法709条）の法的方程式

**過失 ＝ 予見可能性 ＋ 結果回避義務違反**

危険の大小に応じた要求水準の変動



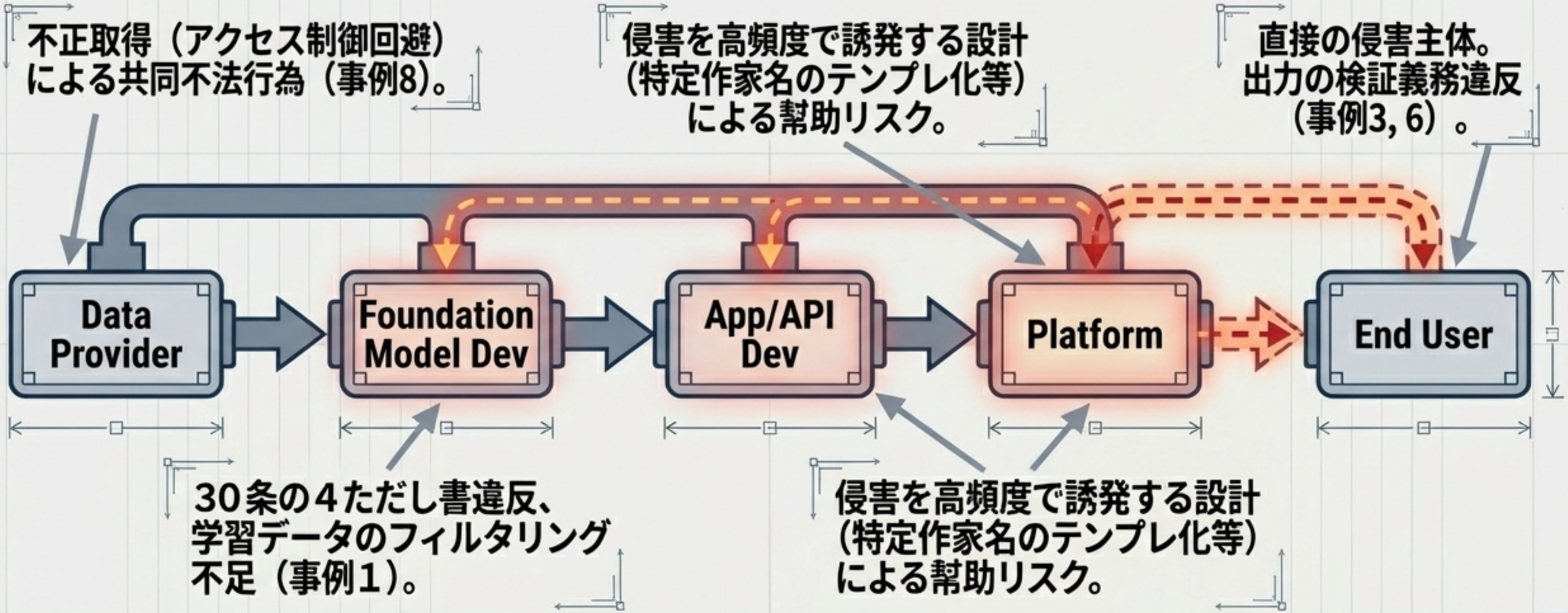

**予見可能性**
侵害リスクを事前に把握できたか？
（例：特定のプロンプトが既存知財を誘発する可能性）

**結果回避義務違反**
リスクを下げる合理的措置をとったか？
（例：フィルタリング、ヒトによる最終確認）

**重要インサイト**

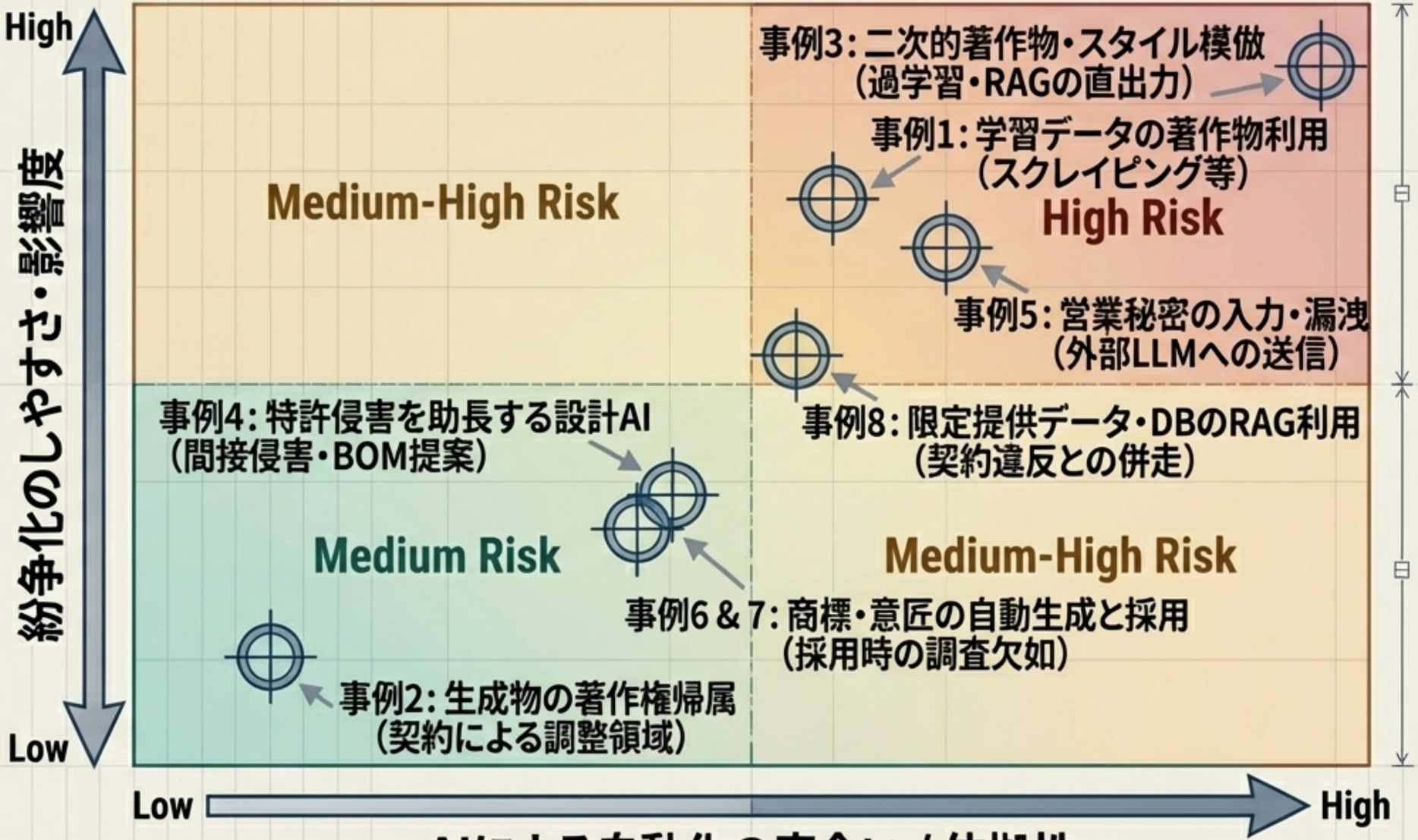
知財侵害のように『財産的利益』が問題となる場面でも、AI関与のしかた（人の介在度合い）に応じて注意義務の要求水準が変動する。

# 共同不法行為の連鎖（民法719条）：バリューチェーンの分業と責任




**単独の不法行為だけでなく、システムの「設計の欠陥」が連帯責任（共同不法行為・幇助）を引き起こす。**

# 知財領域・8つの想定事例リスクマップ

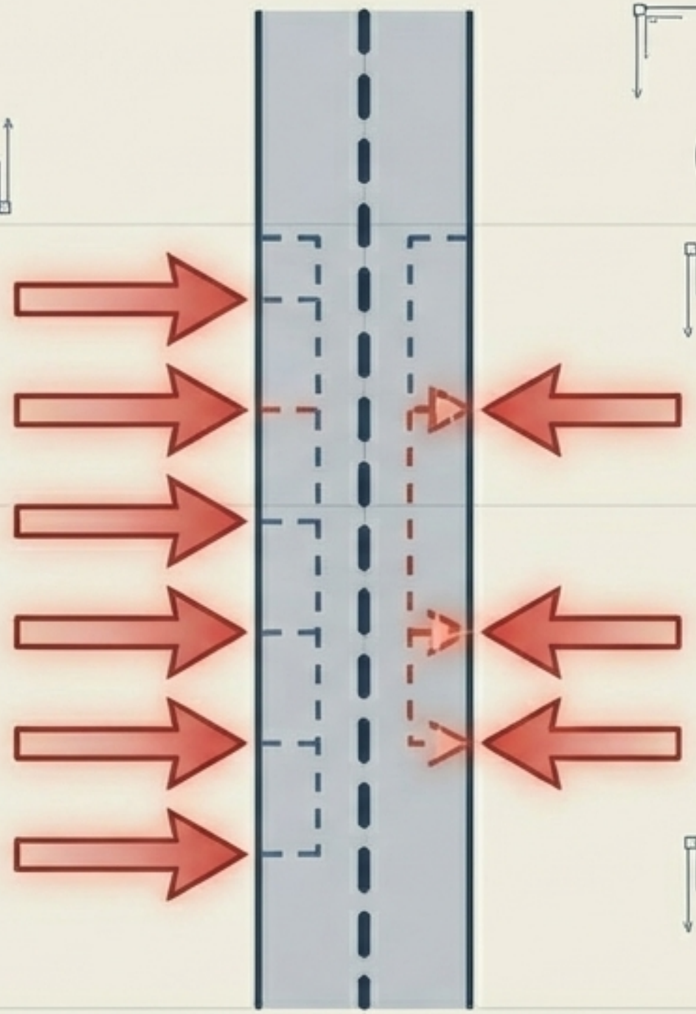


# インプットの脆弱性：データ境界線の突破

## 漏洩リスク
## (Data Escaping)

事例5（営業秘密）：
従業員が外部LLMに未公開ソースコード等を入力。
秘密管理性の喪失リスクと契約違反。

対策：AI利用前提の追加統制（DLP、入力禁止カテゴリ、ログ保持）。



Corporate Firewall
(社内ネットワーク境界)

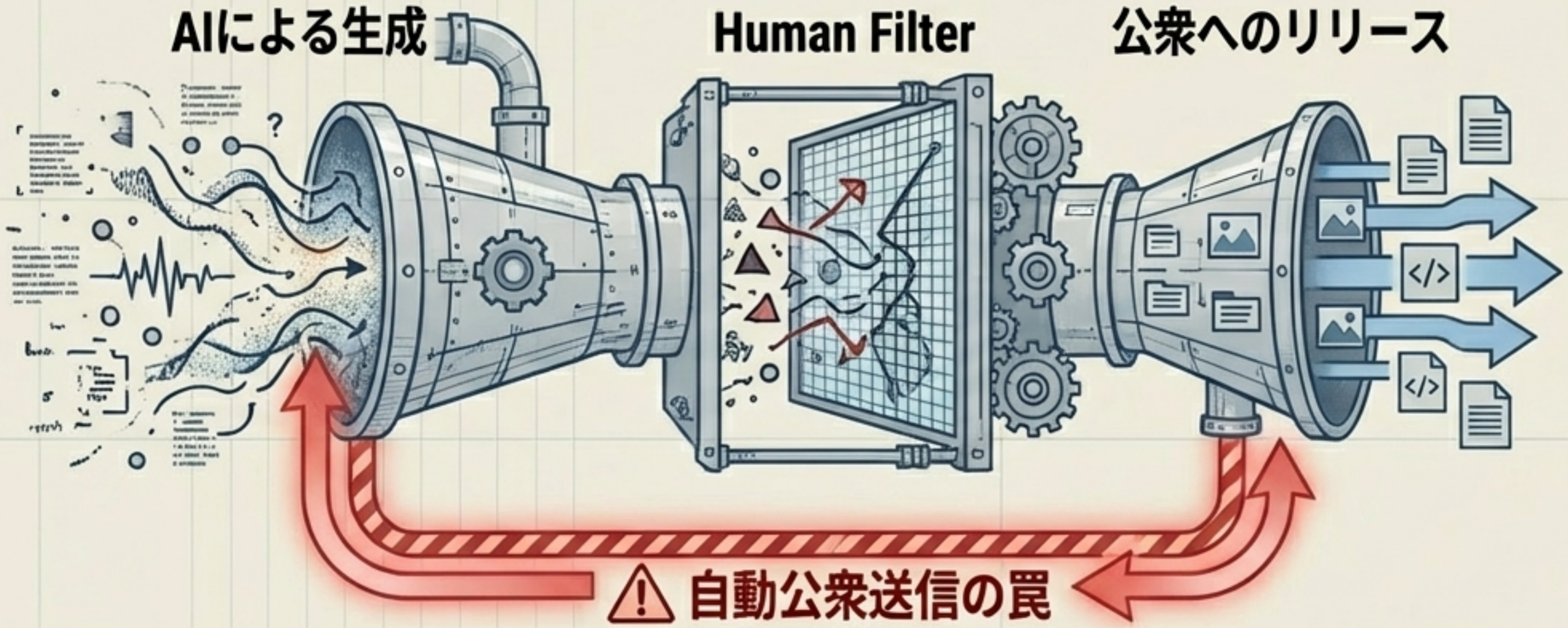
## 侵害取得リスク
## (Restricted Data Entering)

事例1（クローズドデータ）：
認証回避等による学習データの不正取得。30条の4「不当に害する」への該当。

事例8（限定提供データ）：
RAG用に有償DBから大量抽出。
目的外利用禁止条項違反。

対策：robots.txtの遵守、利用市場との市場衝突評価。

# アウトプットの脆弱性：生成物の無検証リリース




依拠/代替型AIとして出力が自動的に公衆へ届く設計は、結果回避義務の重心が**「提供者側の設計・停止機構」**へ跳ね上がる。

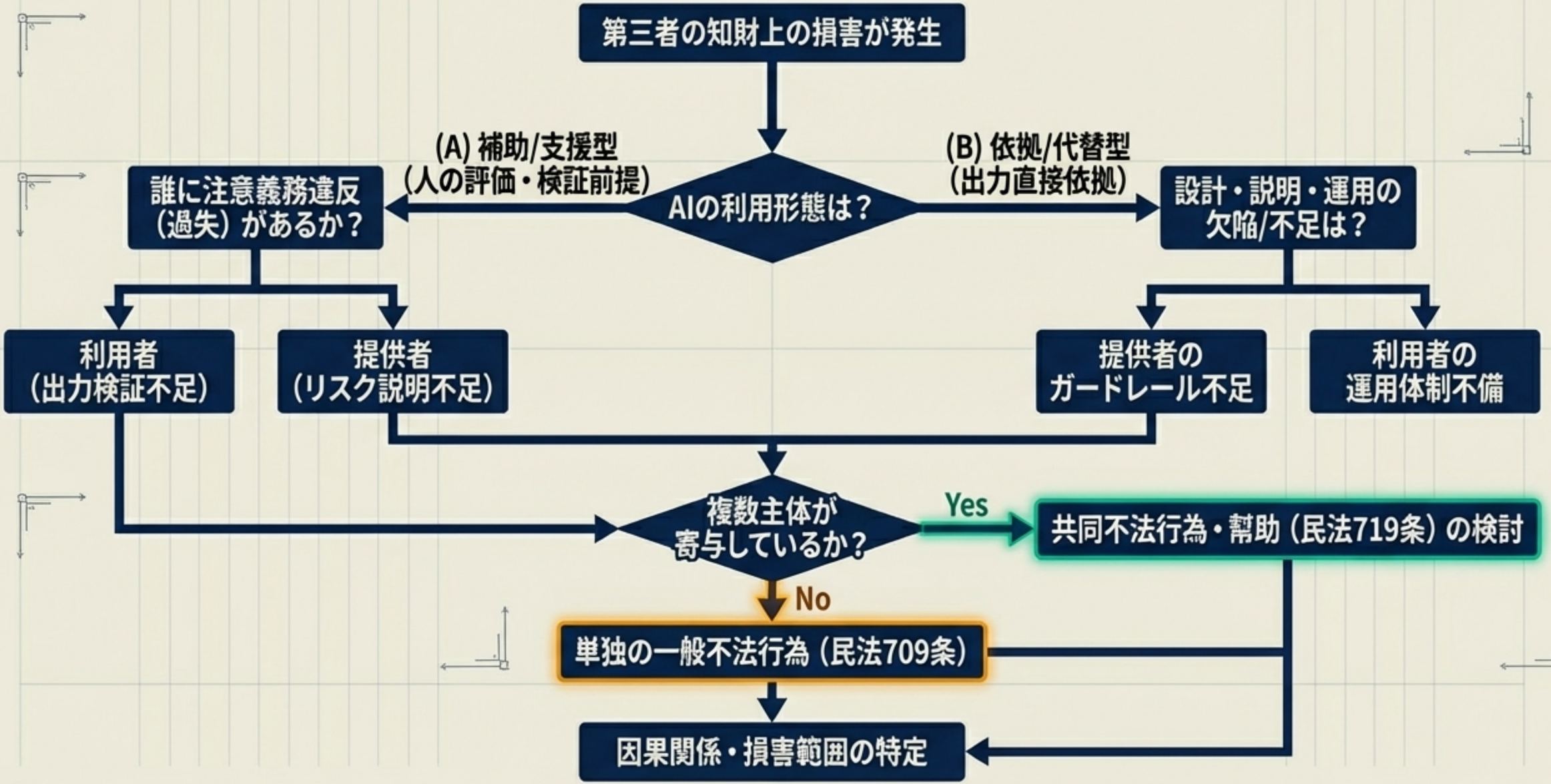
事例3（スタイル模倣）：
LoRA等で特定作家を過学習させ、類似物を大量生成・販売。

事例6/7（商標・意匠）：
ネーミングAIや画像生成AIの出力を、先行調査(クリアランス)なしに製品化。

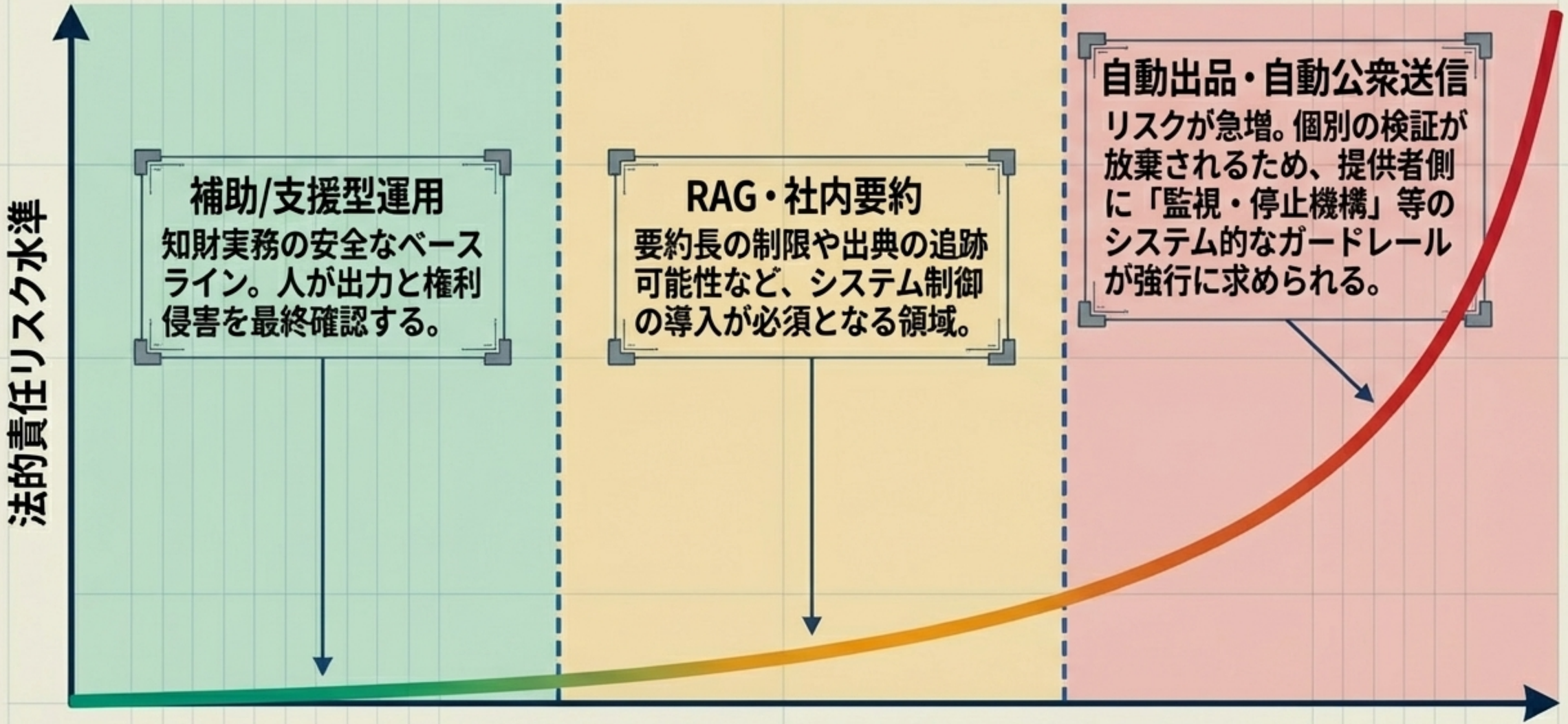
事例4（特許）：
AIが提示した「特許回避BOM」をそのまま製造。

**知財侵害リスクを内包する用途は「補助／支援型」として運用し、公開前審査（人手レビュー・専門家確認）を必須設計とする。**

# 責任評価の意思決定ツリー（実務フロー）



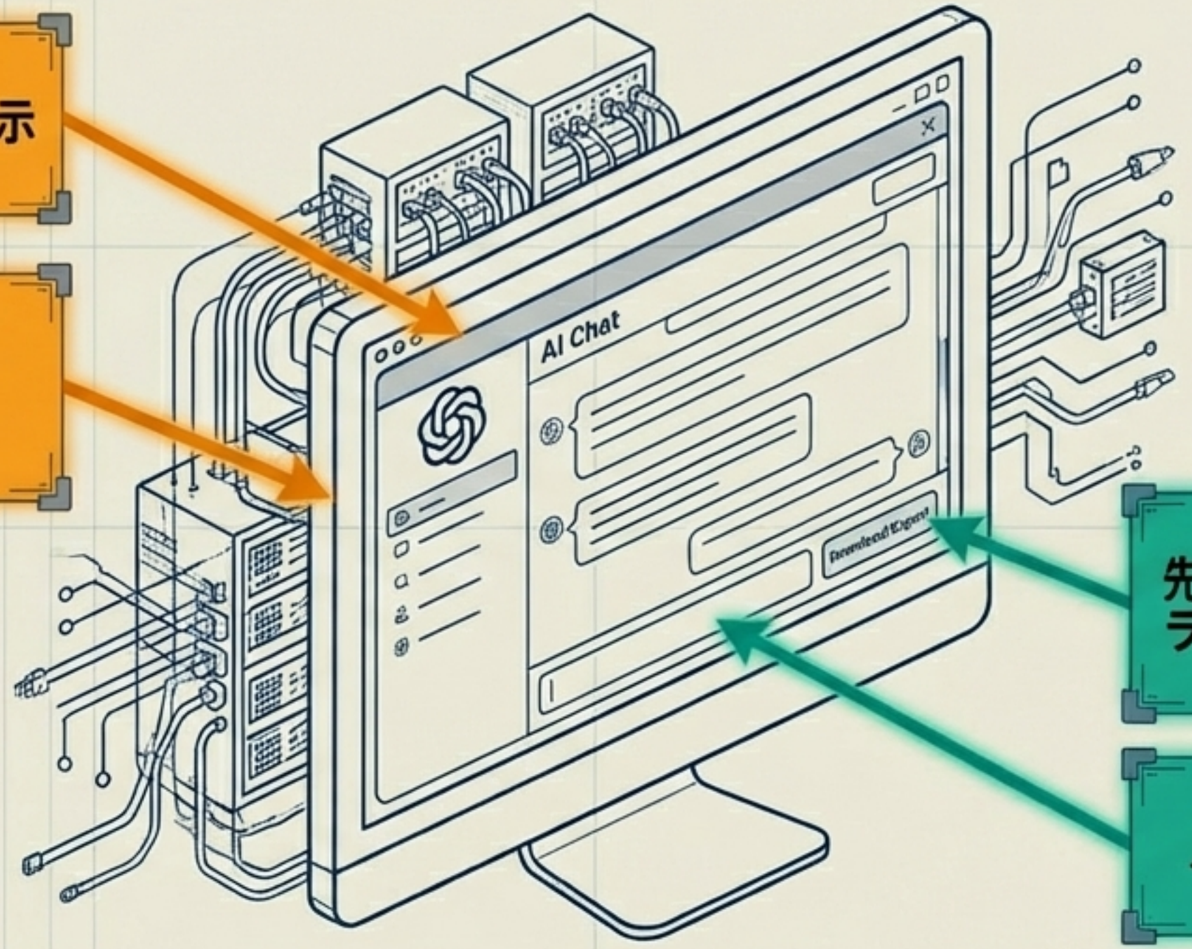
# Human-in-the-Loop (HITL) リスクカーブ



# 注意義務の解剖図：提供者と利用者の境界線

**UI Warning Banner：**
リスク説明・禁止用途の明示
（説明上の注意義務）

**Backend Filter：**
著名商標や特定作家名の
近似生成抑止機能
（設計上の結果回避義務）




**Output Export：**
先行商標・特許調査（クリアランス）の実施、最終成果物の権利侵害レビュー

**Input Field：**
入力データの適法性確認・
営業秘密の除外

**基本原則：提供者は「予見困難なリスクへのガードレール」を敷き、
利用者は「本来負うべき注意義務」を全うする。**

# デフォルトルールの“上書き”：契約責任（民法415条）の設計

手引きは不法行為（デフォルト）を中心とするが、実務では契約による「リスク配分の上書き」が最重要。

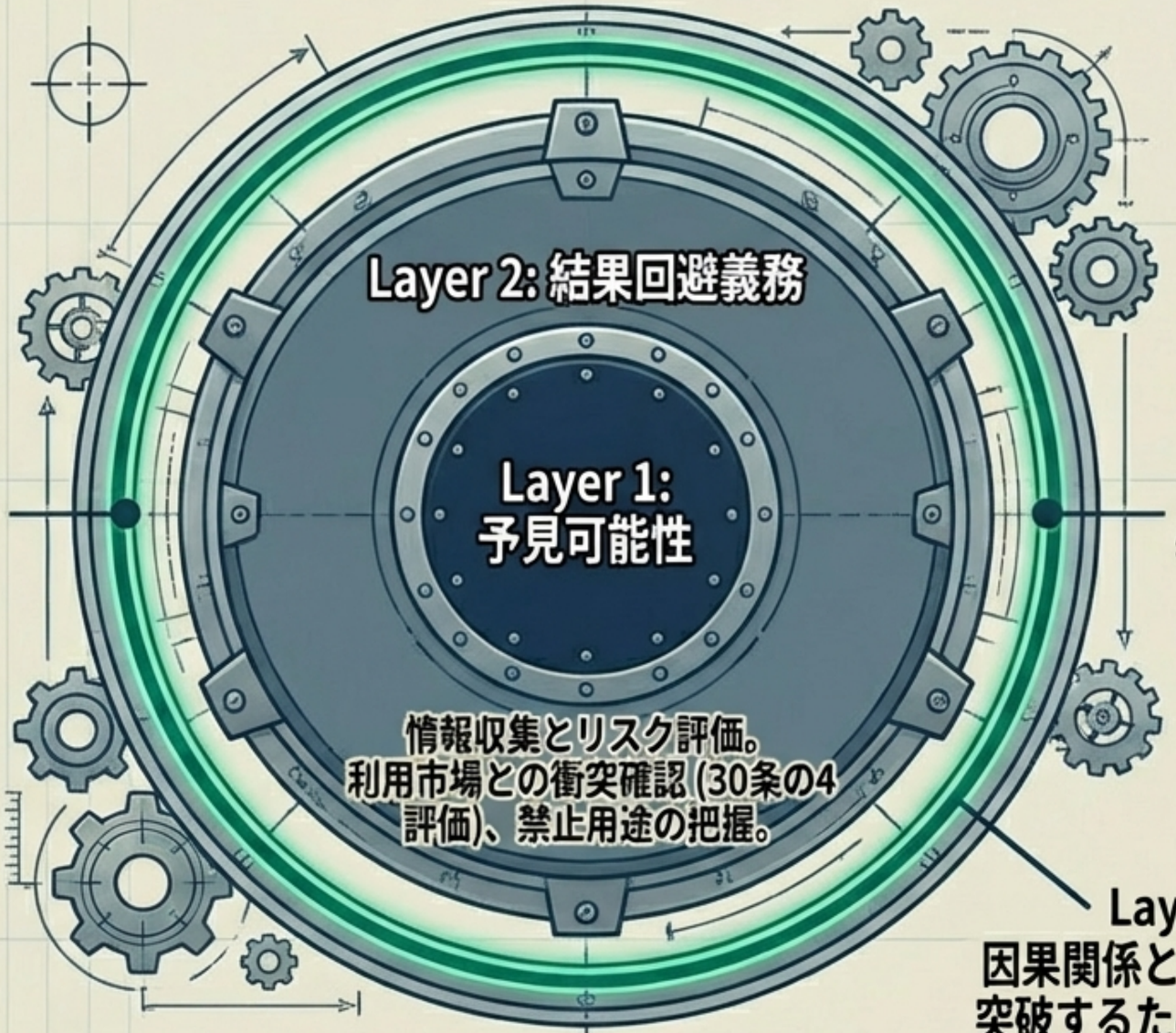
## インプット側統制 (Input Agreements)

- ☑ 入力データの学習への二次利用禁止（オプトアウト）
- ☑ ログ保持期間の制限・削除要求権
- ☑ 限定提供データ・個人情報の除外合意

## アウトプット側統制 (Output Agreements)

- ☑ 第三者権利（知財等）の非侵害保証（または非保証）の明記
- ☑ 第三者請求時の協力義務（ログ提供等）
- ☑ 補償条項（Indemnity）のスコープ設計

# 統合防衛アーキテクチャ：3層の法的シールド



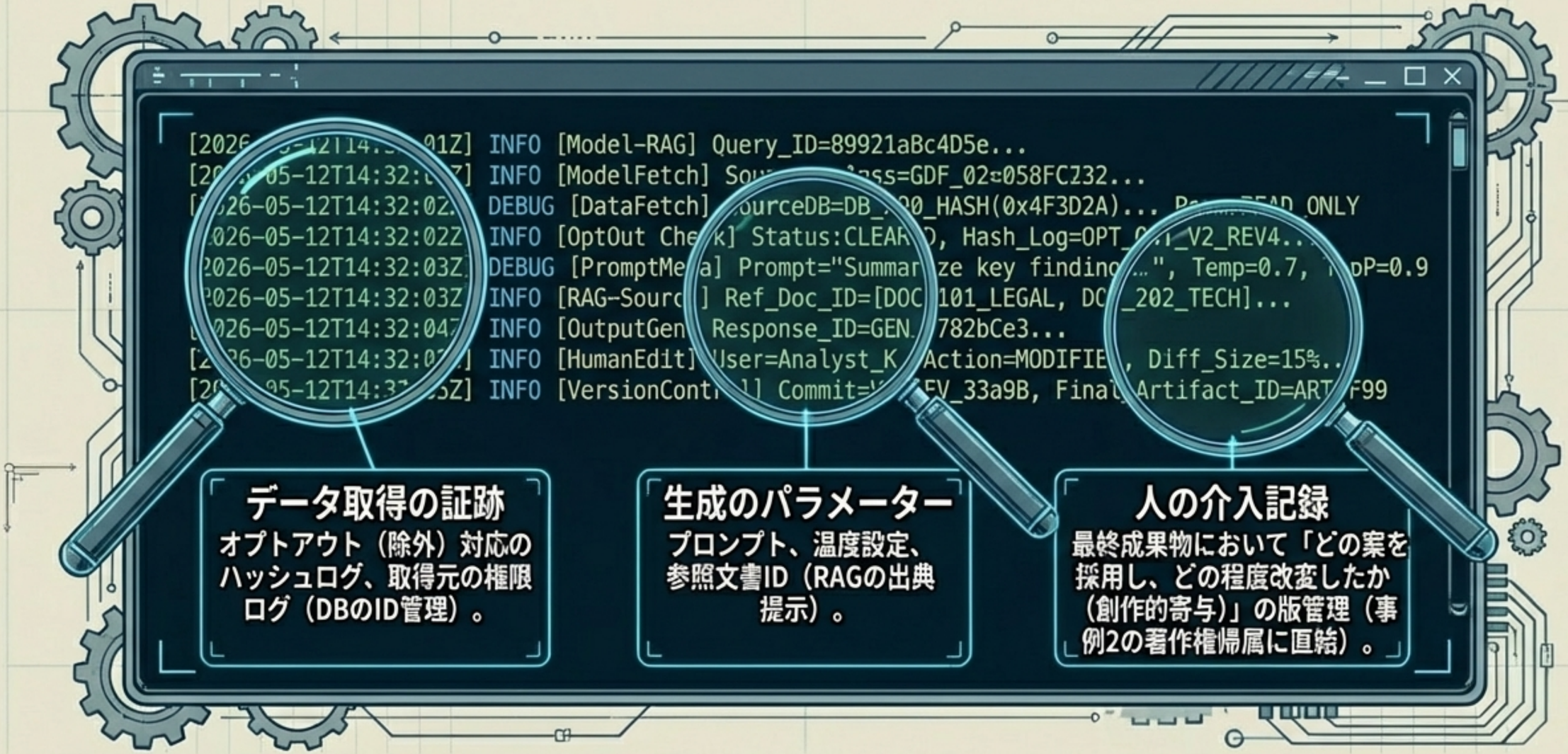

技術的ガードレール:
RAGの引用範囲制御、著名商標フィルタ、DLP導入。

運用的ガードレール:
Human-in-the-loopの必須化、FTO/クリアランス手順の標準化。

Layer 3: 監査可能性
因果関係と過失の「立証困難性」を突破するためのログ保存と版管理。

# 究極の防御としての「監査可能性（Traceability）」

## AI紛争では「どのデータを学習したか」「なぜその出力になったか」の立証負担が極めて重い。



# 実装への青写真（The Implementation Blueprint）

## 1. 運用ルール (Operational Rules)

- ☑ 知財侵害リスクを伴う業務は原則「補助/支援型」に限定。
- ☑ 自動化（依拠/代替型）する場合は体制構築・停止機構を必須化。
- ☑ 公開前審査・商標/意匠調査プロセスの組み込み。

## 2. 技術・契約 (Tech & Contracts)

- ☑ 入力データの学習利用禁止（契約での明文化）。
- ☑ RAGでの要約長制限・実質的複製の抑止設計。
- ☑ 補償スキーム（Indemnity）と保険の検討。

## 3. ガバナンス (Governance & Proof)

- ☑ 「最低限の水準＋α」の営業秘密管理（アクセス制御）。
- ☑ 生成履歴・選択理由・編集履歴の完全な証拠化。
- ☑ AI専用の利用規約・NDAの定期監査。