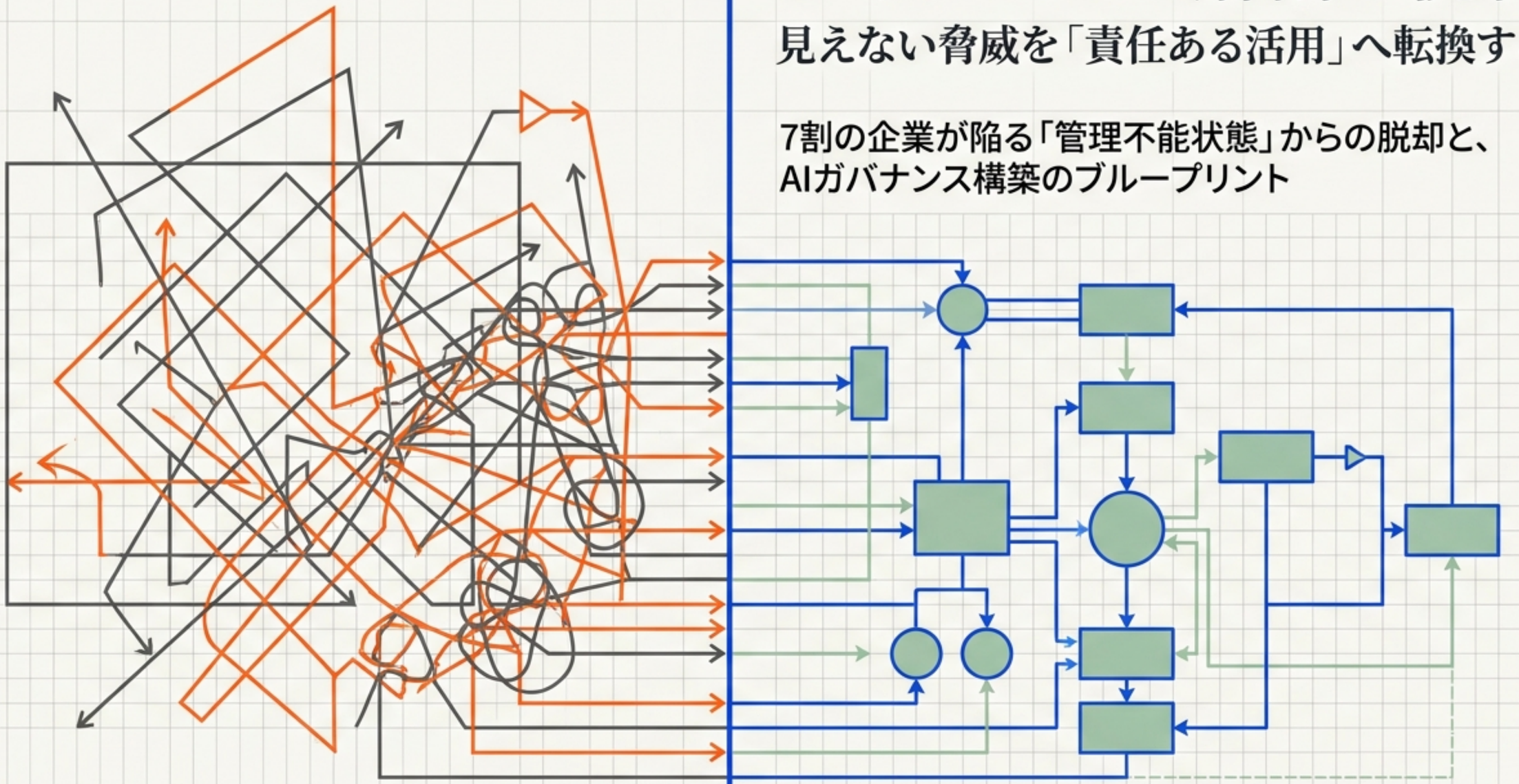
シャドーAIの解剖と統制

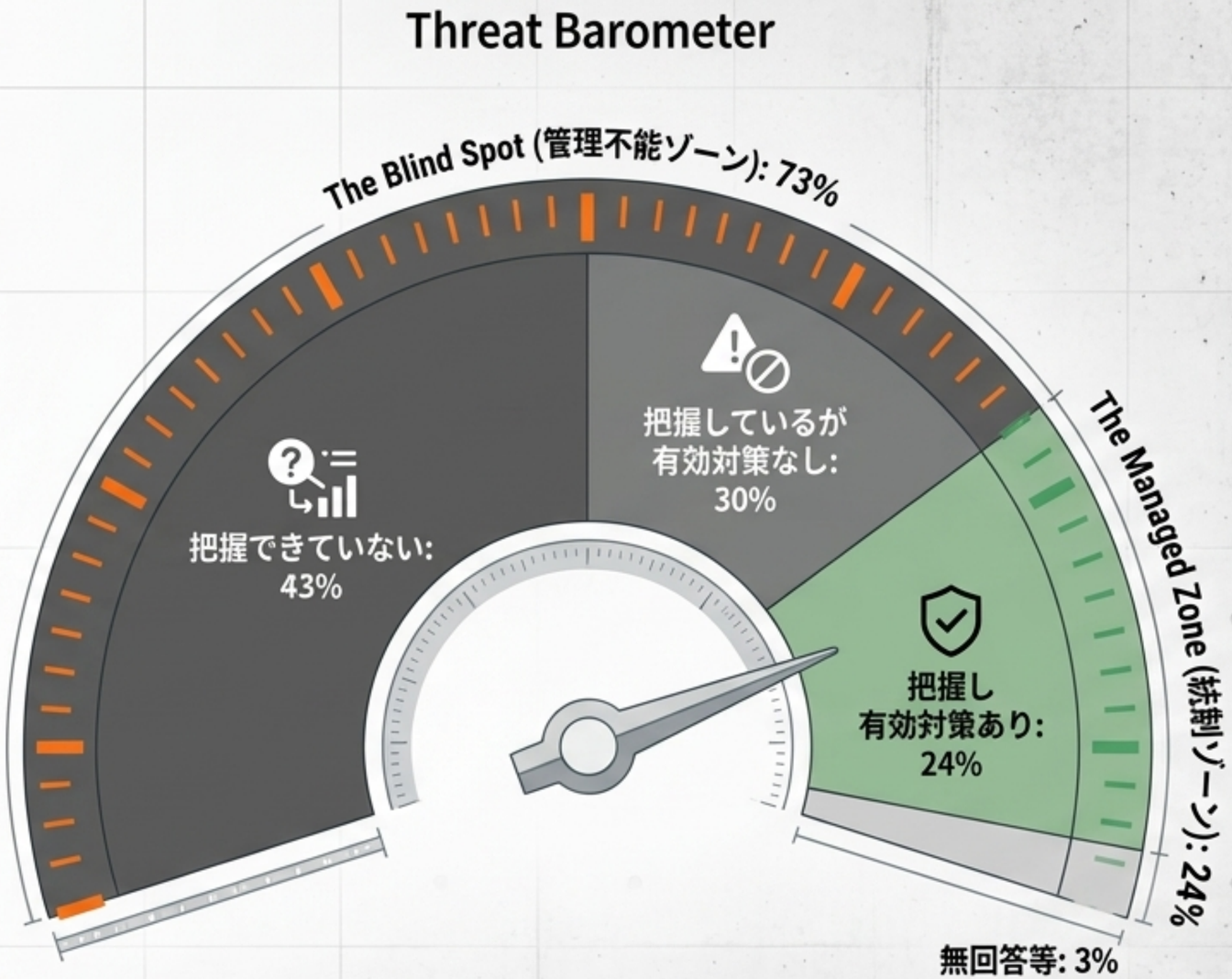
見えない脅威を「責任ある活用」へ転換する

7割の企業が陥る「管理不能状態」からの脱却と、
AIガバナンス構築のブループリント



国内企業の73%が、自社内のAI利用をコントロールできていない

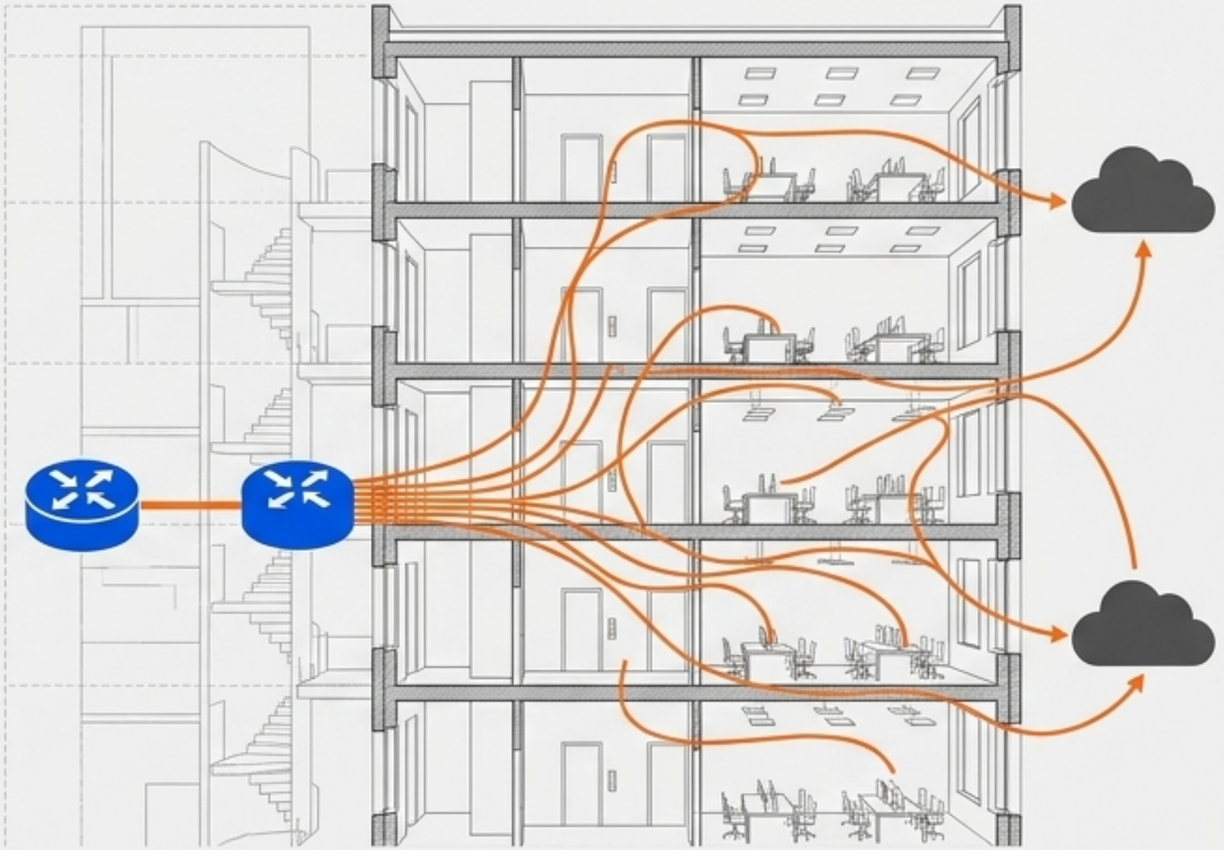
Gartner Japanの調査が示す現実：多くの企業が『部門主導の未承認AI利用』を黙認、あるいは検知すらできていない。あなたの組織は、**この巨大なブラインドスポット（73%）に属していないか？**



Architectural Audit

幻想（The Illusion）

- ルールによる一律禁止
- 公式には誰も使っていないという建前



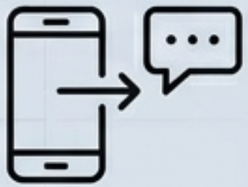
現実（The Reality）

- 世界の知識労働者の78%が、自前のAI（BYOAI）を職場に持ち込んでいる（Microsoft調査）

現場実態として一律禁止は守られない。禁止政策はAI利用を止めるのではなく、利用を『地下化（Shadow化）』させるだけである。

Architectural Audit

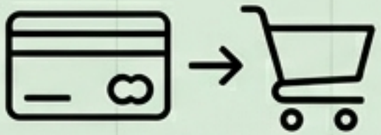
シャドーAIの正体：ブラウザの向こう側へ拡張する「未知の資産」



未承認の一般公開チャットボット
（個人端末からの持ち込み）

SaaSのAI機能の無断有効化
（既存ツールに潜むAI）

Shadow AI



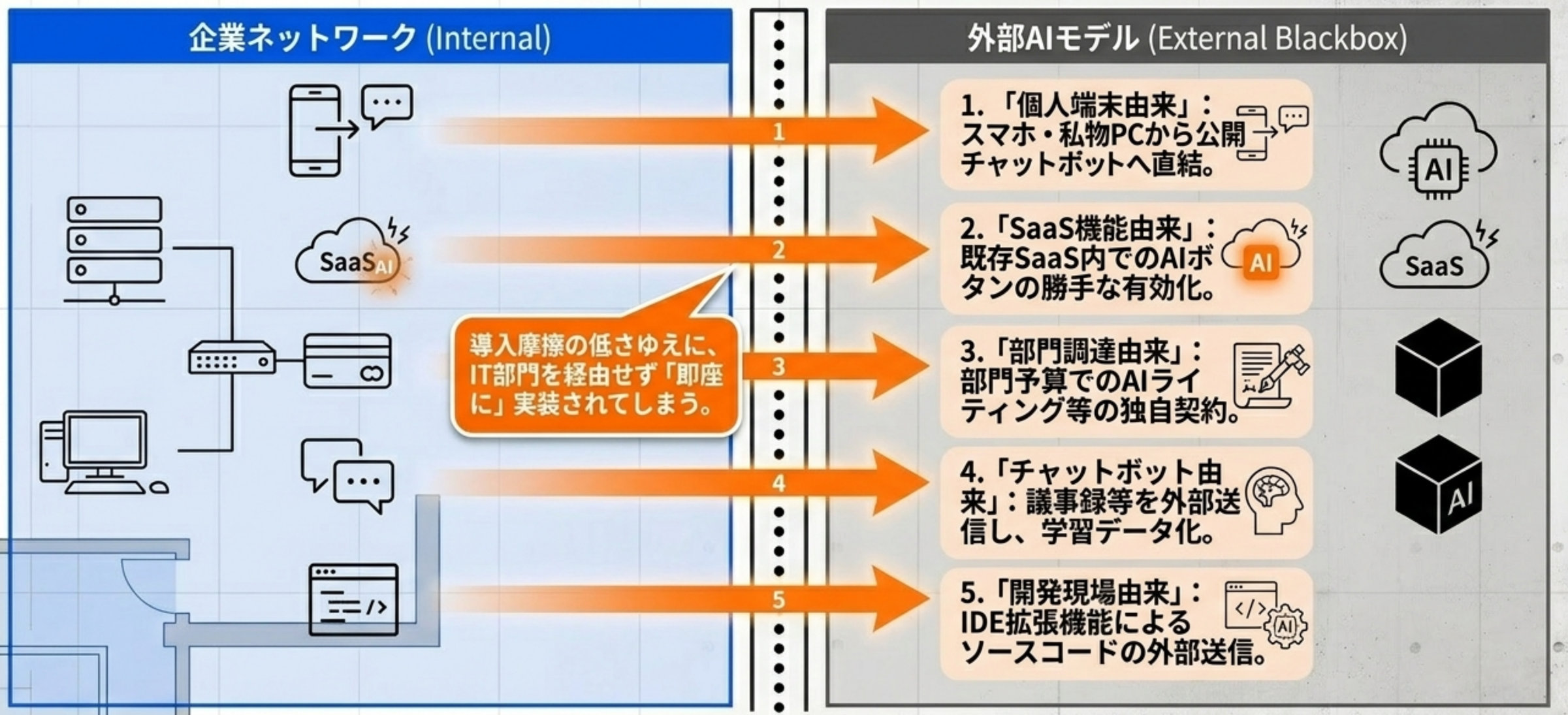
部門単位の自己調達
（クレジットカード決済による野良SaaS）

無許可のAPI連携
（社内ツールから外部モデルへの接続）

**組織の正式承認・監督・リスク評価・契約審査・監査設計の『外』で使われる
すべてのAIツール・機能・連携網。（NCSC / IBM定義より統合）**

Architectural Audit

境界をすり抜ける5つの侵入経路（Invasion Routes）

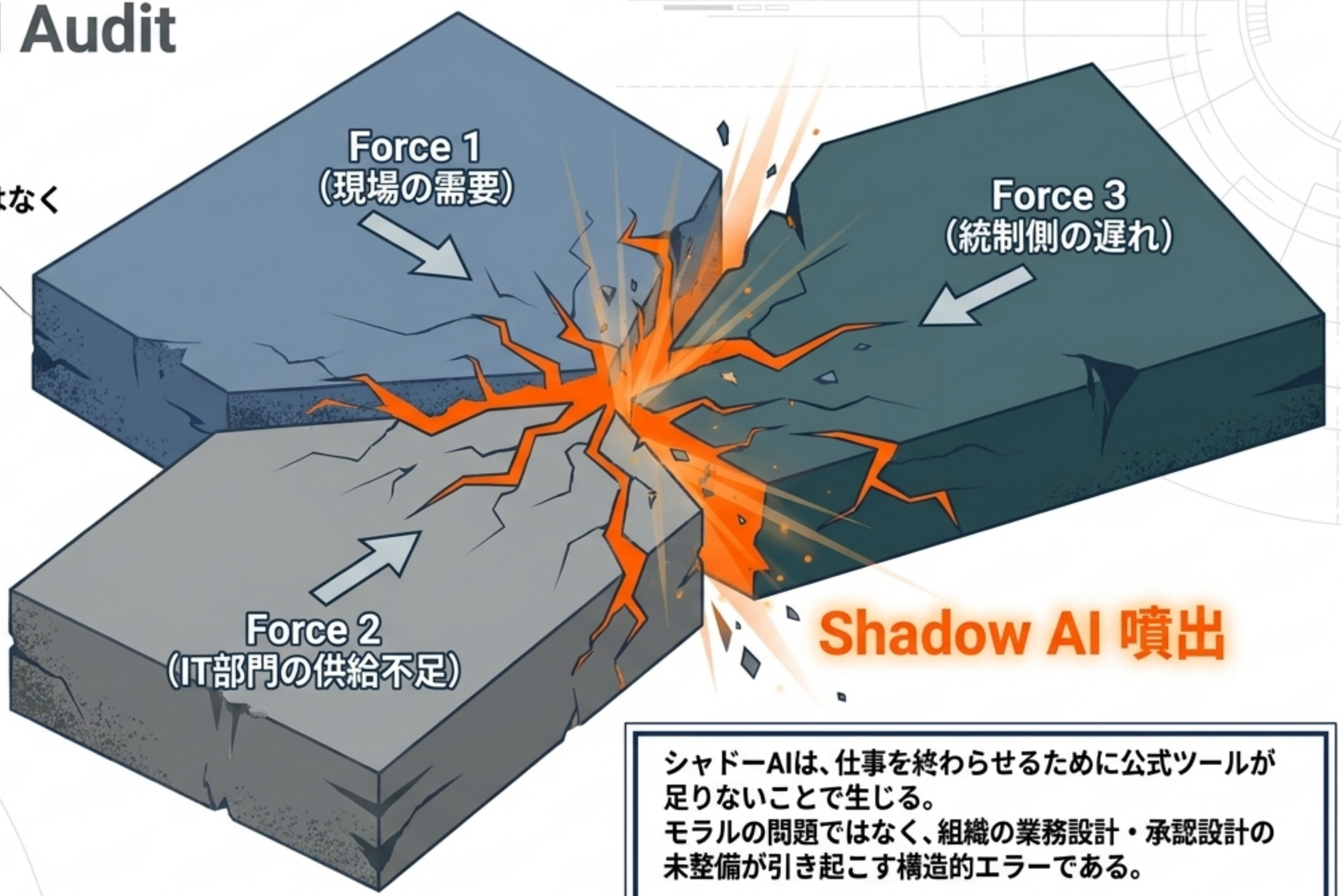


Architectural Audit

Headline

発生のメカニズム：悪意ではなく「需要と供給のギャップ」

- Plate 1 Detail
  スピードと生産性の追求 / 公認ツールの機能不足
- Plate 2 Detail
  承認プロセスの遅れ / 代替手段（公式AI）の不在
- Plate 3 Detail
  法務・セキュリティによるルール未整備 / 過剰な制約



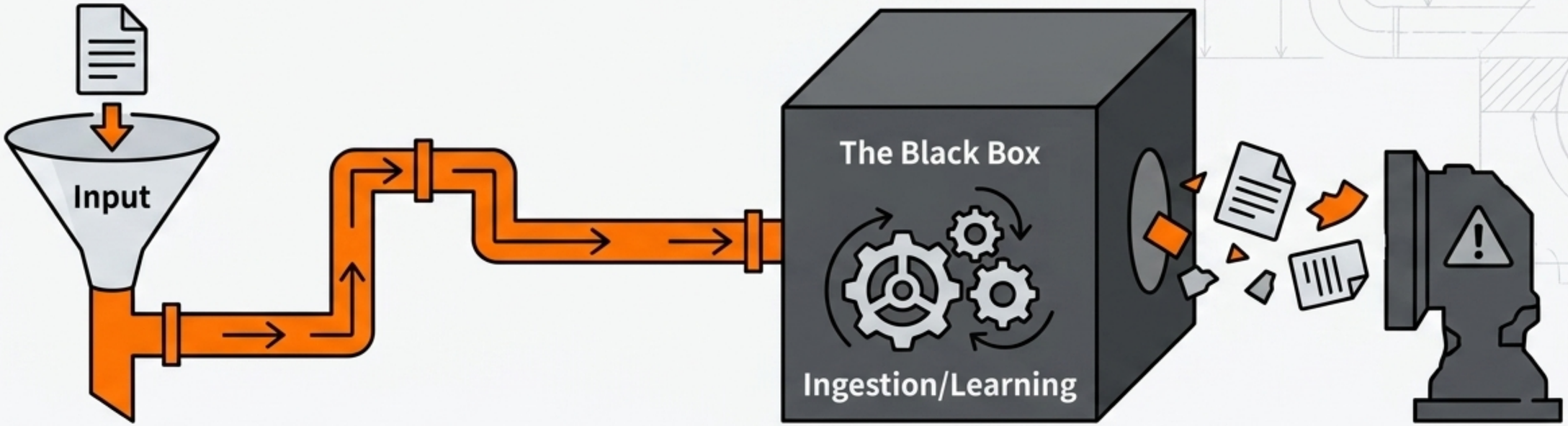
シャドーAIは、仕事を終わらせるために公式ツールが足りないことで生じる。
モラルの問題ではなく、組織の業務設計・承認設計の未整備が引き起こす構造的エラーである。

シャドーAI リスク・プロファイル（脅威診断表）

[リスク分類]	[影響度]	[発生シナリオ]	[国内外の事例・公開例]
情報漏洩	High	設計書・顧客情報のAI投入	Samsung（機密コード流出）
法令違反	High	個人データの目的外外部送信	PPC注意喚起、 イタリアGarante制裁
知財・機密漏洩	High	営業秘密・ノウハウの流出	文化庁（AIと著作権 チェックリスト）
サイバーセキュリティ	High	プロンプトインジェクション	OWASP上位リスク、 EchoLeak研究
業務品質・監査	Med	幻覚の流用、証跡不備	Mata v. Avianca (架空判例提出による制裁)

Architectural Audit

不可逆なデータ流出：ブラックボックス化のメカニズム



ユーザーが便利なAIに「社外秘の議事録」や「ソースコード」をペーストする。

データが外部事業者のサーバーへ渡り、ブラックボックスに入る。

AIの学習データとして吸収され、利用規約に基づき権利が移動・保持される。

最悪のシナリオ：他社のプロンプトに対する「回答」として自社の機密が出力される。

便利な外部検索箱ではない。外部事業者・外部保存先・外部規約を伴う「情報処理基盤」への無防備なデータ移転である。

Architectural Audit

パラダイムシフト：「禁止」から「可視化・承認・統制」の分業モデルへ

過去のアプローチ

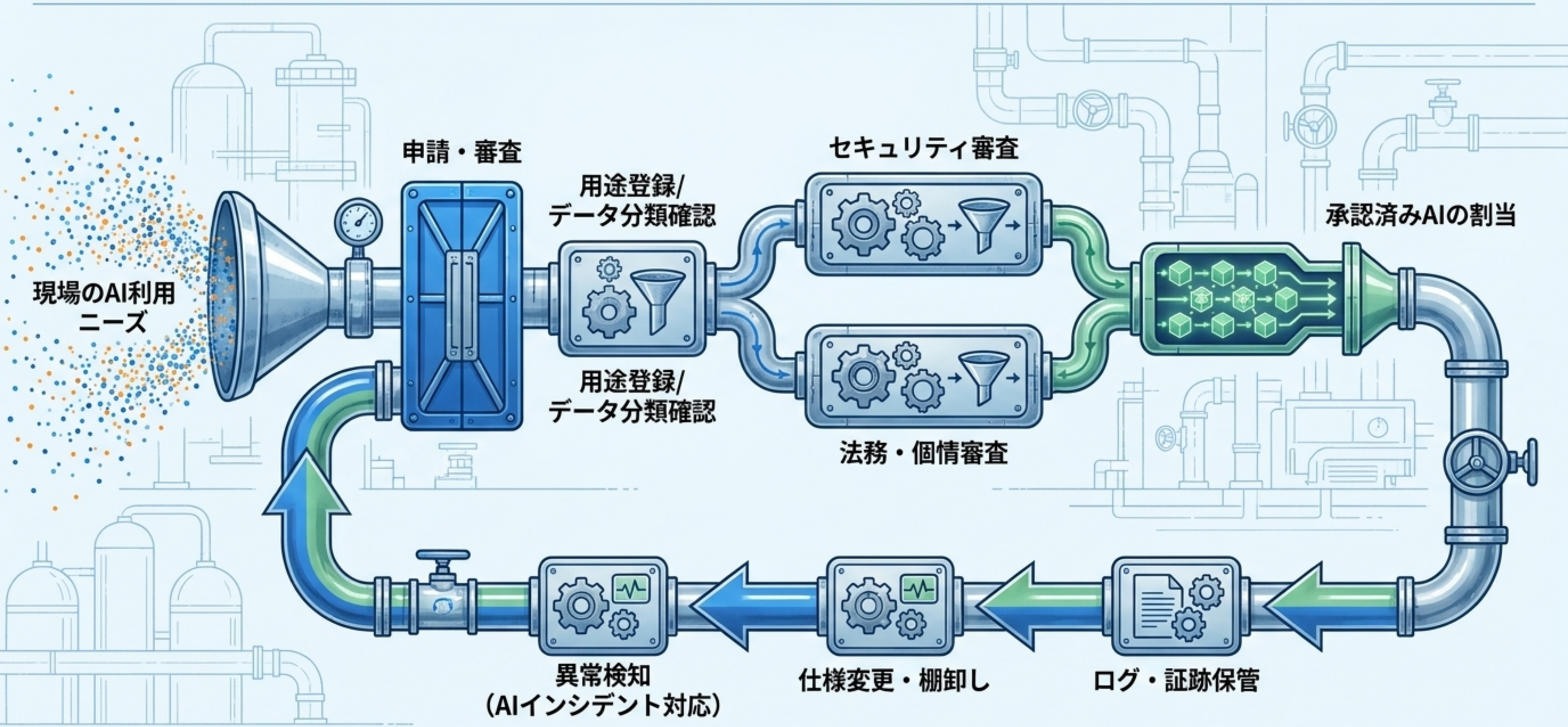
「禁止による抑圧」→
「利用の地下化（シャドー化）」の悪循環

新時代のアプローチ

「代替手段（公式AI）の提供」
→「利用の可視化と統制」の好循環

**「シャドーAIを防ぐ最善の手立ては、見つけ出して罰することではない。
『安全で使い勝手の良い公式AI（承認済み代替手段）』を先に提供することである。」**

承認・監視の全体ワークフロー（The Governance Pipeline）



NIST AI RMF, CSF, AISIが推奨する「Govern / Map / Measure / Manage」を統合した継続的ループ。

Architectural Audit
対策手法のROI評価カタログ [技術統制編]

検出・可視化 **(CASB/SaaS管理)**	 コスト: 中	 「何が使われているか」を特定。禁止より先の必須基盤。	 個人端末の通信は取りこぼすリスク。
データ保護 **(DLP/ブラウザ制御)**	 コスト: 中	 高機密データ（コピペ等）の送信を直接抑止。	 誤検知の調整が必要。SaaSごとの実装差。
アクセス制御 **(SSO/Private Marketplace)**	 コスト: 中〜高	 野良AIより先に「公式AI」を提供し、調達を統制。	 現場ニーズに追いつかないと抜け道化。

Architectural Audit
対策手法のROI評価カタログ [制度・プロセス統制編]

ポリシー策定と分業	コスト: 低	教育・契約の土台。 禁止/許可/申請の明確化。	代替手段がないと 形骸化。
契約条項・法務審査	コスト: 中	学習利用の有無、データ所在 の確認。「規約未読」を防ぐ。	無料・個人利用ツール には効きにくい。
監査・インシデント対応	コスト: 中〜高	AI-IRSに基づく異常検知・ 封じ込め。監査対応。	コンシューマ向けAIでは ログ取得に限界。
従業員教育・認定	コスト: 低〜中	「判断の質」を向上。 職種別の認定で機能開放。	継続的な運用が必要。

Architectural Audit

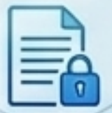
アクション・ロードマップ：短期対応（止血とトリアージ）

優先度: 高

Step 1:
利用実態の棚卸し
（情シス/CISO）
現在のAIアプリとデータフローの把握。

優先度: 高

Step 2:
「入力禁止データ」の定義
（法務/事業部）
個人データ、ソースコード等の禁則明文化。



優先度: 高

Step 3:
暫定ポリシーの策定
（経営/CISO）
全面禁止ではなく、許容・禁止・申請制の区分け。

優先度: 高

Step 4:
承認済みAIの一次リスト公開
（情シス/調達）
現場が「すぐ使える公式AI」の提供。

最優先事項は「禁止すること」ではなく「見える化し、安全な代替路を敷くこと」である。

Architectural Audit

アクション・ロードマップ：中長期対応（ガバナンス構築）

Mid-Term 1: 制御の導入

CASBによる新規AI検出、ブラウザDLPによる送信抑止。

Mid-Term 2: 審査の標準化

契約審査テンプレートの整備と申請フロー簡素化。

Long-Term 1: 統合的運用

NIST AI RMF/CSFおよびAI-IRSの統合運用。

Long-Term 2: 委員会の常設と定期棚卸し

AIガバナンス委員会の設置、四半期ごとの再承認プロセス。

業種固有の規制（金融・医療等）がある場合は、ログ管理と説明責任の強度を一段引き上げる。

シャドーAIの本質は「利用を止めること」ではない

・シャドーAIは、AI活用が本格化した組織で必ず生じる『運用課題』である。
・統制なき導入促進は放任に堕し、代替手段なき全面禁止は地下化を招く。
・いま求められるのは、見えない脅威を恐れることではなく、組織横断のガバナンスという光を当て、AIを『責任ある活用』へと導くことである。