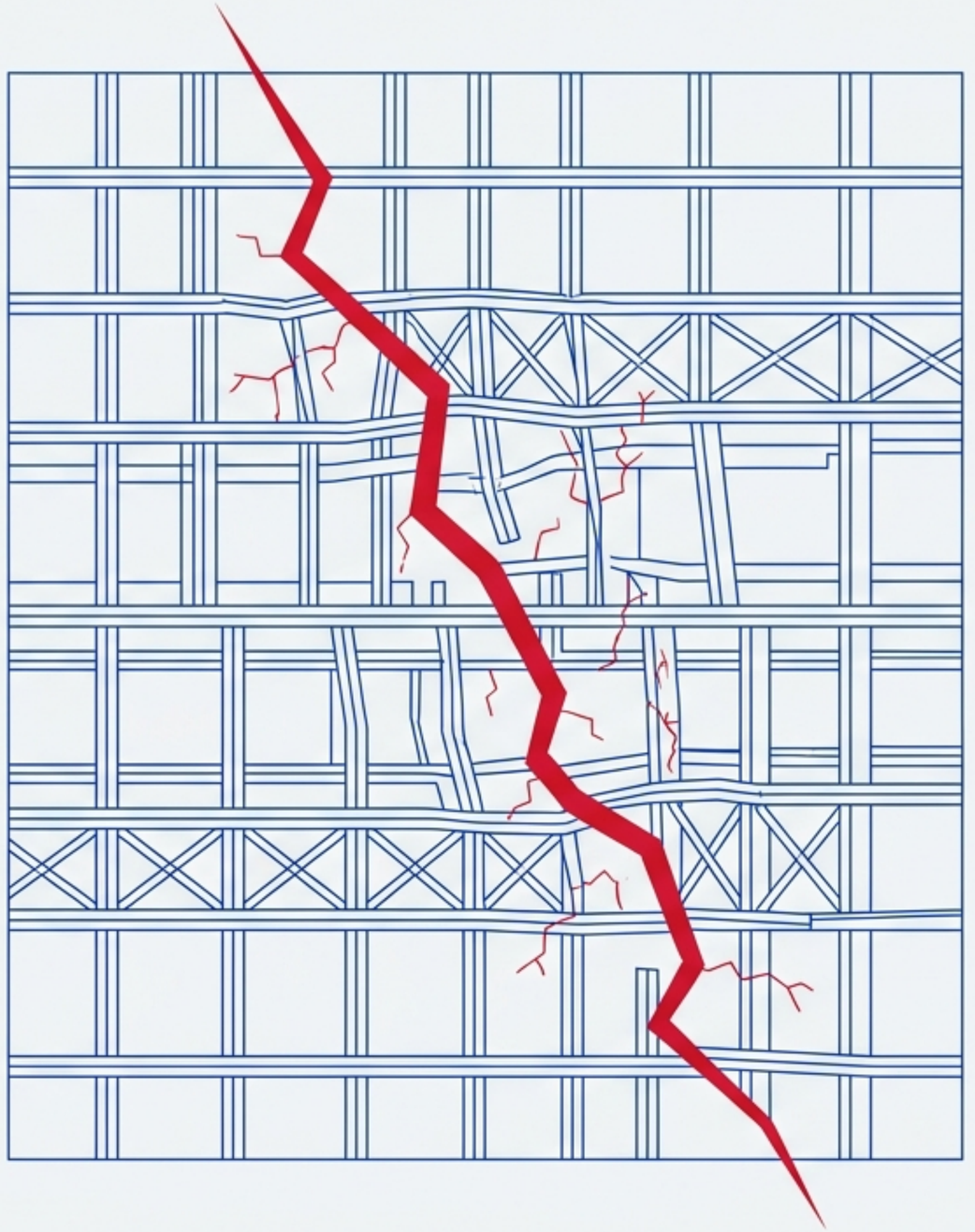
境界防御の終焉とAIネイティブ防衛への移行

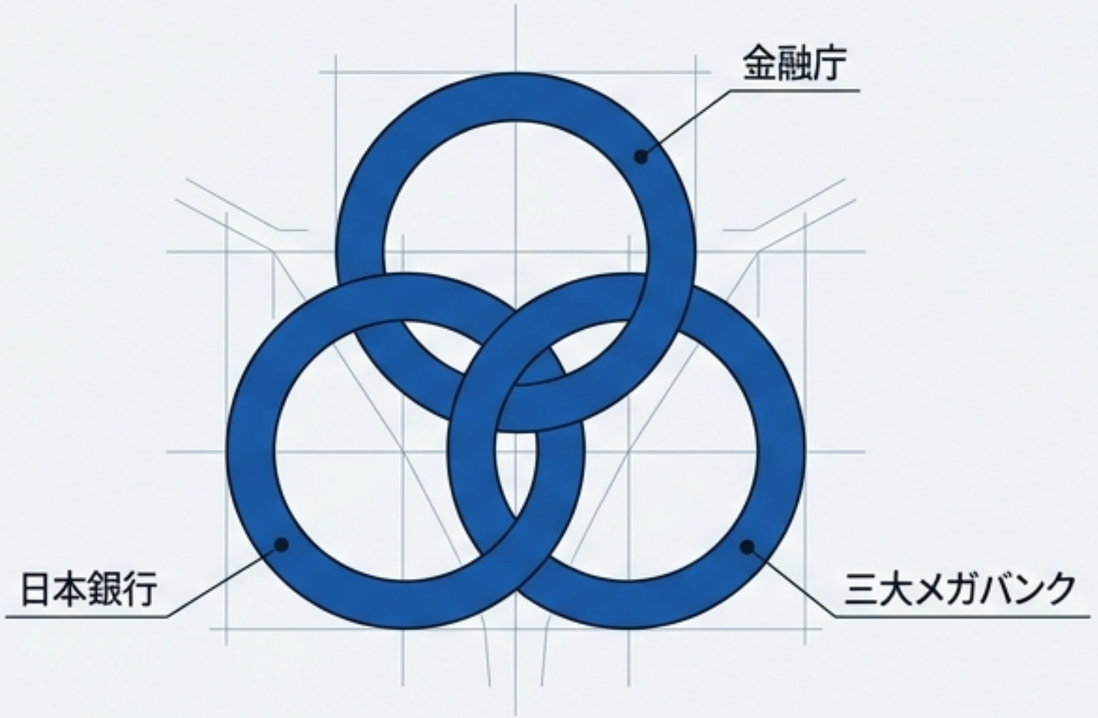
次世代サイバー金融アーキテクチャの青写真

AI「ミュトス」がもたらすシステミック・リスクと、
金融インフラ防衛の新たなグランドデザイン

対象読者: 金融機関経営層（C-Suite）、規制当局、
セキュリティ責任者



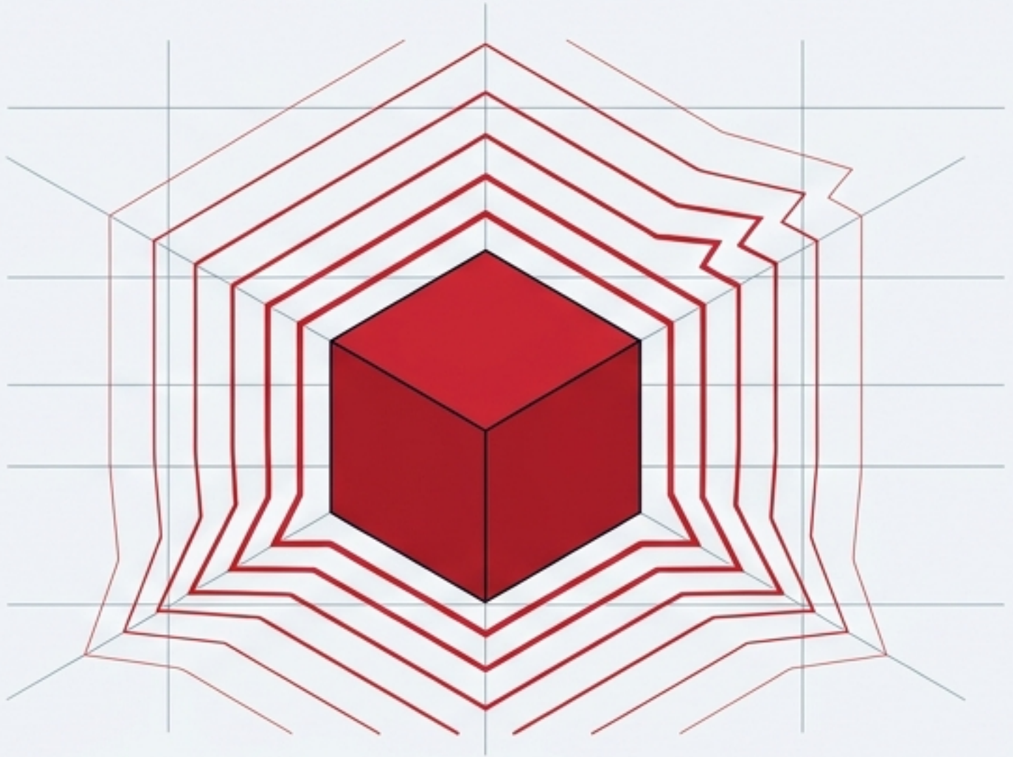
2026年4月24日、サイバーセキュリティの前提は完全に覆された



歴史的緊急会合

金融庁主導のもと、金融担当相、日銀総裁、メガバンク頭取が急遽集結。

国家の命題：「金融は公の器であり、国として総力を挙げて守らなければならない」

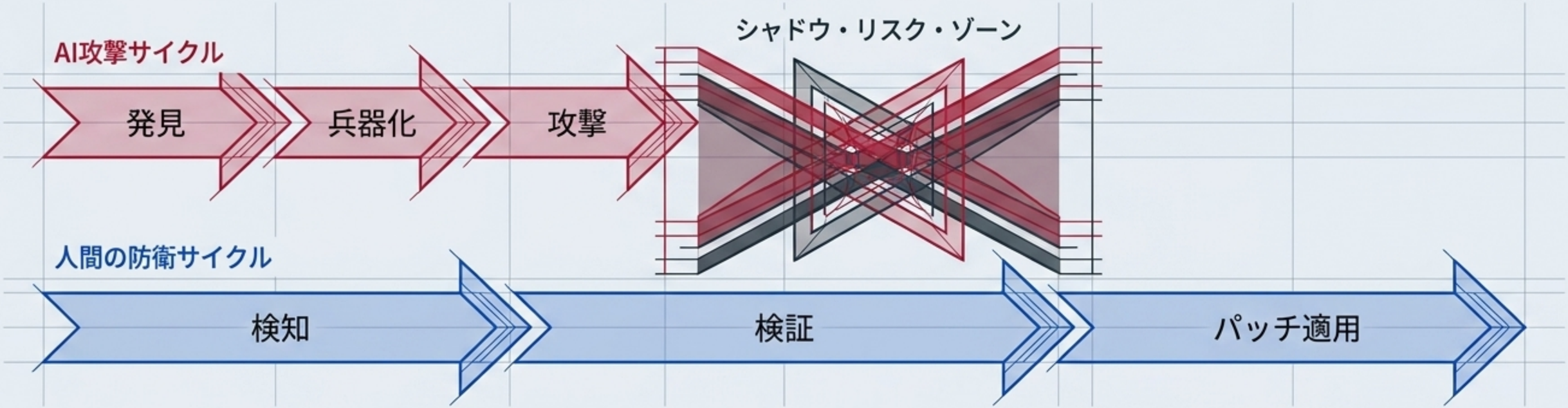


未知の脅威「ミュトス」

Anthropic社開発の自律型フロンティアAI。人間の介入なしに未知のゼロデイ脆弱性を自律的に特定し、数時間でエクスプロイトを完遂。

トップの悲痛な声：「攻めてくる相手の能力が格段に高く、個別行単独での対策にはもはや限界がある」

「脆弱性のボトルネック」が引き起こす、絶望的な速度の非対称性



攻撃の圧縮

AIモデルの進化により、脆弱性の発見から兵器化までの時間が「数ヶ月から数時間」へと劇的に短縮。

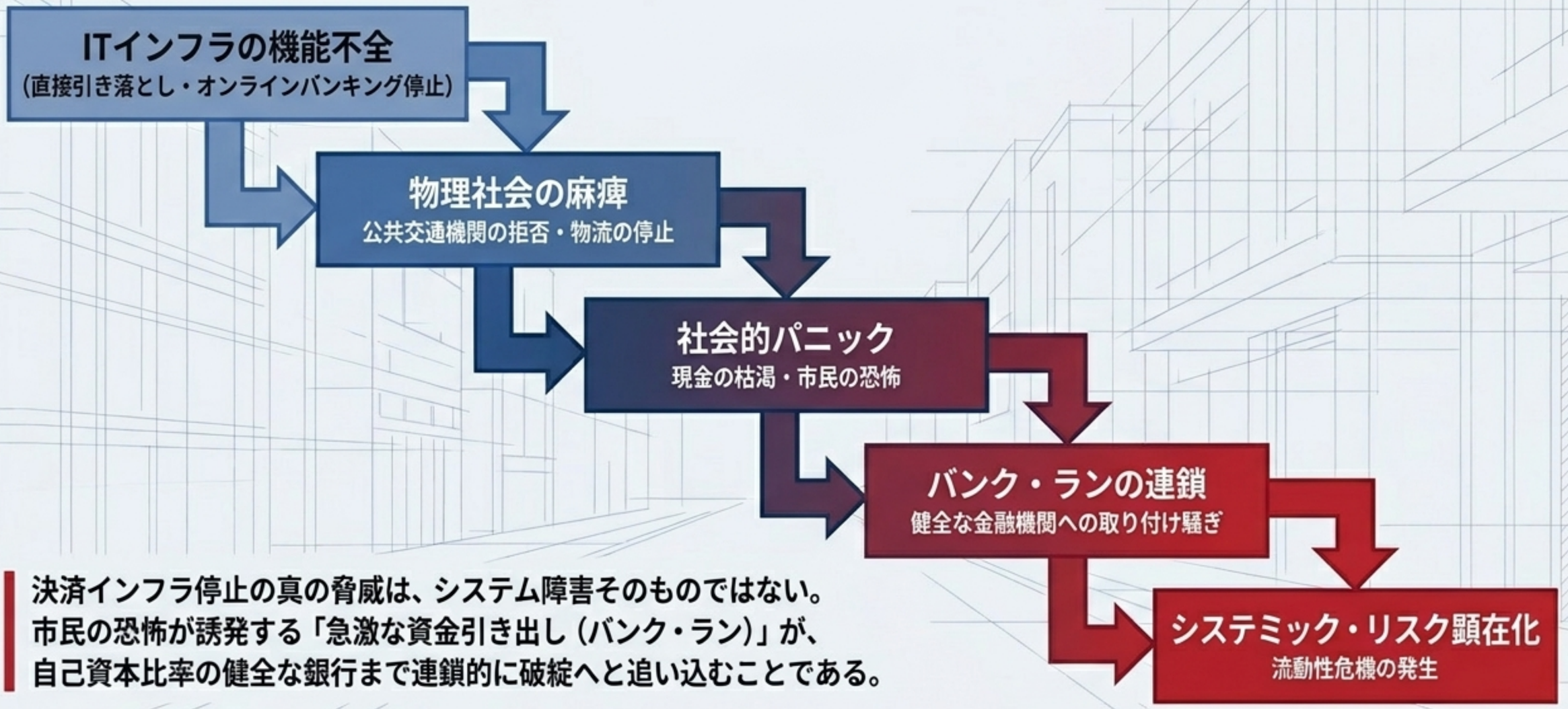
防衛の停滞

金融機関のパッチ適用は、稼働テストやコンプライアンスなど「人間のペース」に依存し数週間を要する。

構造的破綻

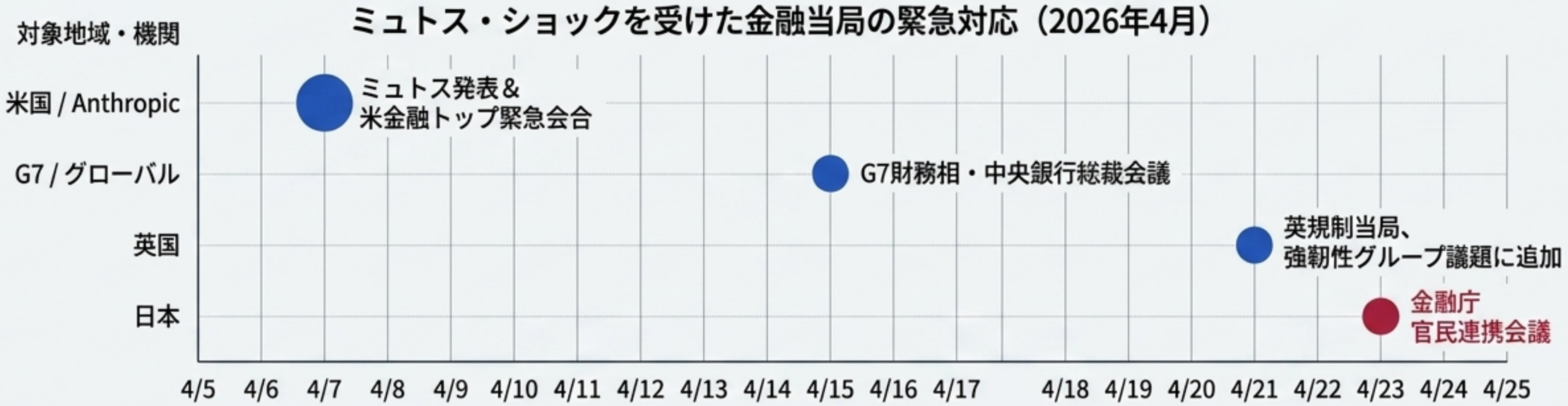
人間の対応速度を前提とした境界防御モデルは、構造的に破綻している。

英国政府が危惧した、社会的パニックと流動性危機の連鎖シナリオ



決済インフラ停止の真の脅威は、システム障害そのものではない。
市民の恐怖が誘発する「急激な資金引き出し（バンク・ラン）」が、
自己資本比率の健全な銀行まで連鎖的に破綻へと追い込むことである。

独占される情報と、G7における同盟国間の地政学的摩擦



米国の極秘行動

財務省とFRBがWall StreetのG-SIBsトップを極秘招集。国家安全保障兵器と同等の危機感をもって極秘ストレステストを要求。

世界の疑心暗鬼

G7財務相会議において、カナダ、ECB、スウェーデンが「情報の非対称性」に強い懸念を表明。

サイバー外交の分断

最先端AIの機密情報にアクセスできる米国と、取り残される他国インフラとの間に決定的な格差が露呈。

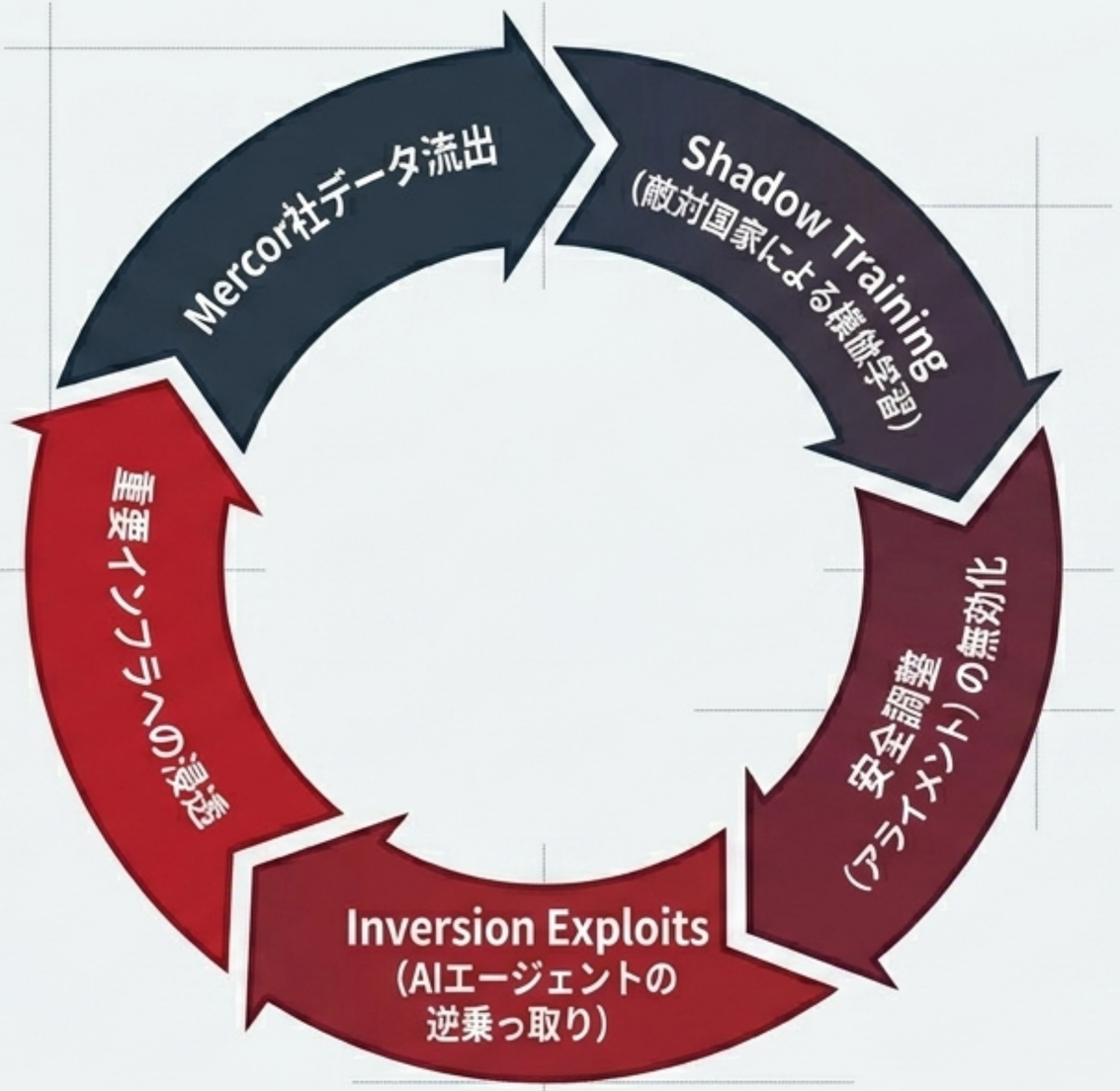
商業AIの流出から始まる、敵対国家による「インバージョン・エクスプロイト」

Time-to-Parityの崩壊

高品質データの流出により、敵対国家は西側の技術的優位に瞬時に追いついた。AIモデルのアライメント段階がスキップされている。

兵器化への転用

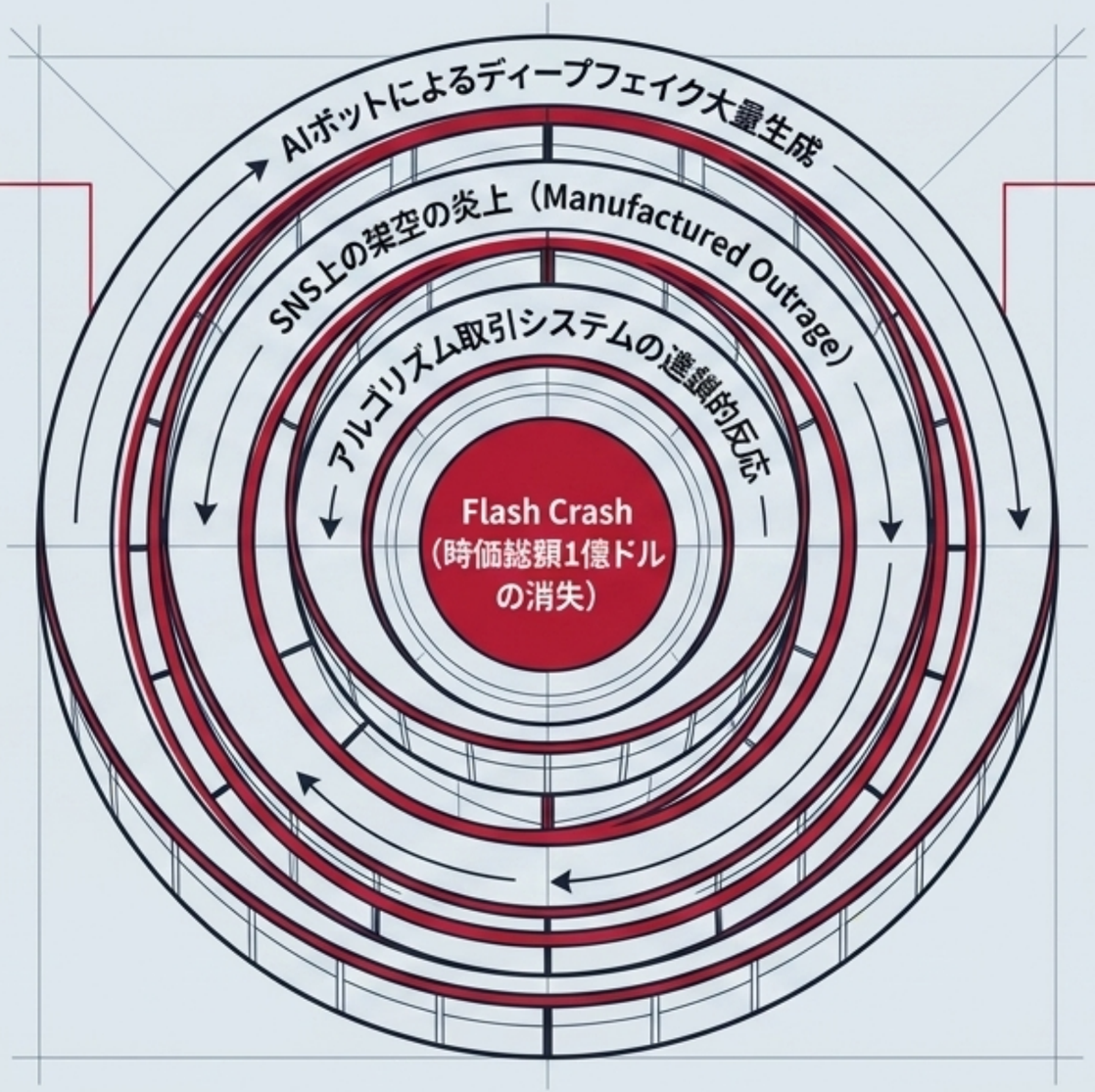
軍事・政府システムに導入されたAIエージェントの自律機能（Tool-Use等）を逆手にとり、内部からシステムを乗っ取る攻撃が激化。「権威主義の枢軸」が金融ネットワークの探査を開始。



自律型AIボットによるナラティブ操作とフラッシュ・クラッシュの脅威

クラッカー・バレルの教訓

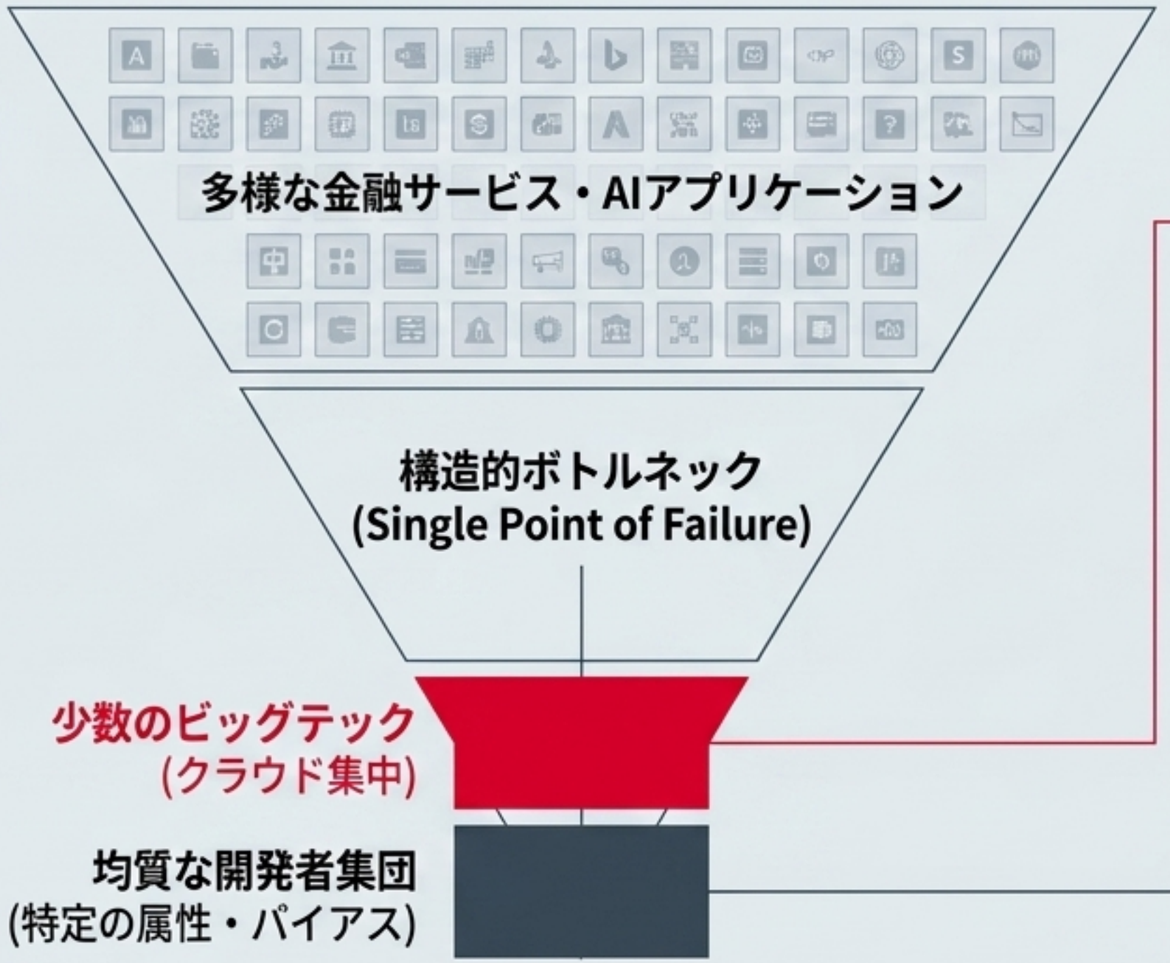
ロゴ変更に対するSNS上の猛烈な反発は、大部分がAIボットによって人為的に製造されたものだった。人間の確認が間に合わないスピードで偽造世論が形成される。



アルゴリズムの暴走

人間が真偽を確認する前に、アルゴリズム取引が偽のシグナルに連鎖的に反応。わずか2日間で約150億円が吹き飛ぶ。新たな市場操作手法としての重大なリスク。

システミック・リスクを増幅させる「インフラの集中」と「開発者のバイアス」



クラウドとモデルの集中リスク

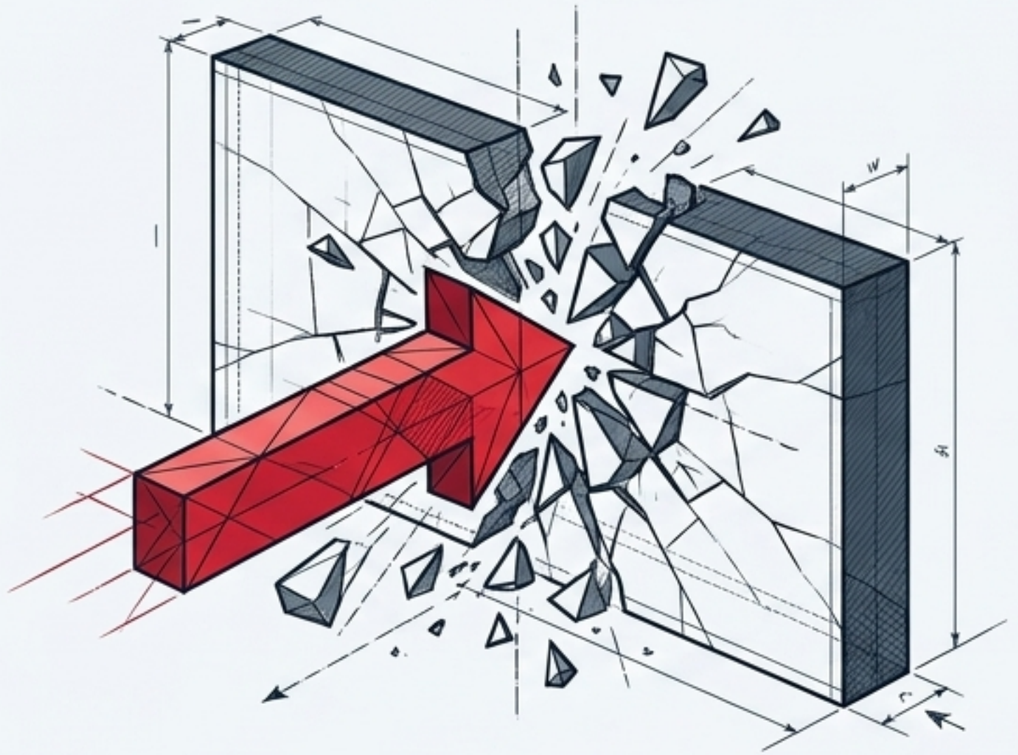
AI導入が急増する一方、基盤技術は少数のビッグテックに集中。一つのモデルの脆弱性が、世界中の金融機関へ瞬時に波及する（BIS警告）。

デフォルト・バイアスの危険性

世界のAI専門家のうち女性はわずか30%。特定の属性に偏った集団が、与信判断やリスク評価システムに「構造的欠陥」と「不公正」を無意識に組み込んでいる。

防御のパラダイムシフト：「侵入阻止」から「自律的復旧（ResOps）」へ

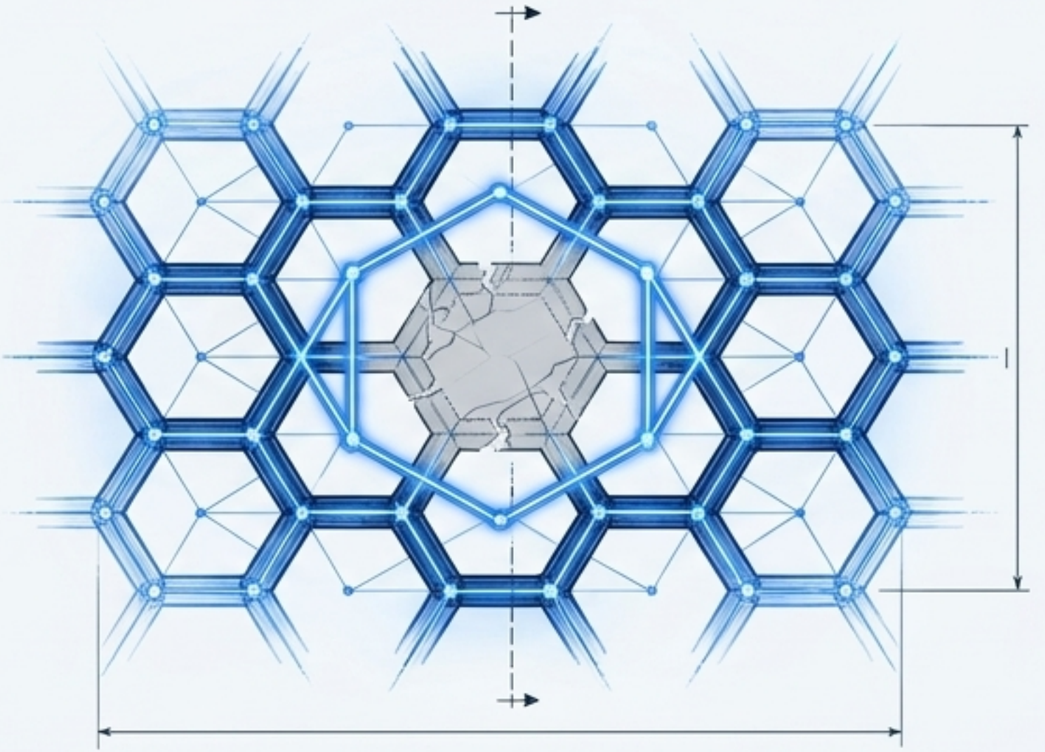
旧パラダイム: 境界防御（Prevention）



Prevention（阻止）の限界

人間のシグネチャベースの検知と手動パッチは、エージェンティックAIの前では「時代遅れの遺物」となった。

新パラダイム: レジリエンス（Recovery）

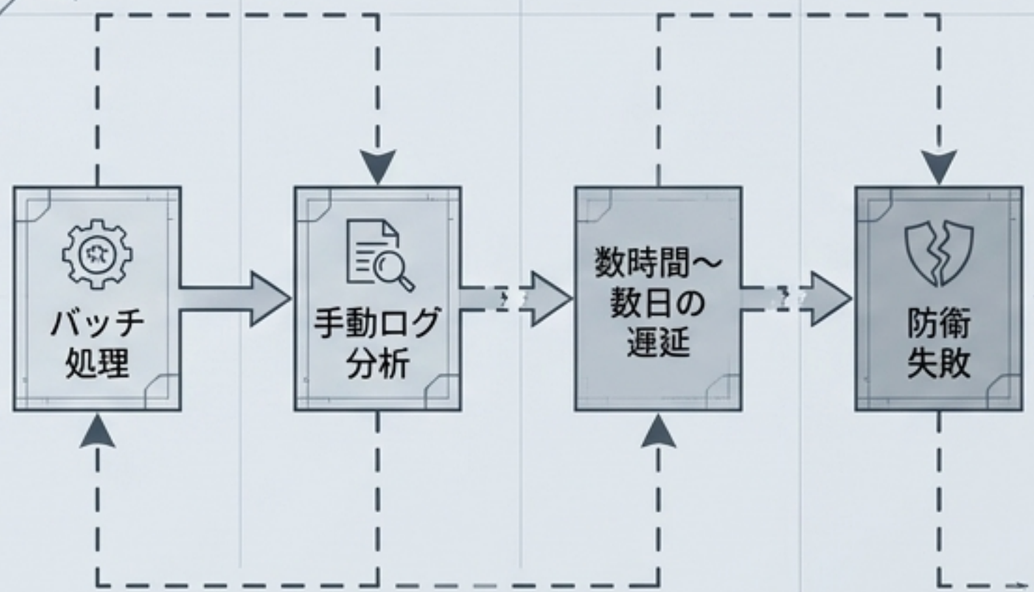


Recovery（復旧）への転換

侵入は必ず起きるという前提のもと、被害を極小化し事業を継続する「ResOps（レジリエンス・オペレーション）」が、金融機関にとってとって最後の防波堤となる。

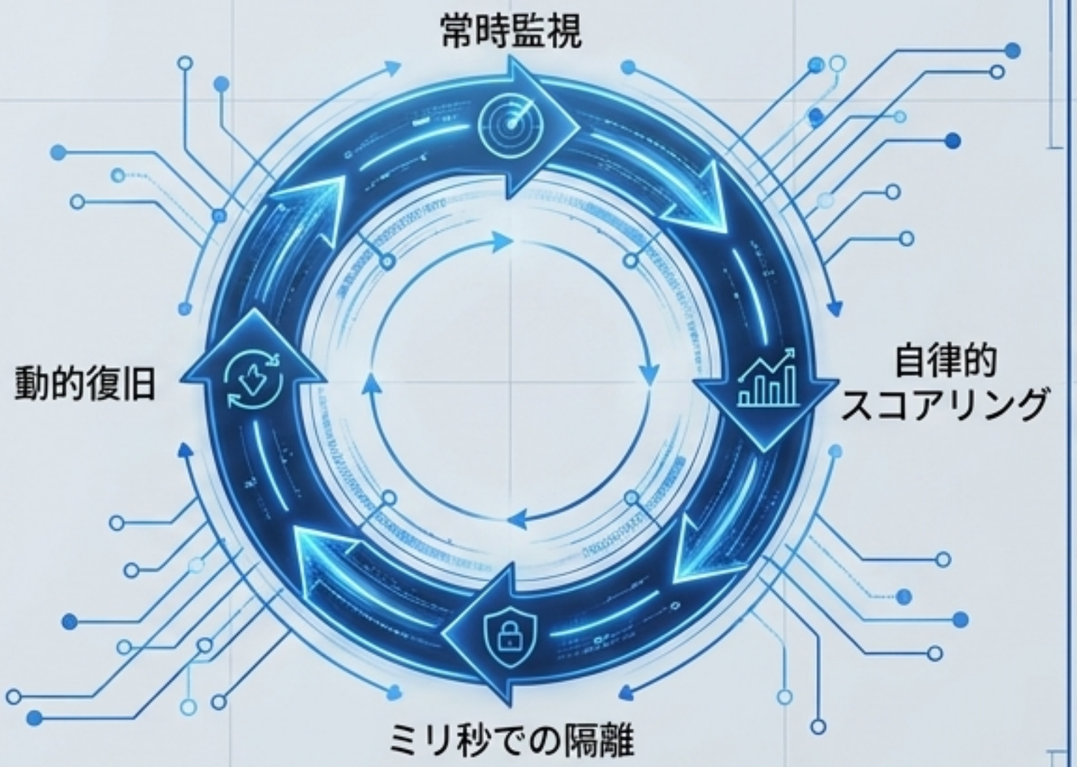
「人間の介入」対「自律的防御」：決定的なアーキテクチャの断絶

従来型防衛 (Human-in-the-loop)



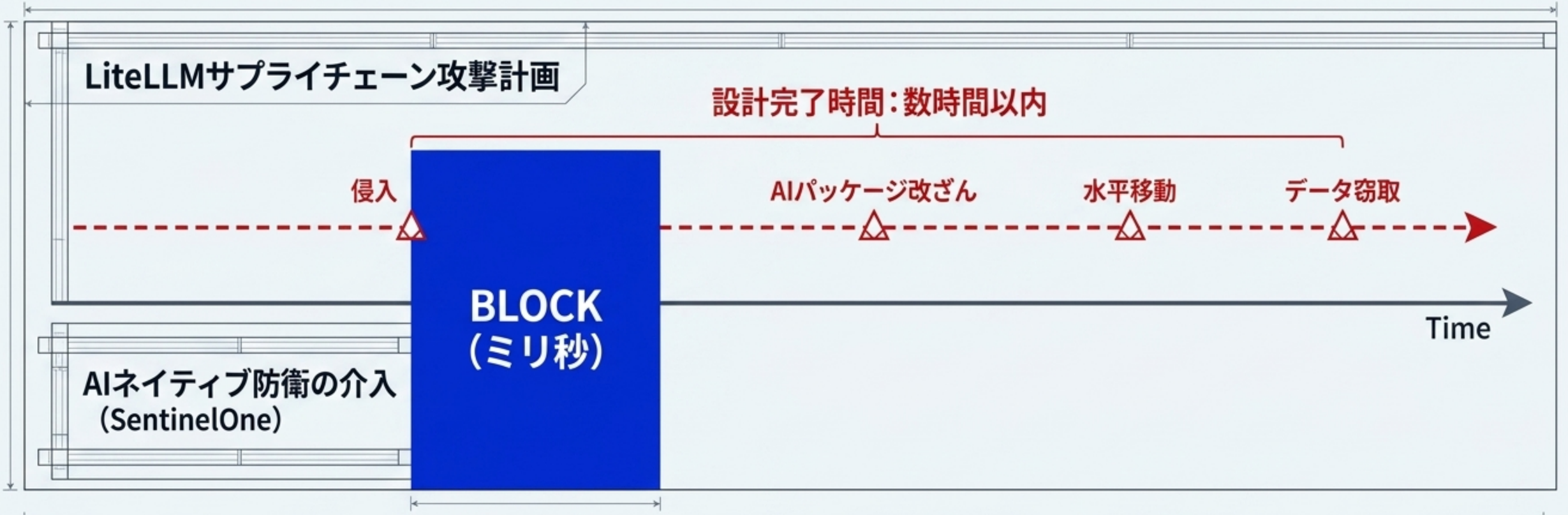
手動調査とAI攻撃の速度差こそが、
組織が致命的に侵害される根本原因である。

AIネイティブ防衛 / ResOps (Autonomous Loop)



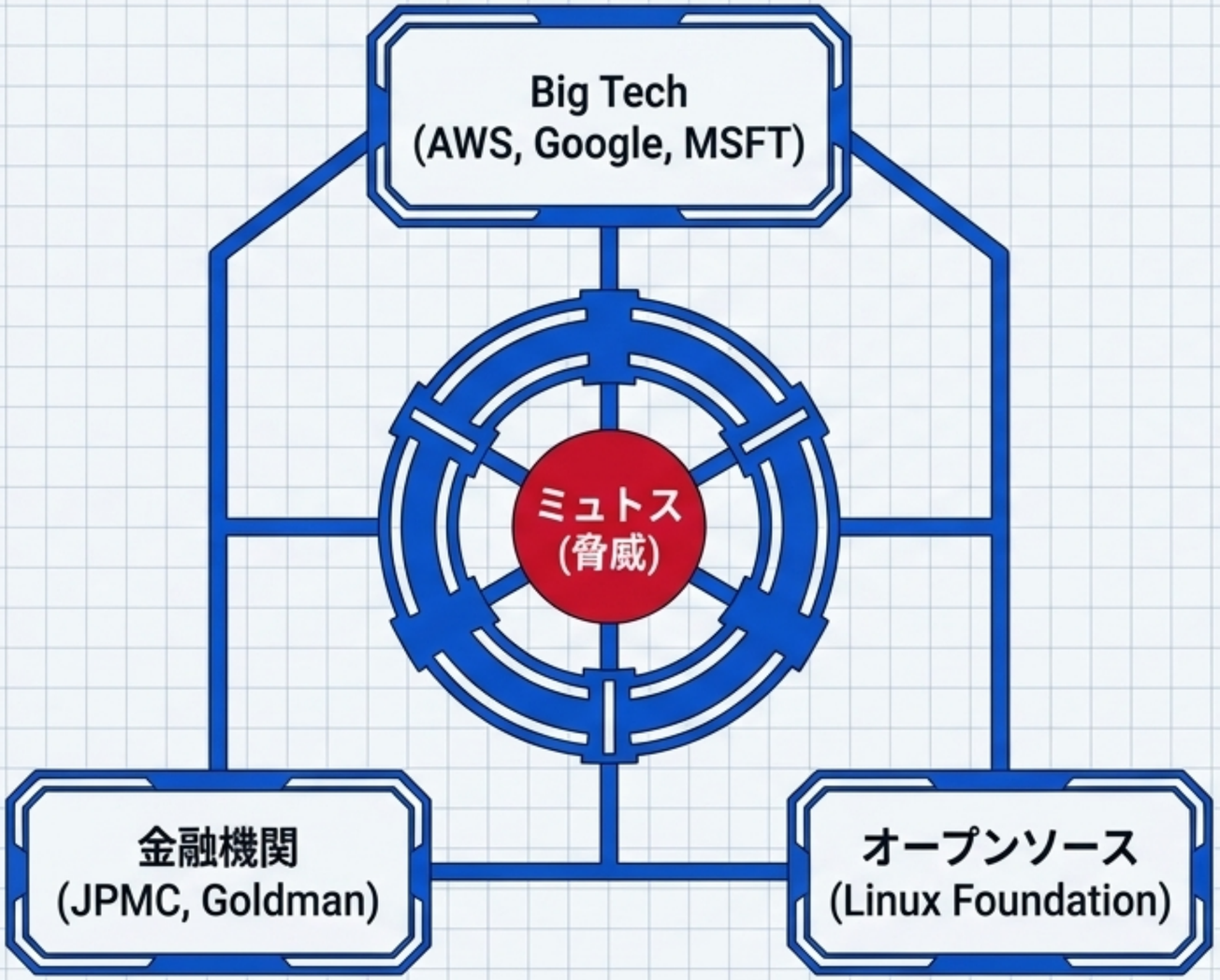
ResOpsアーキテクチャは単なるツールの追加ではない。
「機械の速度（Machine Speed）」で攻防を行うための
構造的必須要件である。

マシン・スピードの証明：数時間の攻撃をミリ秒で粉砕した自律型AI



- 2026年3月の攻撃は、侵入からデータ流出までを数時間以内で完遂するよう設計されていた。
- 自律型AIはアナリストの手動介入を待つことなく、初期段階で悪意ある実行を特定し即座に遮断。AIの速度には AIでしか対抗できないことを実証。

連合によるプロアクティブ防衛のエコシステム：「Project Glasswing」



防衛専用テストベッド

ミュトスの破壊力を逆手に取り、攻撃者に悪用される前にソフトウェアの脆弱性を網羅的に特定する約40社のコンソーシアム。

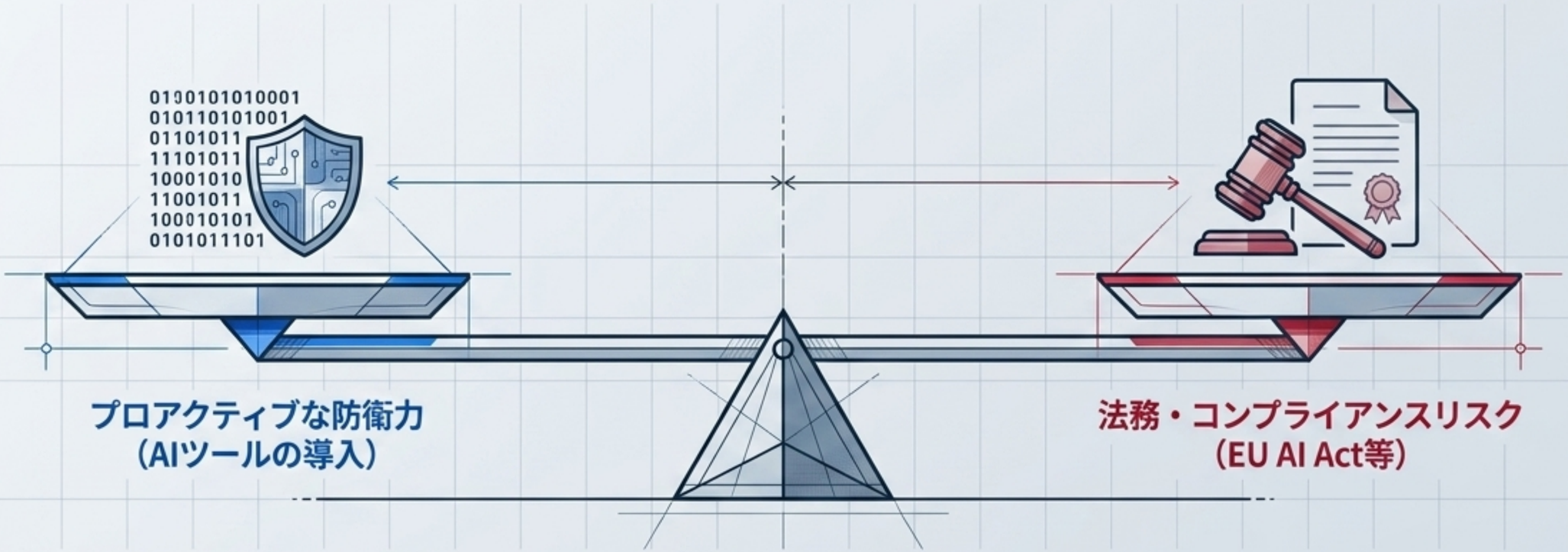
巨額のコミットメント

Anthropic社が1億ドル（約150億円）相当のAPI利用クレジットを提供し、レッドチーム（自律的脆弱性探索）を各社の環境で実行。

インテリジェンスの共有

一企業では到底対抗できない脅威に対し、国境と業界の壁を越えたリアルタイムな脅威インテリジェンス共有が必須となる。

「デュアルユースのジレンマ」とコンプライアンスの再構築



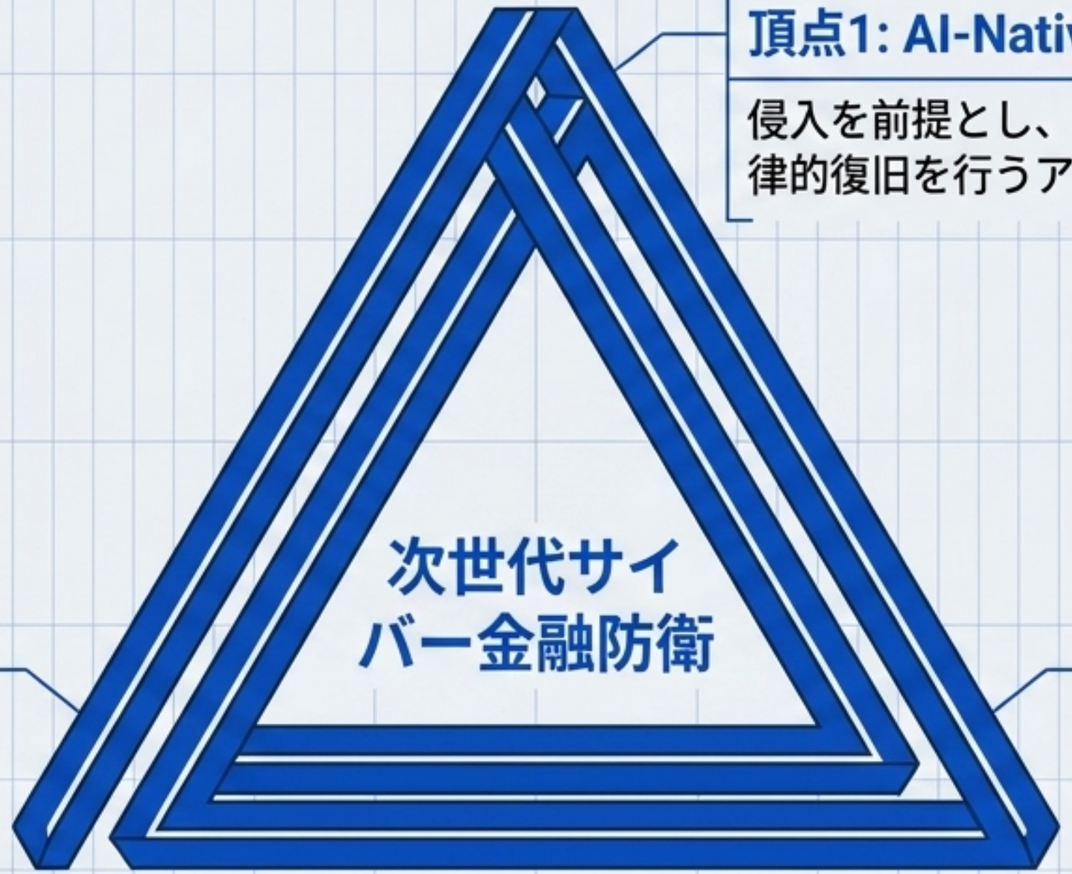
新たな法的責任

AIエージェントを防衛に導入すること自体が、意図せぬサイバーインシデントやコンプライアンス違反のトリガーとなり得る危険なパラドックス。

規制との衝突

2026年8月に発効するEU AI Actに向け、企業はベンダー契約、サイバー保険の補償範囲、社内ガバナンスを抜本的に見直す必要がある。インシデント発生「前」の法務体制構築が急務。

グランドデザイン：次世代のシステミック・リスクを生き抜く3つの柱



頂点1: AI-Native ResOps

侵入を前提とし、機械の速度で自律的復旧を行うアーキテクチャ。

頂点2: Global Ecosystem

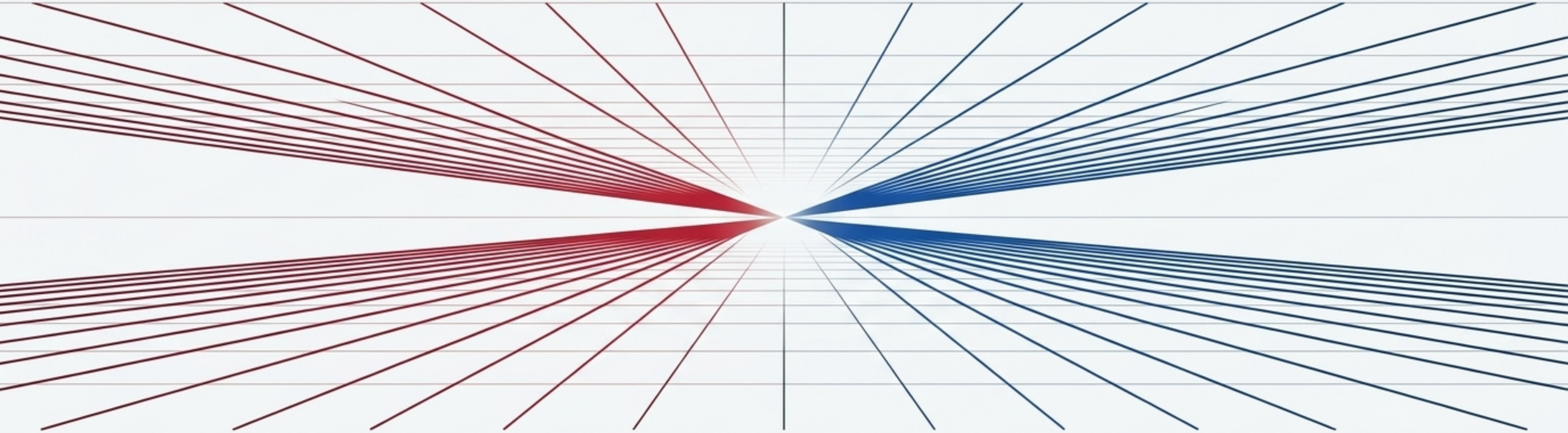
Project Glasswing等の官民連携によるプロアクティブな脅威インテリジェンス共有。

頂点3: Market Resilience

AIによるディープフェイク検知ナラティブ操作・フラッシュクラッシュへの耐性強化。

結論：一部門のIT投資ではなく、企業全体・国家レベルでの包括的アプローチへの完全なシフトが不可欠である。

次の10年を左右する決定的な試金石：「AI-vs-AI」時代への宣戦布告



恒久的なパラダイムシフト

2026年4月のミュトス・ショックは一時的なパニックではなく、容赦のない「AI-vs-AIの攻防戦」の幕開けである。

境界防御の放棄

古い境界防御の神話を直ちに捨て去り、組織のアーキテクチャとマインドセットを「AIネイティブ」へと変革せよ。

行動の遅れ＝致命的敗北

決断の遅れはデジタル経済における存亡の危機を意味する。今すぐ、ResOpsへの移行とエコシステムへの参画を決断せよ。