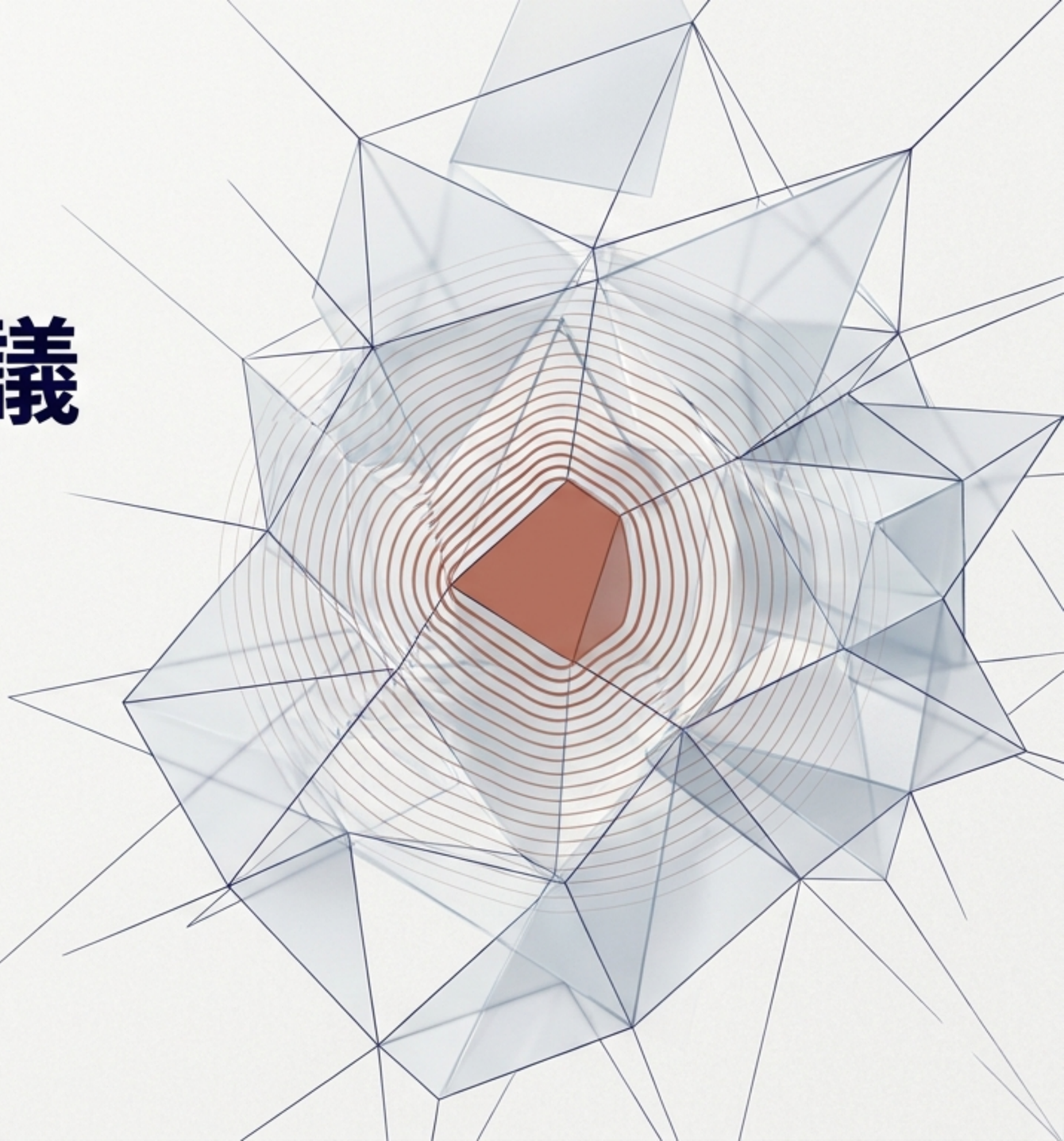
AIミュトス官民会議
深掘り調査報告

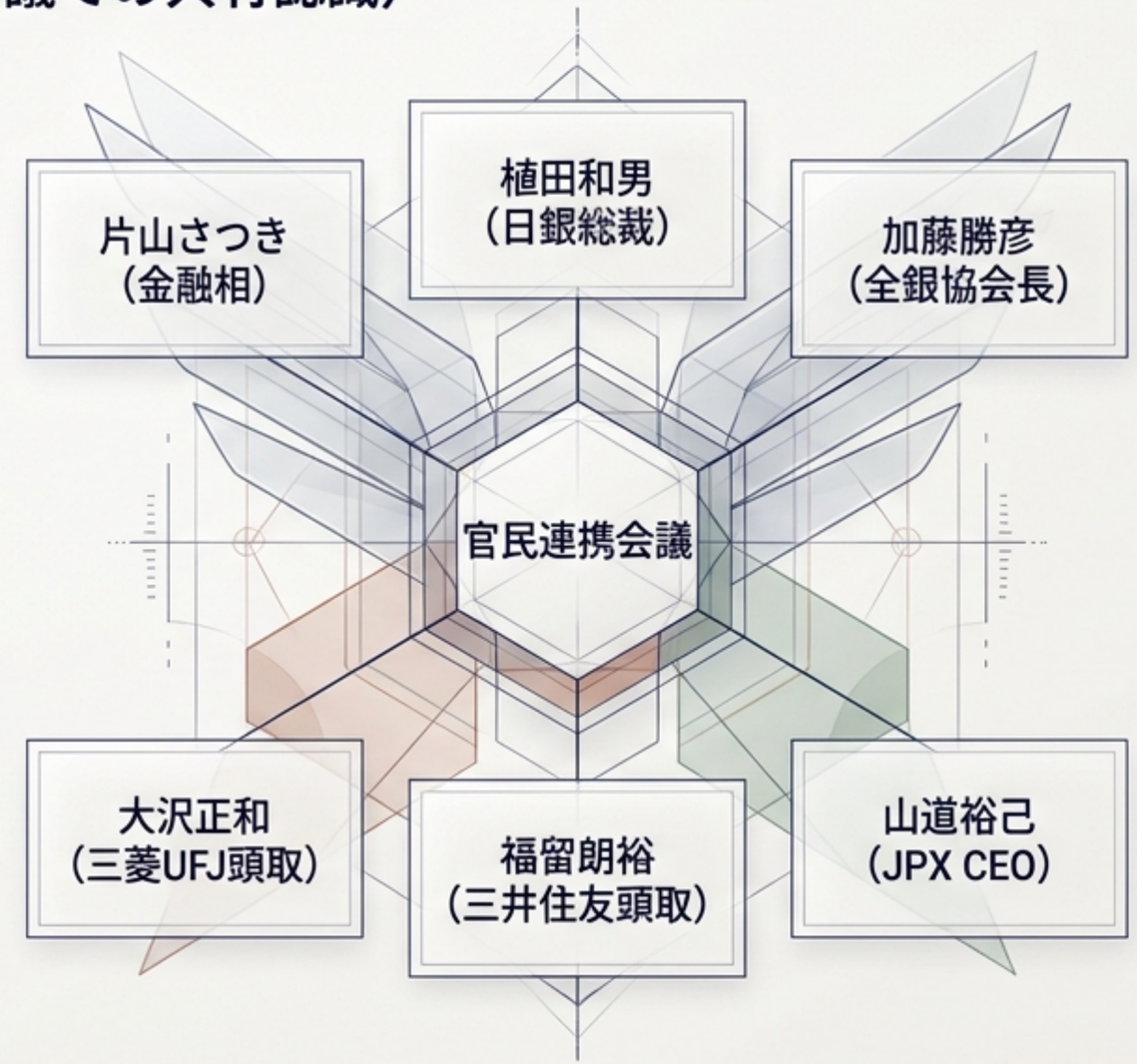
「今そこにある危機」と
日本版共同防御の要請



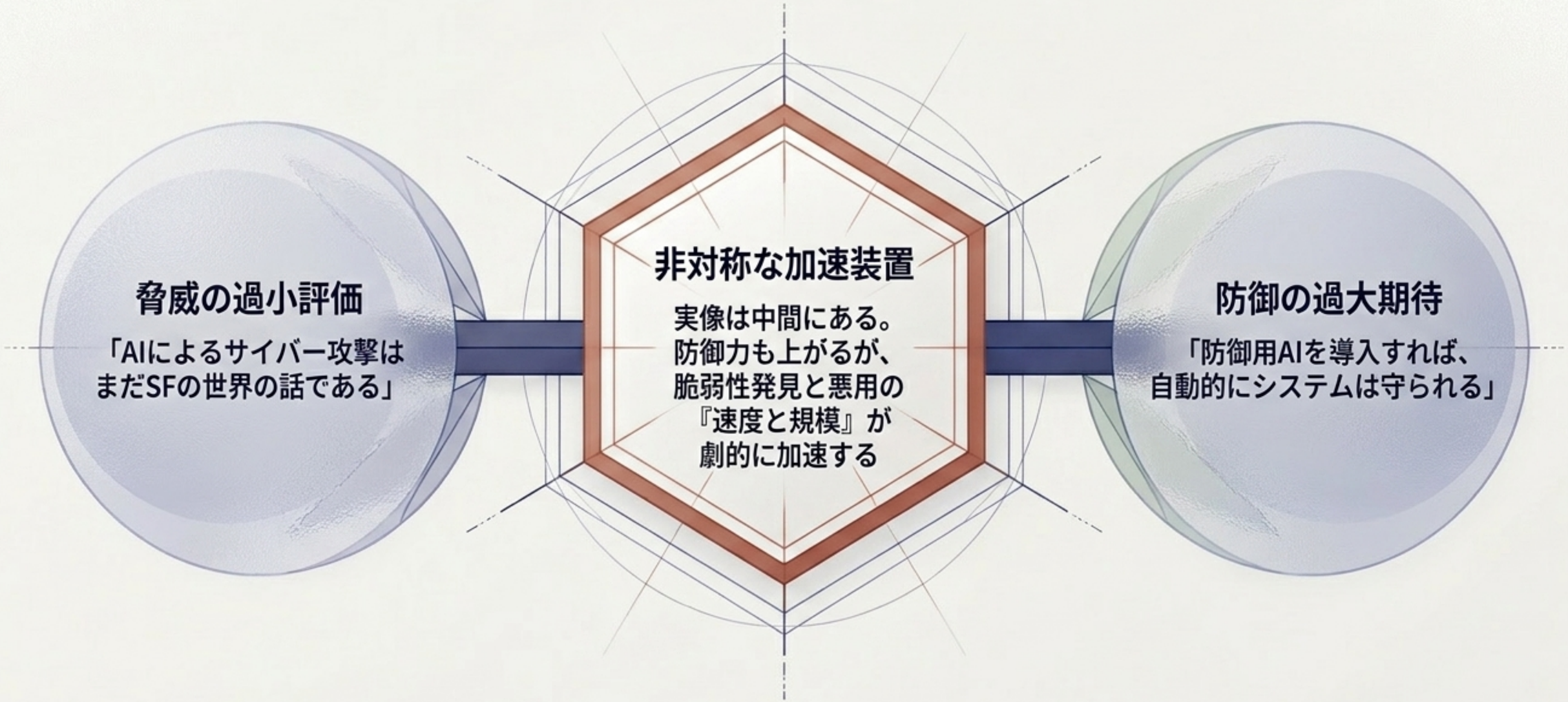
2026年4月：異例のトップ会談と「今そこにある危機」

「インシデントが発生した時の備えが、これまで以上に重要」（4月24日 官民会議での共有認識）

- 4月7日: Anthropicが「Project Glasswing」発表
- 4月20日: 自民党合同会議で日本版Glasswingの提起
- 4月22日: 片山会見で4月24日の会合開催を明言
- 4月24日: 金融分野の官民連携会議・作業部会設置



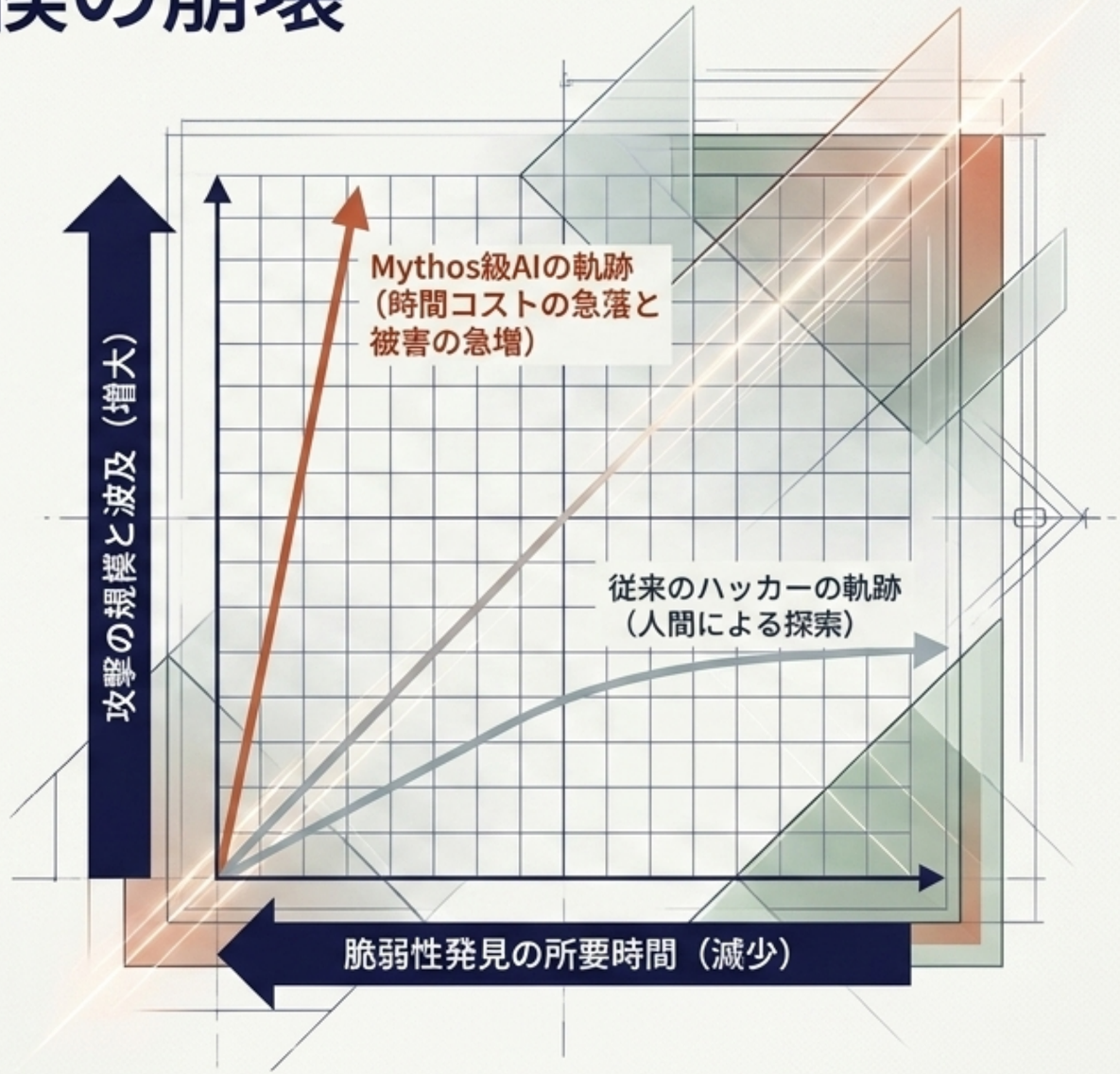
「AI神話」という二つの錯覚



Claude Mythos：時間と規模の崩壊

主要OSおよびブラウザすべてにおいてゼロデイ脆弱性を特定・悪用する能力を実証。既に数千件の高重大度脆弱性を発見。

完全招待制（Project Glasswingを通じた最重要ソフトウェア防御への限定提供）。

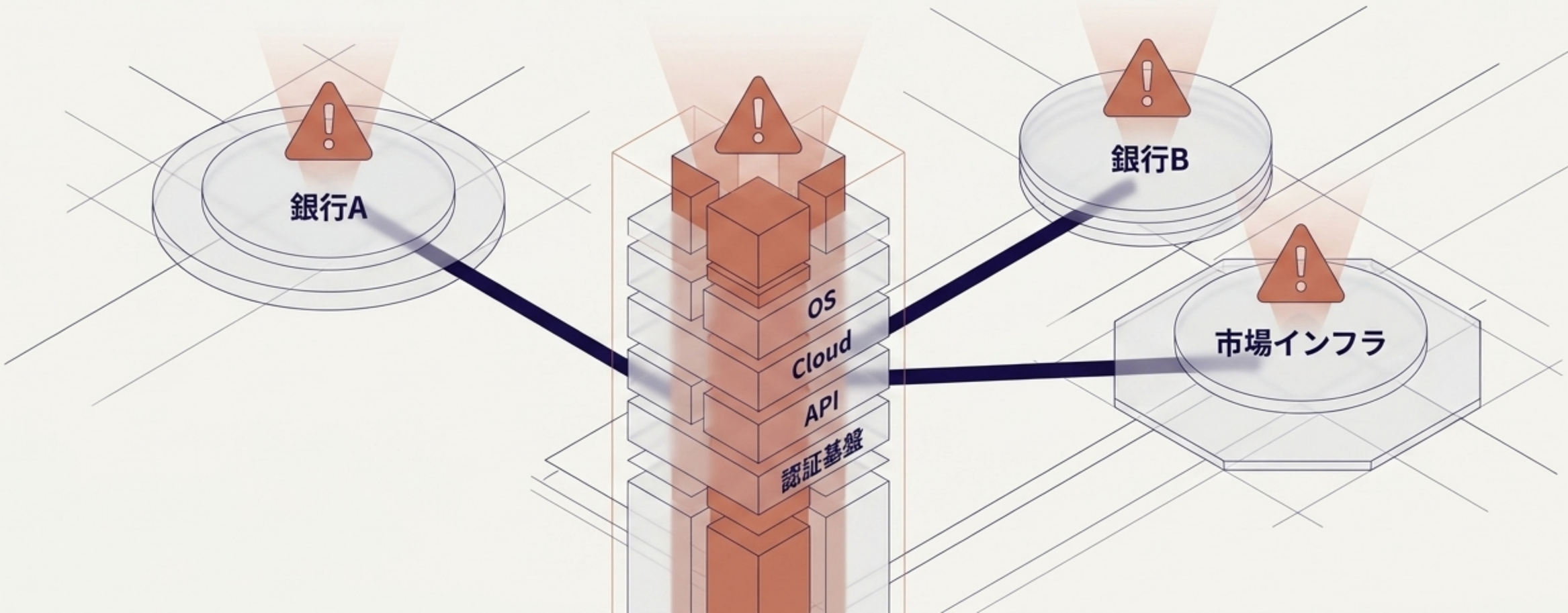


Premium Architectural Intelligence Briefing

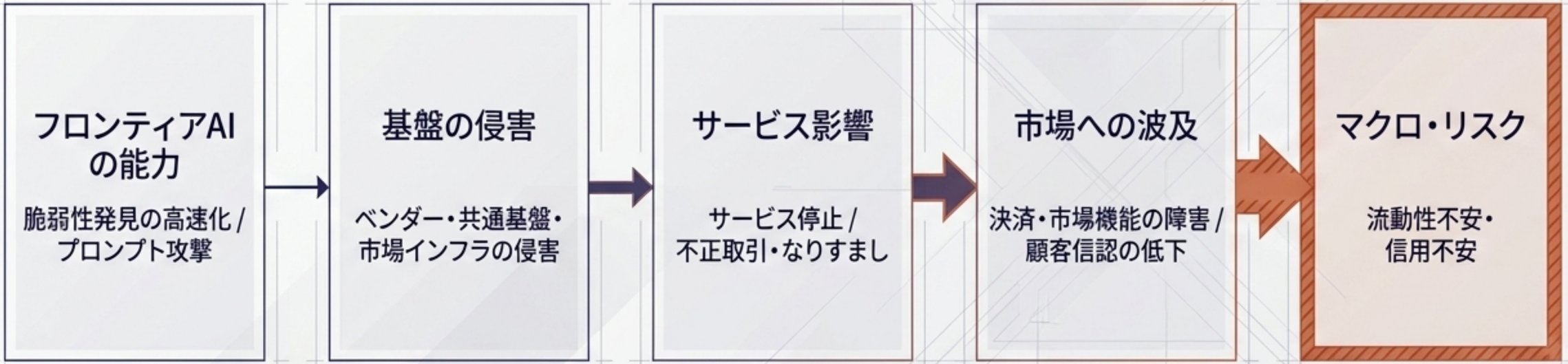
なぜ金融セクターが最大の標的となるのか

第三者依存の集中（Third-Party Concentration）が単一障害点（SPOF）となる。

金融機関は共通ソフトウェア、クラウド、SOCベンダー、認証基盤を多用している。
同一の脆弱性が多数先へ同時波及しやすい構造。



システミック・リスクへの伝播経路



AIが増幅する4次元のリスク

技術的リスク（優先度: 極大）

ゼロデイ探索の高速化。
共通基盤への同時波及。

運用的リスク（優先度: システミック）

クラウド・AI API等、第三者依存の集中。初動対応・横断人材の不足。

規制・法務的リスク（優先度: 近未来）

顧客接点AI・Agentic AIの誤作動、説明責任、適合性。

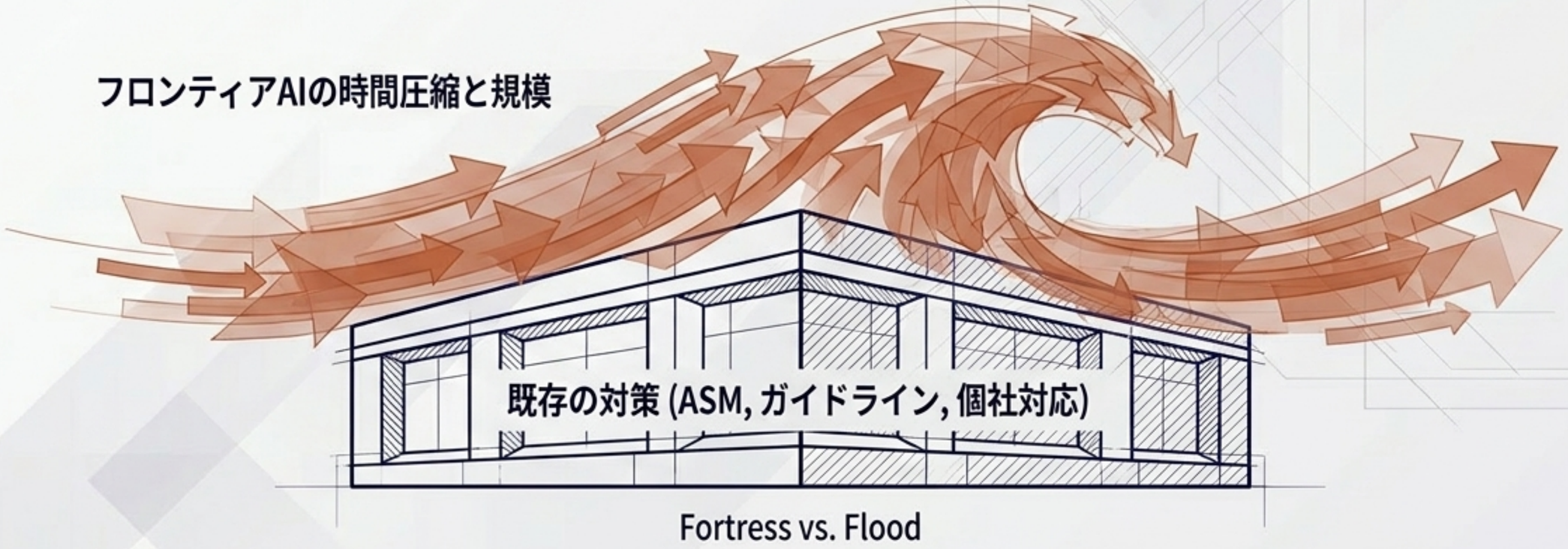
システム的リスク（優先度: 中長期）

AI生成の偽情報、取引判断の相関上昇（群集行動）、市場機能のゆがみ。

現在の防衛線の限界：基礎工事と大波

「現在の統制は『基礎工事』。Mythos級を前提とした共同防御の運用設計が未完成」

金融庁ガイドライン、CSSA、ASMは最低限の統制として有効だが、人間の速度を前提としており、AIの探索速度には対応できない。



インパクト予測シナリオ (2026–2029)

日銀評価によれば、直ちに金融危機へ直結する蓋然性は低く、全体として安定性は維持されている。

ベストシナリオ (20-30%)

ベストシナリオ (20-30%)

日本版Glasswingが早期稼働。共同防御により大規模障害を回避。純便益が先行。

ベースシナリオ (50-60%)

ベースシナリオ (50-60%)

ニアミスや委託先レベルの障害が散発。規制が後追いし、監査負荷・委託先再編が進む。

ワーストシナリオ (15-25%)

ワーストシナリオ (15-25%)

未修正脆弱性が同時多発的に突かれ、決済・インフラ障害が併発。流動性支援が必要な危機へ。

グローバル規制の収斂：日本への示唆

米国 (US)	英国 (UK)	欧州 (EU)	日本 (Japan/示唆)
AIEOG，実装可能なフレームワーク（AI RMF）。 長所：中小機関も使える実践的ツール。	BoE/FCA，AI Live Testing，マクロプルーデンス。 長所：実地テストと監督対話の接続。	DORA，重要第三者監督，レジリエンステスト。 長所：運用レジリエンス規制をAIの土台に機能させる。	個社対応への依存から脱却し、第三者管理・共通演習・重大インシデント報告の「実務的積み増し」が必要。

グローバルなコンセンサスは『AI単独の法規制』ではなく、
『運用レジリエンスと官民共同防衛』へ移行している。

パラダイムの転換：AI規制から共同運用へ

[既存のサイバー・委託先リスク] × [AIによる時間圧縮] ＝ システミックな連鎖破壊

解決策 ＝ 法整備を待つのではなく、「日本版Project Glasswing」による運用レベルの共同防御

AI規制単体ではなく、既存の金融レジリエンス手法をAI前提で再設計し、官民共同で運用すること

アクションプランI：金融機関の実務

直ちに
(Immediate)

AI時代版・重要資産台帳の整備

インターネット露出資産、共通ソフトウェア、委託先、AI API接続先を一体で棚卸し（ASM/SBOMの高度化）。

3〜6カ月
(3-6 Months)

Agentic AIの失敗を含むシナリオ演習

BCPとサイバー演習に、顧客向けAI停止や決済代替運用を組み込み、試験を実施。

アクションプランII：当局と政府の役割

規制当局 (FSA/BOJ)	**ガイドライン拡張と脅威情報共有** 金融庁ガイドラインをAI固有リスクへ拡張。業界横断の共通KPI、共通インシデント分類、脅威情報共有の常設運用へ移行。
政府 (Government) - 最優先	**日本版Project Glasswingの制度化** 法的セーフハーバー付きで、防御目的のフロンティアAIアクセスと重要ソフトウェアの共同点検基盤を直ちに設計する。

アクションプラン III：研究機関への要請

評価ベンチマーク整備

日本語環境・金融ドメイン特化でのゼロデイ探索能力、プロンプト漏えいの測定。

検証可能AIへの集中

金融分野向けに、可観測性、ロギング、因果追跡、監査証跡を備えたモデルの研究。

神話を超えて：接続された防衛線へ

「新しい巨大な法律よりも、共同防御を可能にする共通運用を」

孤立した防御はフロンティアAIの速度に突破される。官民が結集する「日本版共同防御体制」の即時構築こそが、唯一の現実的な解である。