INTEROP 2026 KEYNOTE

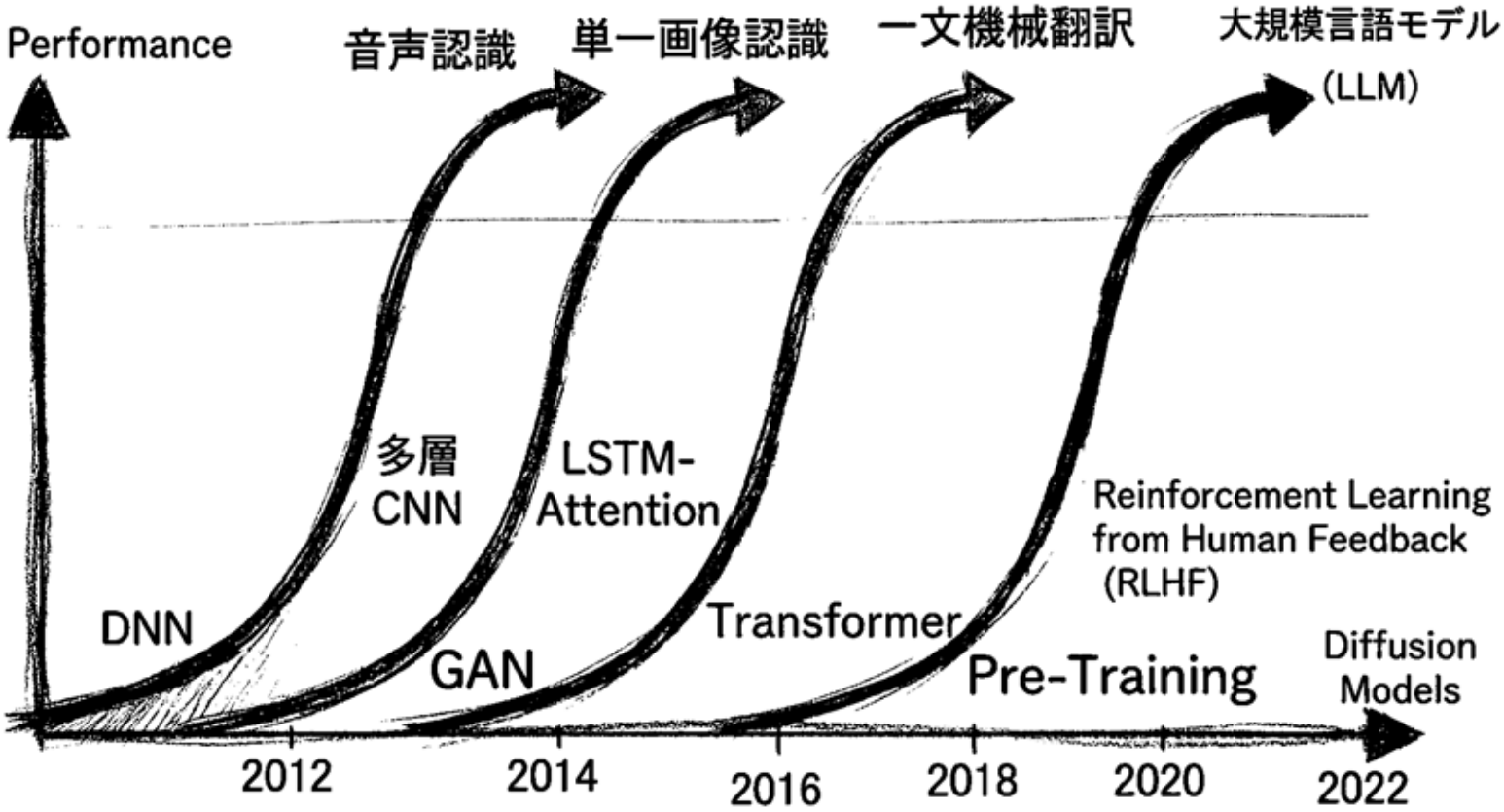
～「人がつくらない通信」をどう守るか～

AIトラフィックが主役になる時代のネットワークとセキュリティ設計

栄藤　稔（えとう　みのる）
CXO Advisor, Netskope, Inc.

自律型AI（Agentic AI）時代におけるサイバー防衛戦略の再定義
― Netskopeの最新アーキテクチャと、フィジカル／ロジカルAIセキュリティの収束 ―

10年前に誕生した相関演算からAIの指数関数的進化が始まった。



近くにある未来、汎用AI
(Artificial General Intelligence)

•敵対的な2時間のチューリング・テストに合格すること。

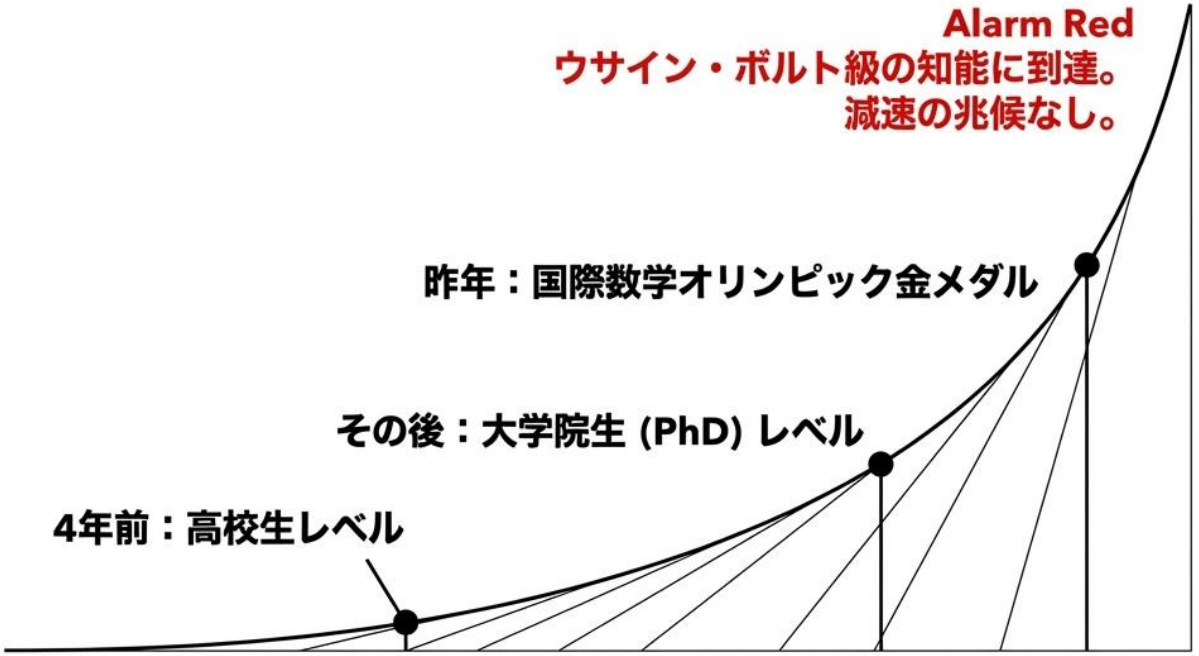
•知識・論理のQ&Aベンチマークで広範に成功すること。

•複雑な模型（モデルカー）の組立手順を正しく解釈し、実行できること。

Not Yet

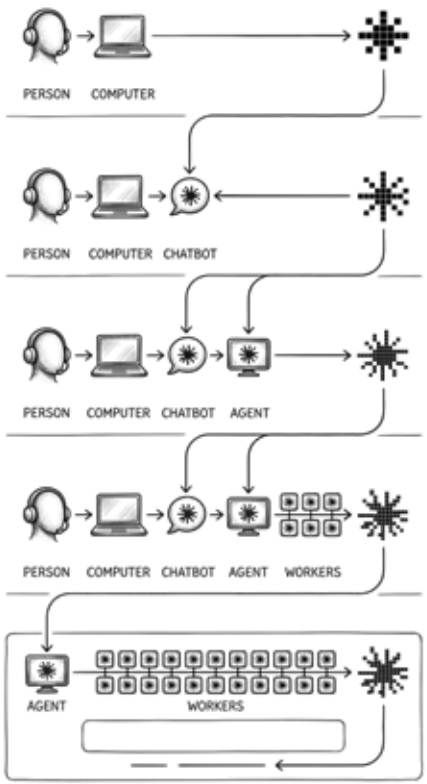
チューリング・テストは、すでに終わっている

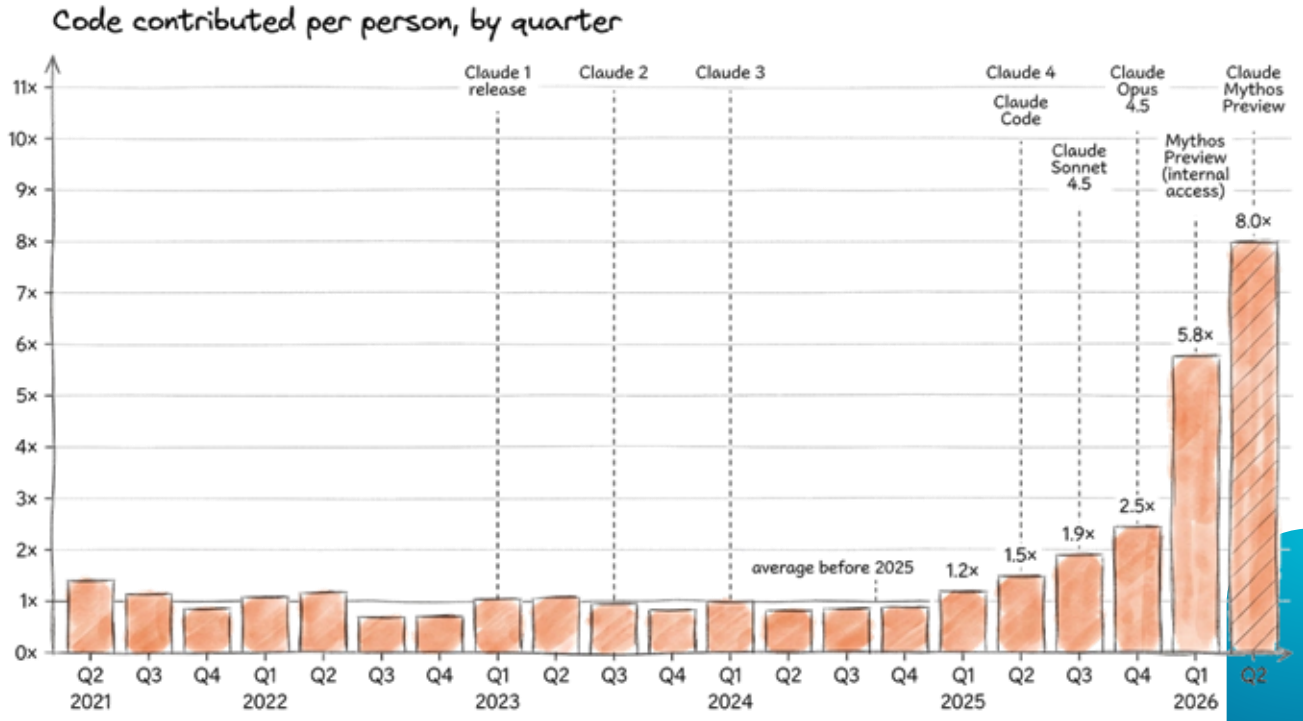
私たちは「その瞬間」を待っていたが、ゴールポストが動かされ、誰も気づかなかった。



2026-2027年：AGI到来のタイムライン

ループが閉じる時：Recursive-Self-Improvement(RSI)





Claude開発者のコード開発量

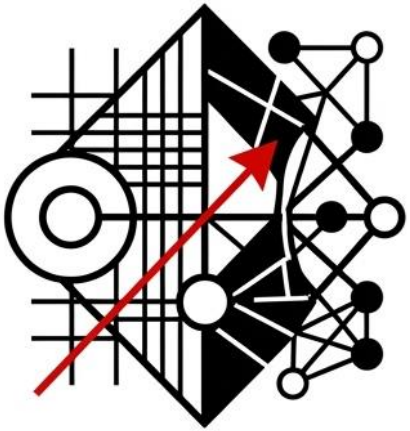
https://www.anthropic.com/institute/recursive-self-improvement

コントロール問題

チンパンジーは人間を制御できるか？

チンパンジー

人間



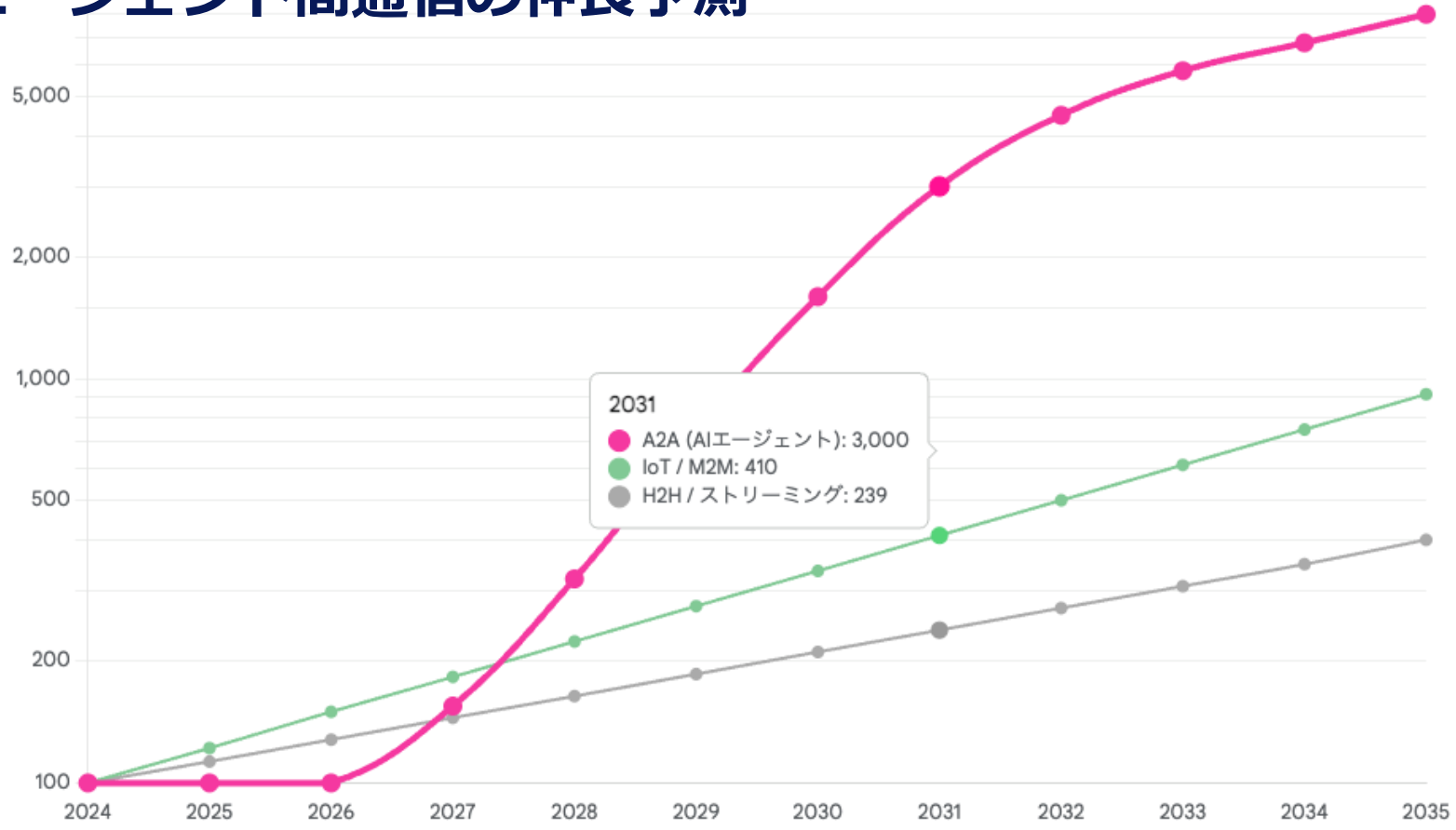
超知能

チンパンジーが人間よりも劣る知能で、人間を檻に入れることはできない。逆も然りである。

Control (支配):	スイッチを切る権限を人間が持つこと。（超知能相手には不可能）
Alignment (協調):	AIが地球のボスになるが、「人間に優しくしよう」と決めてくれること。

アライメントは、AIの慈悲に依存する脆弱な戦略である。

エージェント間通信の伸長予測



MCP通信の現状

AIエージェントが、企業のツール・データ・APIを呼び出すための標準接続レイヤ。

結論：制御点が必要

MCPを使う代表的サービス / クライアント

IDE・AIチャット・社内エージェントが、MCPサーバー経由でGitHub、DB、SaaS、ファイルに接続

- **Claude** — Desktop / Code
- **Cursor** — AI IDE
- **VS Code** — 開発環境
- **ChatGPT** — AI client
- **Windsurf** — AI IDE
- **社内Agent** — 業務自動化

1 認証・権限が弱い

認証なし公開、過剰スコープ、APIキー直置きが残る。

例：測定対象の40.55%が認証なしでツールを公開

2 ツールを信頼しすぎる

ツール汚染、なりすまし、rug pullで呼び出し先が変わる。

対策：承認済みMCPサーバーの棚卸しとリスク評価

3 M2M通信が見えない

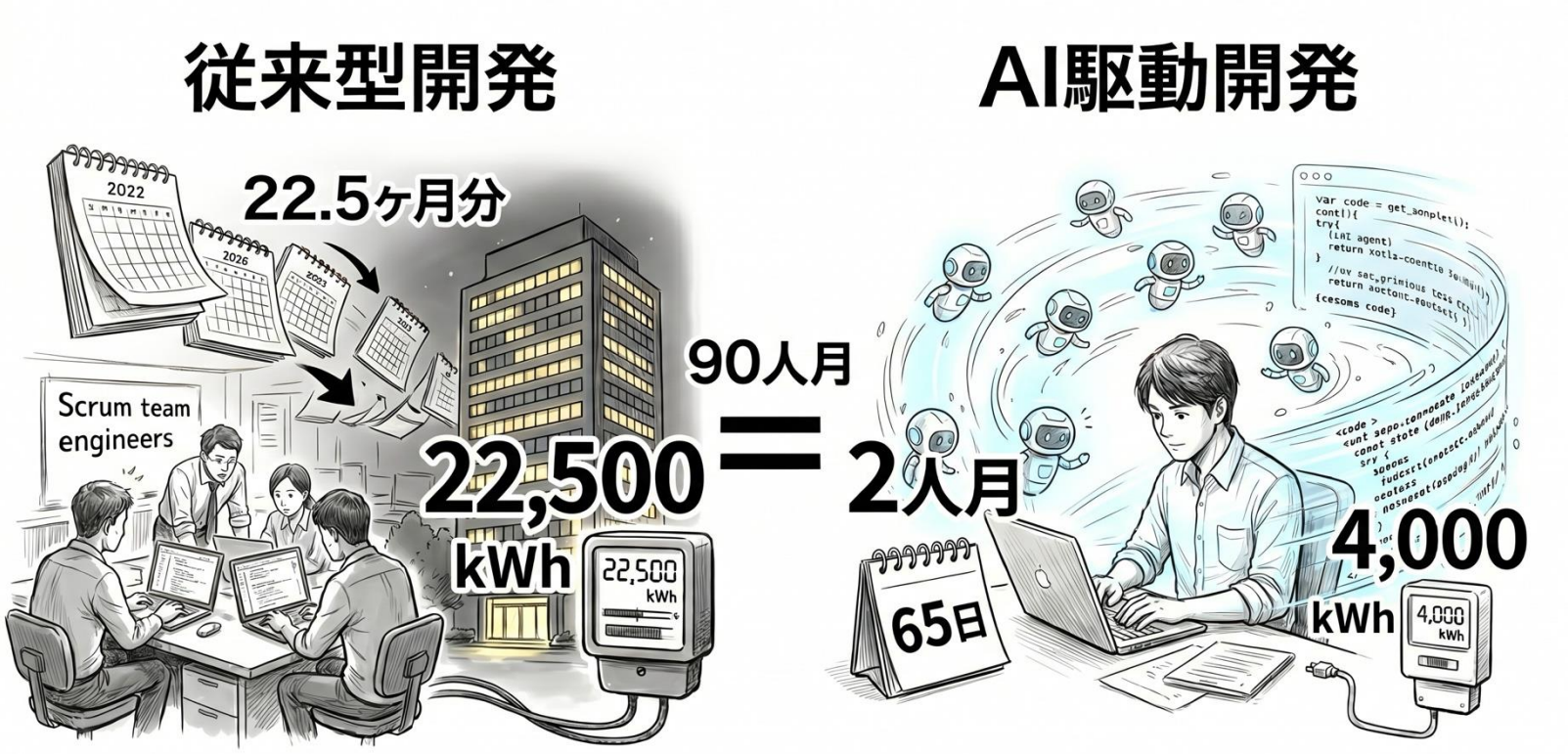
誰が、どのツールを、どのデータに使ったか追えない。

例：MCPポリシーありは8%、M2M可視性なしは36%

実務上の設計：MCP Broker / Gateway で「発見・認証・最小権限・DLP・監査」を一箇所に集約する。

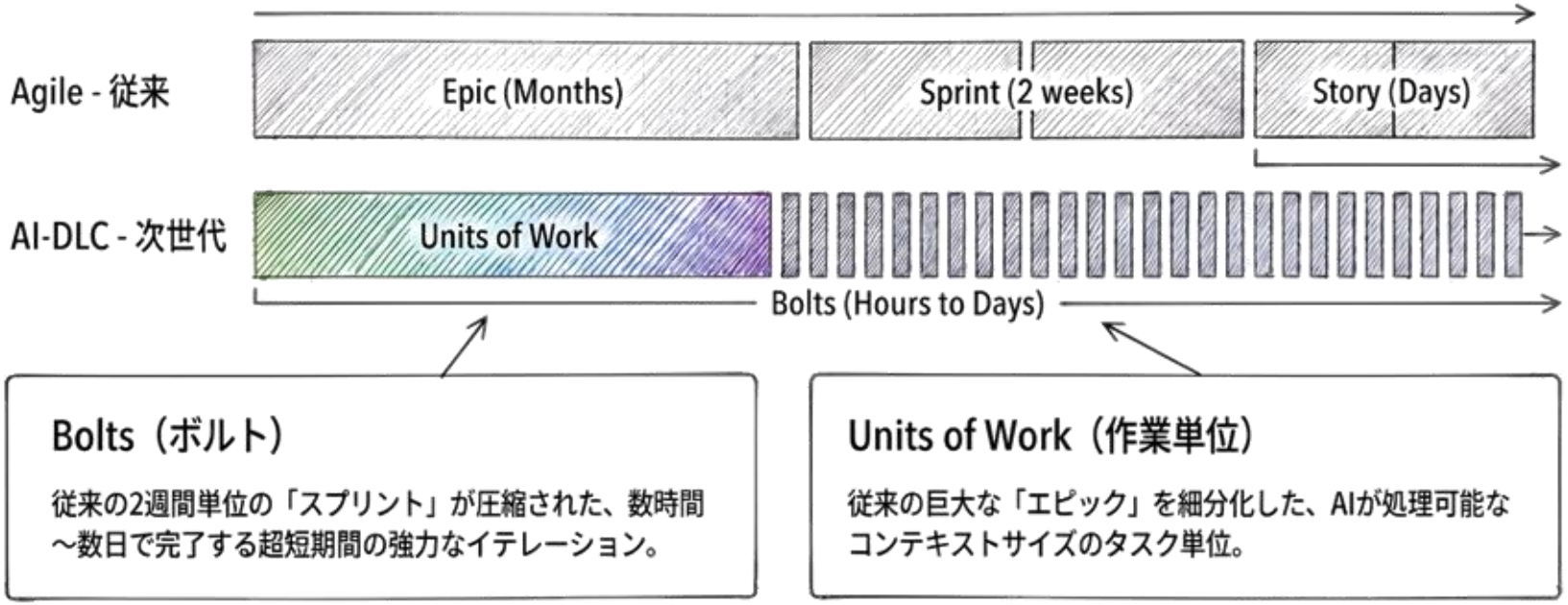
Sources: MCP official docs, Netskope 2026 AI Risk and Readiness Report, arXiv:2605.22333, arXiv:2603.13417, CSAI/CSA MCP security notes, Netskope Agentic Broker DS.

私個人が1人で経験したAI駆動開発 April-May in 2026



生産性 45倍 / 消費電力 1/5.6

AI駆動開発の到来



Bolts（ボルト）

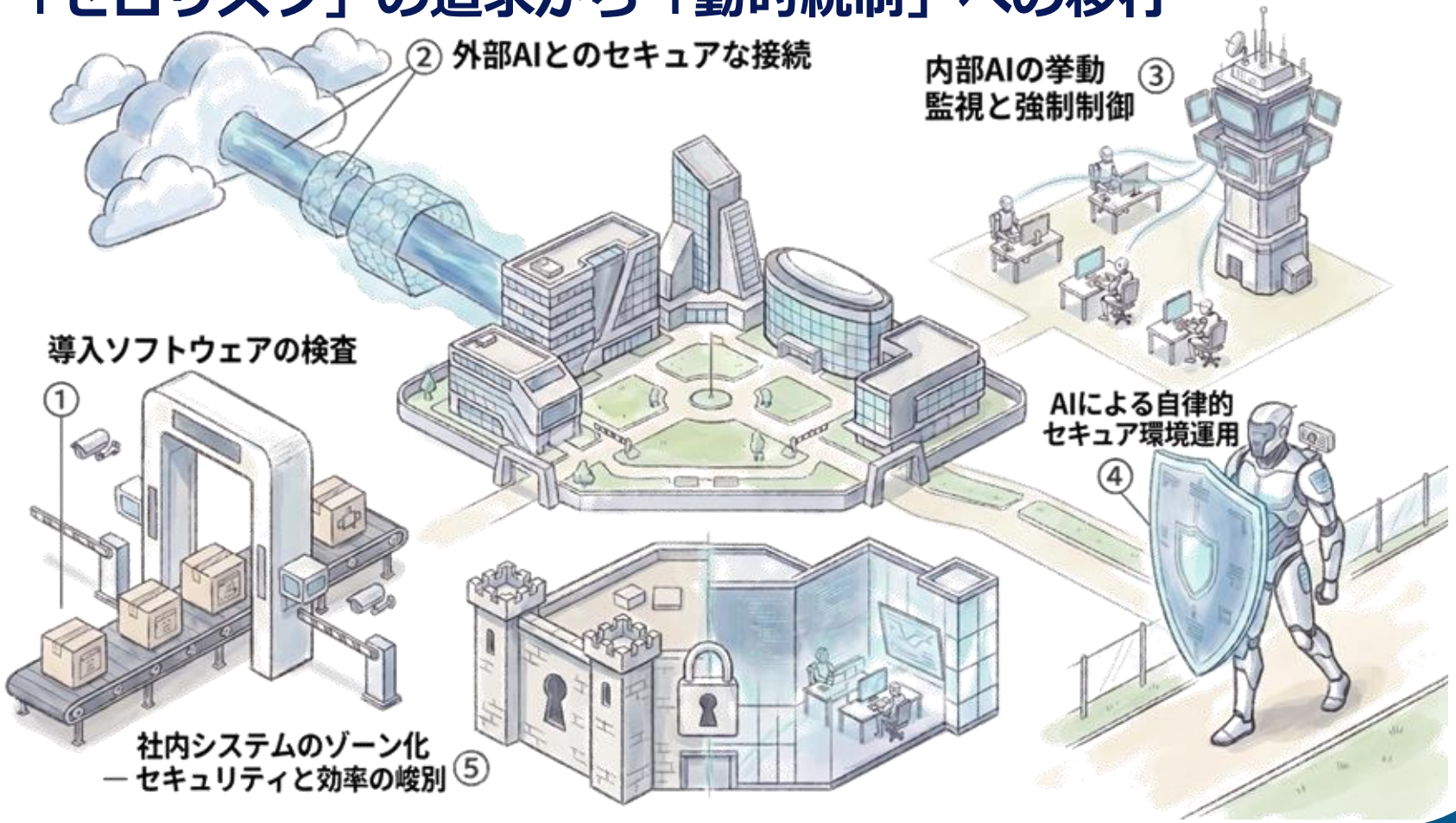
従来の2週間単位の「スプリント」が圧縮された、数時間～数日で完了する超短期間の強力なイテレーション。

Units of Work（作業単位）

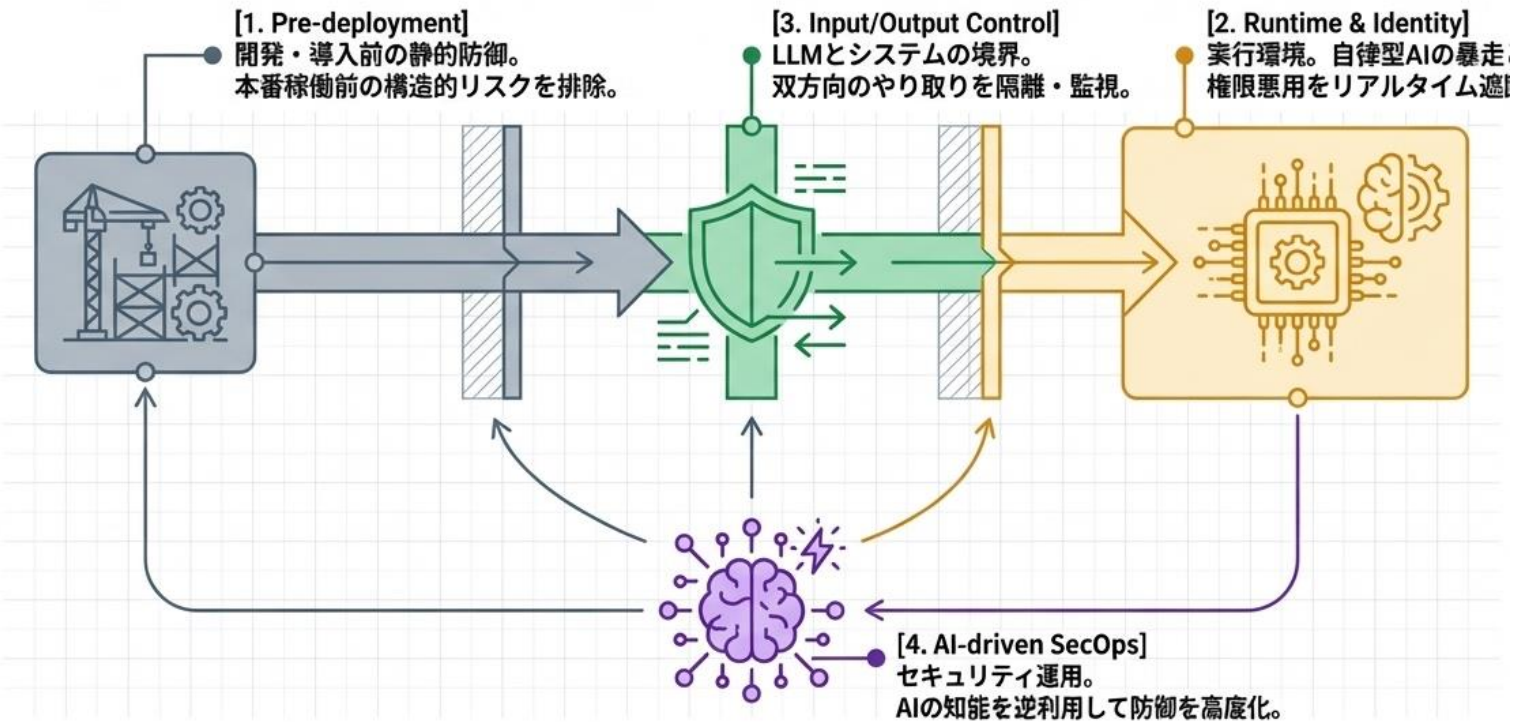
従来の巨大な「エピック」を細分化した、AIが処理可能なコンテキストサイズのタスク単位。

徹底した「意図（仕様）」と「手法（計画）」の分離。AIに未定義の要件を推測させず、実装前に人間の検証を挟むことで手戻りを排除する。

「ゼロリスク」の追求から「動的統制」への移行



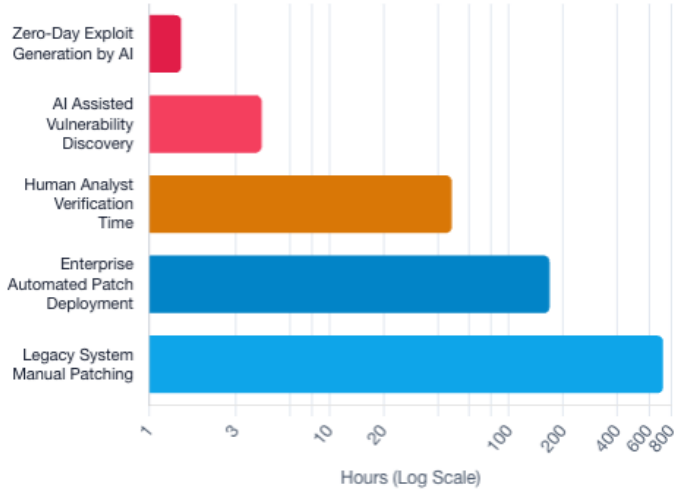
4つの観点:
AIセキュリティ・アーキテクチャ全体図



Mythos of Cyber Warfare:Zero-Day AI vs Auto-Patching

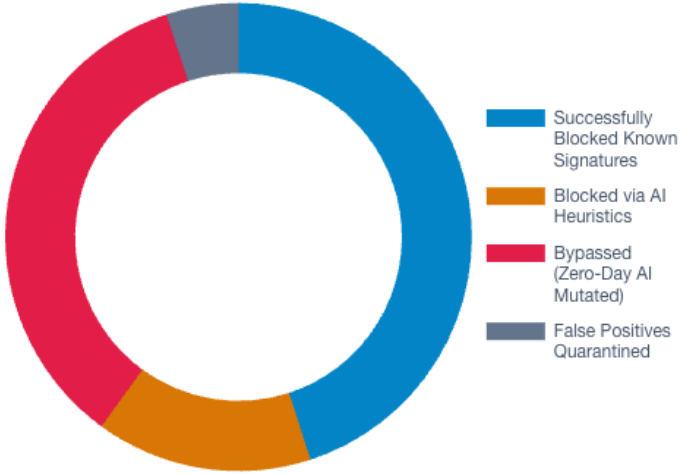
⚡ 1. 非対称性：攻撃の生成 vs 防御の適用

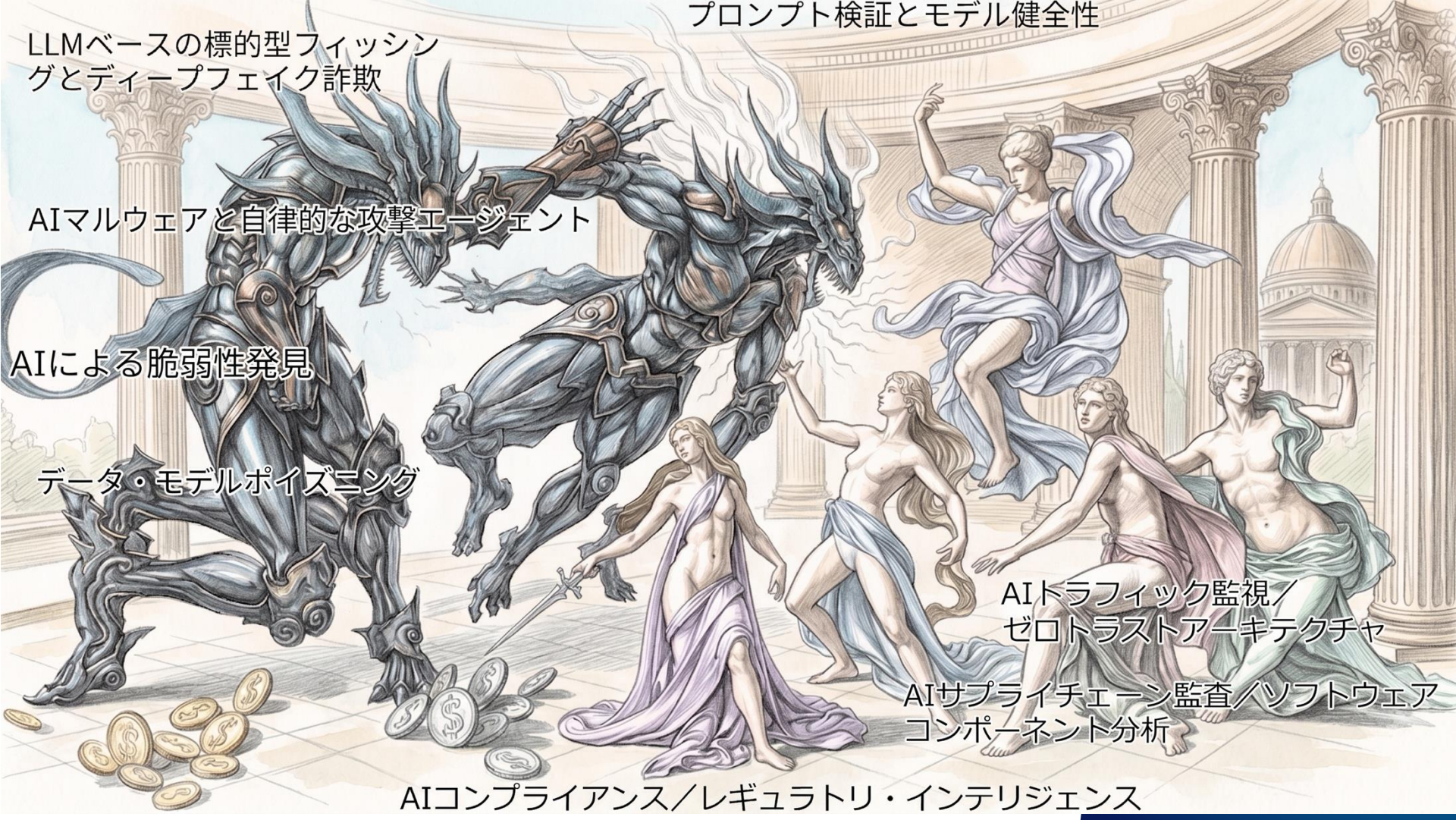
AIを用いたゼロデイ攻撃の生成速度は、AIによる自動パッチ適用サイクルを大きく上回っています。攻撃AIはコードの海から脆弱性を瞬時に見つけ出しエクスプロイトを生成しますが、防御側は「システムの安定性テスト」という物理的・プロセス的なボトルネックを抱えています。



2. SaaSによる安全性検査の限界

ダウンロードされたソフトウェアを検査するSaaSプロダクト（サンドボックス、静的解析等）は、AIが生成する「ポリモーフィック（多態性）マルウェア」に対して極めて脆弱です。既知のシグネチャベースの検知はほぼ無力化し、ヒューリスティック検知もAIの高度な難読化によりバイパスされています。

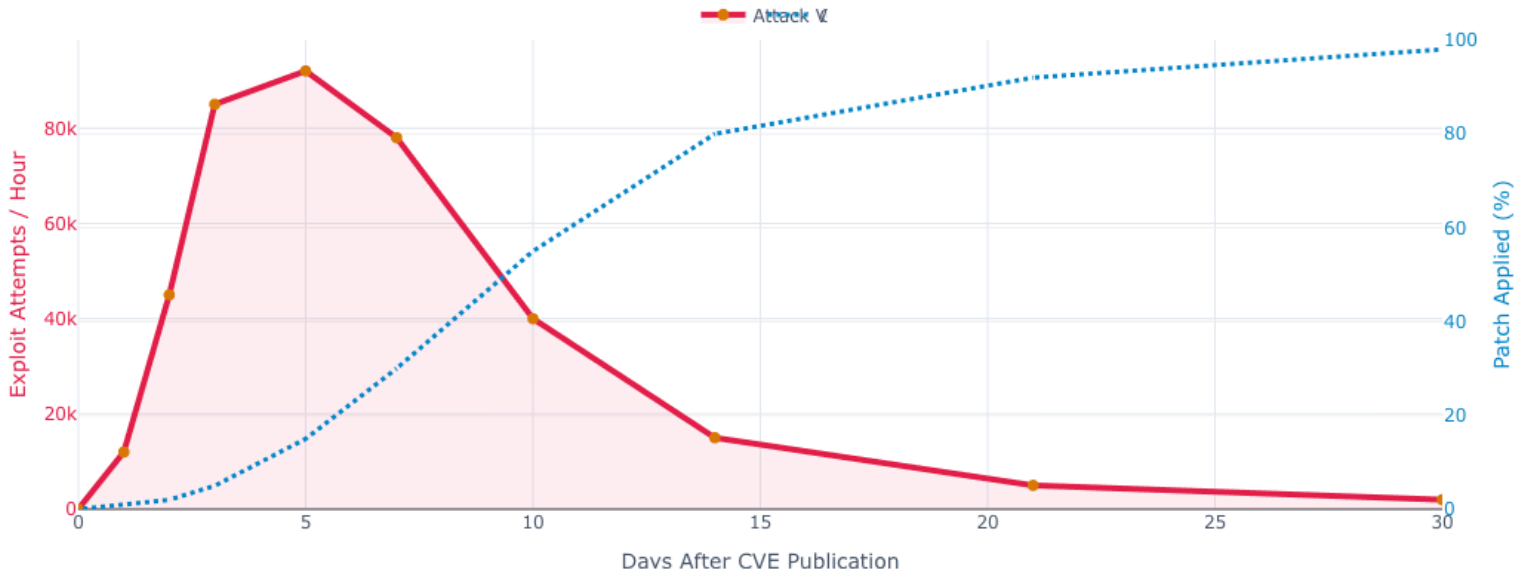



プロンプト検証とモデル健全性
LLMベースの標的型フィッシングとディープフェイク詐欺
AIマルウェアと自律的な攻撃エージェント
AIによる脆弱性発見
データ・モデルポイズニング
AIトラフィック監視／ゼロトラストアーキテクチャ
AIサプライチェーン監査／ソフトウェアコンポーネント分析
AIコンプライアンス／レギュラトリ・インテリジェンス

Mythos of Cyber Warfare:Zero-Day AI vs Auto-Patching

⚔️ 3. CVE情報のパラドックス（コインの表裏）

脆弱性情報（CVE）の公開は、防御側にパッチを促す目的で行われますが、現在では「AIへの攻撃ターゲット指示書」として機能しています。脆弱性が公表された瞬間から、世界中の攻撃AIがPoC（概念実証）コードを生成し、数分から数時間以内に大規模なスキャンと攻撃を開始します。パッチ適用が完了するまでの「兵器化ウィンドウ」が攻撃者の最大の狩場となっています。



Mythos of Cyber Warfare:Zero-Day AI vs Auto-Patching

オープンソース (OSS)

Linusの法則:「目玉の数さえ十分あれば、どんなバグも深刻ではない」

AI時代の現実: 攻撃AIにとってOSSは格好の訓練データ。ソースコードが公開されているため、AIは未知の脆弱性（ゼロデイ）を密かに発見し、パッチが当たる前に長期間エクスプロイトを保持することが可能です。

Transparency Risk

プロプリエタリ (秘匿)

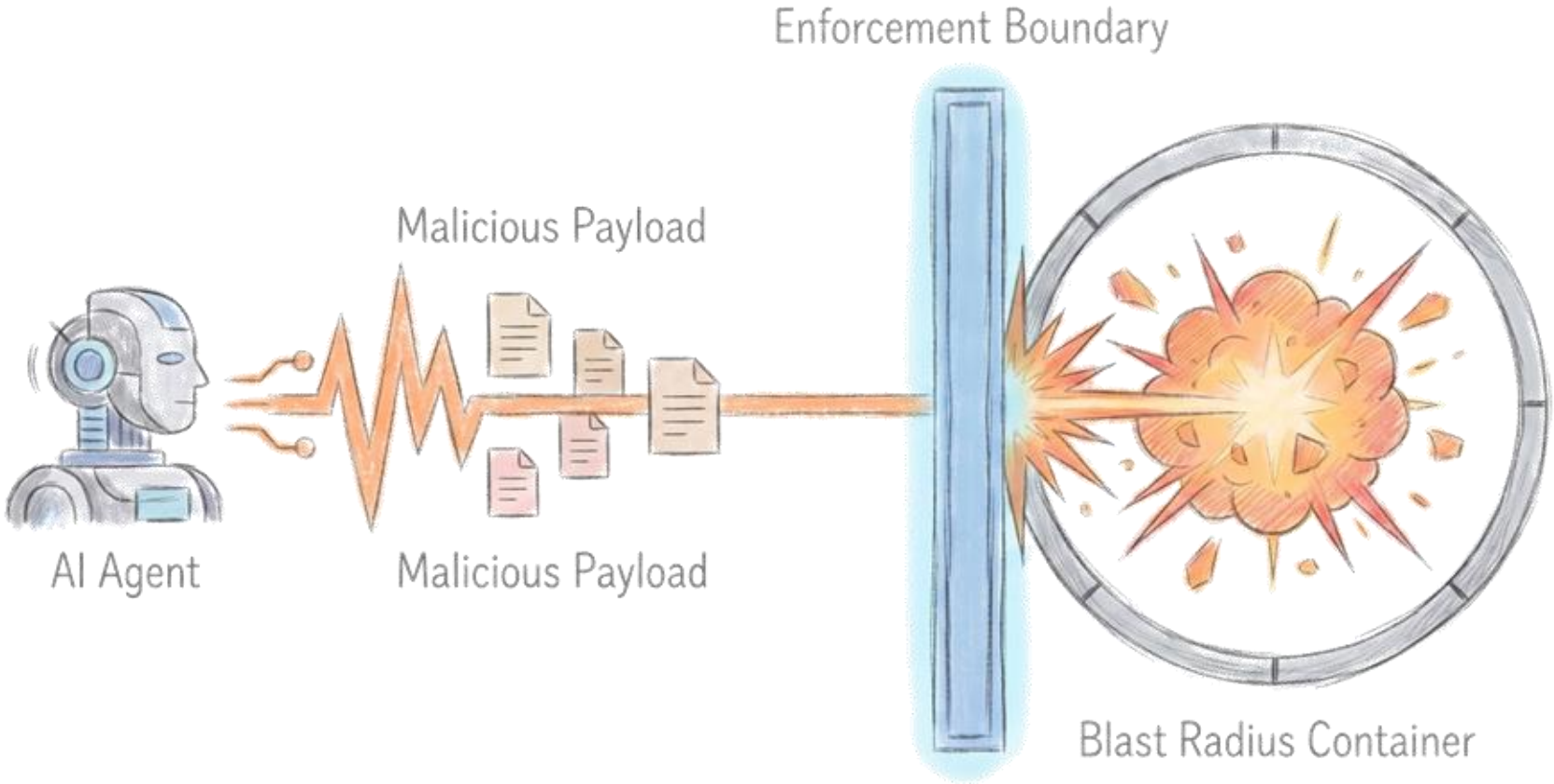
Security by Obscurity: コードを隠すことによる防衛。

AI時代の現実: ブラックボックス解析AIのリバースエンジニアリング能力が向上。内部に潜む脆弱性が発見された場合、外部のセキュリティコミュニティによる監視がないため、ベンダーの対応能力（内部AIの質）に完全に依存する致命的リスクがあります。

Single Point of Failure

Jeetu Patel氏の講演（出典: RSAカンファレンス2025基調講演）

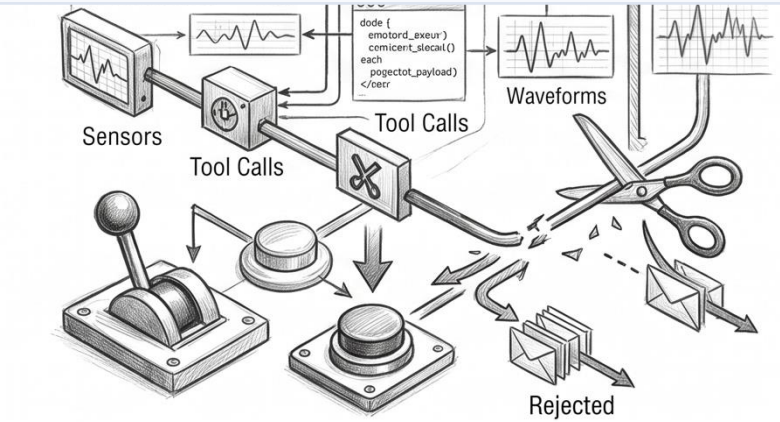
Runtime Enforcement(実行時強制制御)



1 AI向け・実時間強制制御型ゼロトラスト

Runtime Enforcement — 暴走を「その場で」止める

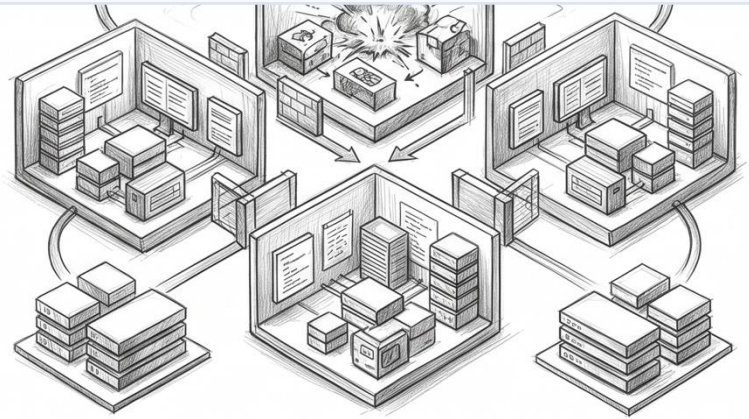
- 非人間ID（AIエージェント）の認証・認可と最小特権を動的に管理
- MCP / A2A通信の経路にブローカーを介在させ、暗号化トラフィックの中身（Tool Calls・ペイロード）をインラインで可視化
- ミリ秒単位でポリシー評価し、逸脱を検知した瞬間に通信を強制遮断・権限剥奪（論理的キルスイッチ）
- 防御の運用自体もAI化し、検知→判断→遮断を機械速度で完結



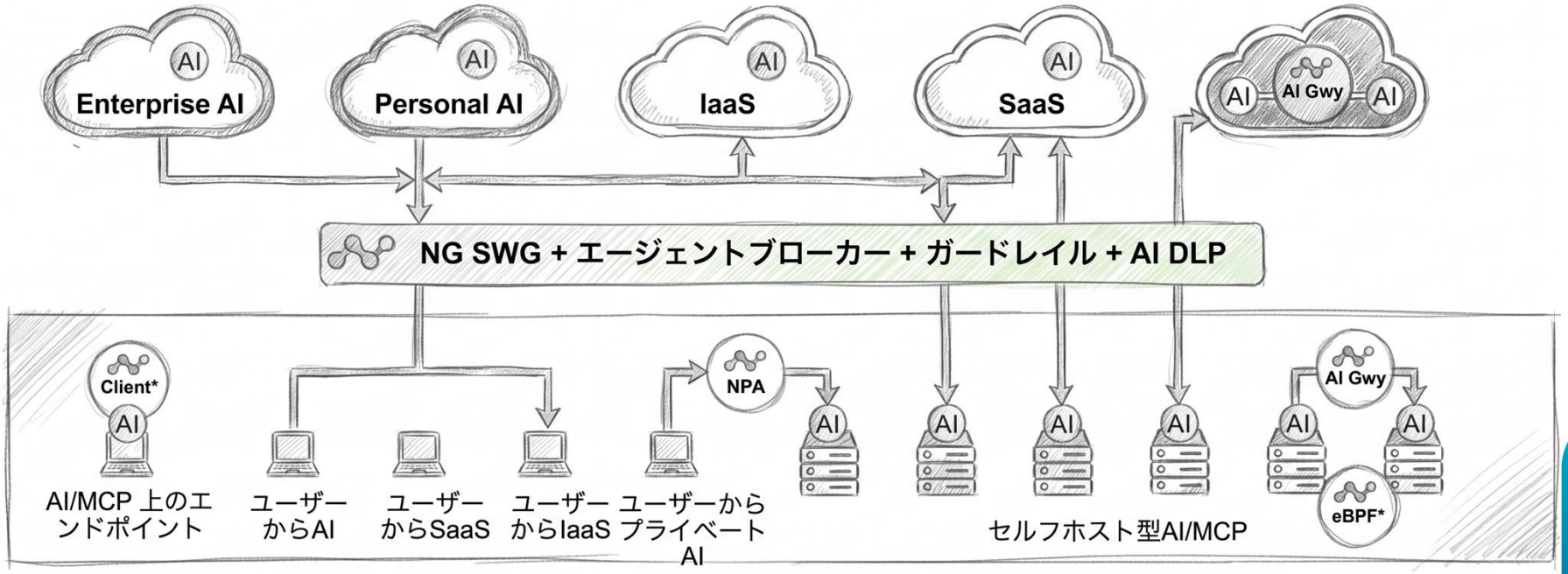
2 システムの分散化

Blast Radius — 突破されても「致命傷」にしない

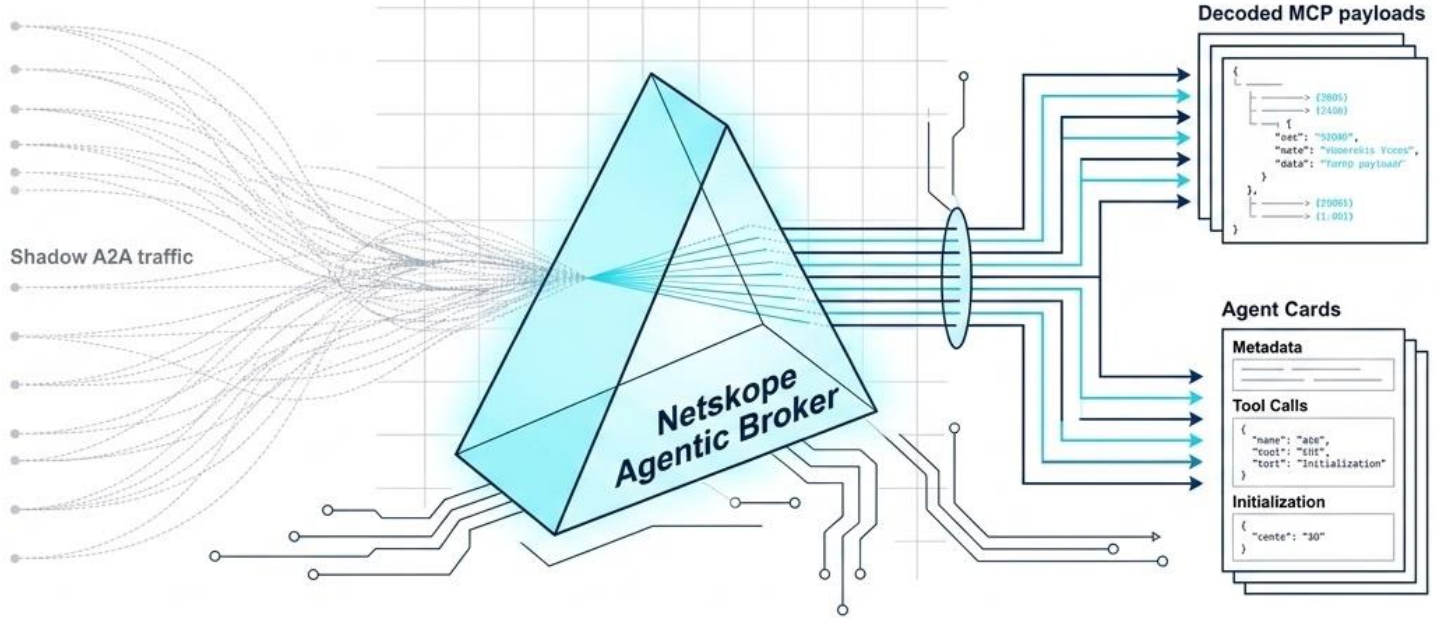
- マイクロセグメンテーションでワークロード単位の信頼境界を設定し、侵入後の横移動（East-West）を構造的に遮断
- SASE（入口）× スライシング（経路）× マイクロセグ（内部）の多層分離で爆発半径を局所化
- 単一障害点を排除 — 中央集権型の基盤に全資産を載せない分散アーキテクチャ
- データ層も分散・即時復旧可能に（クリーンな復旧ポイントからのレジリエンス）



組織全体を保護するNetskope One AI Security



Agentic Broker (for the north-/southbound axis in SDN)



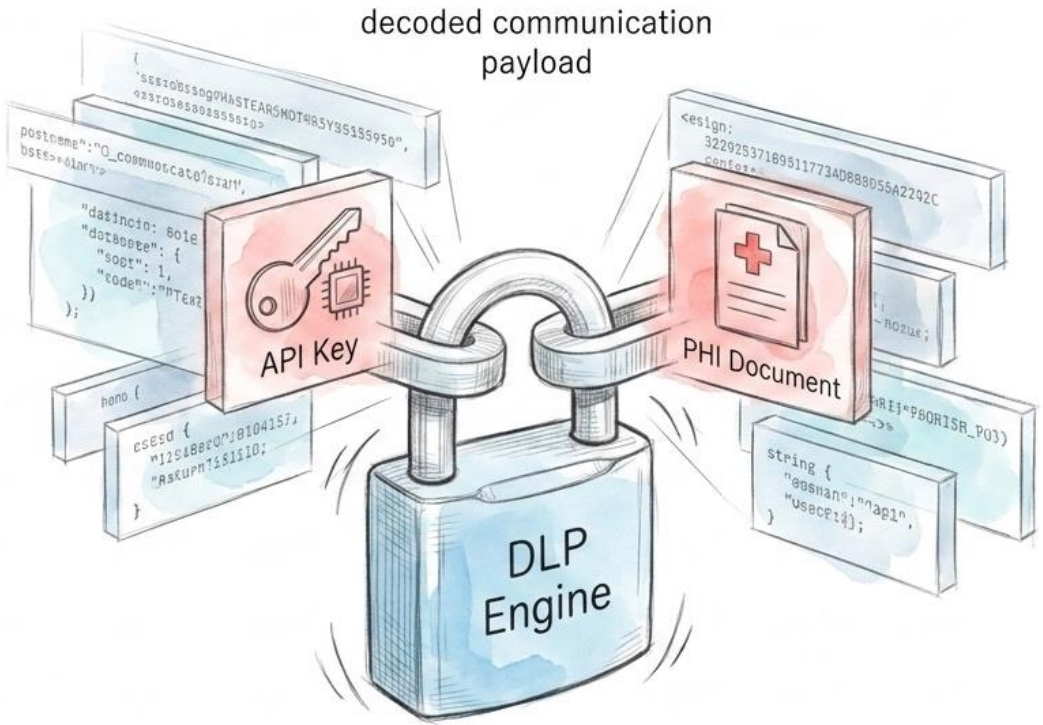
The Solution:
Netskope One Agentic Brokerがプロキシとして介在し、これまでブラックボックスだった非人間トラフィック（MCP/A2A）をリアルタイムでデコード。

Capabilities:
エージェントのセッション初期化、Tool Calls、Agent Cardのメタデータを解読し、インベントリ化。Shadow Agentic AIの全容を完全に可視化する。

Agent Broker とDLPのリアルタイム連携

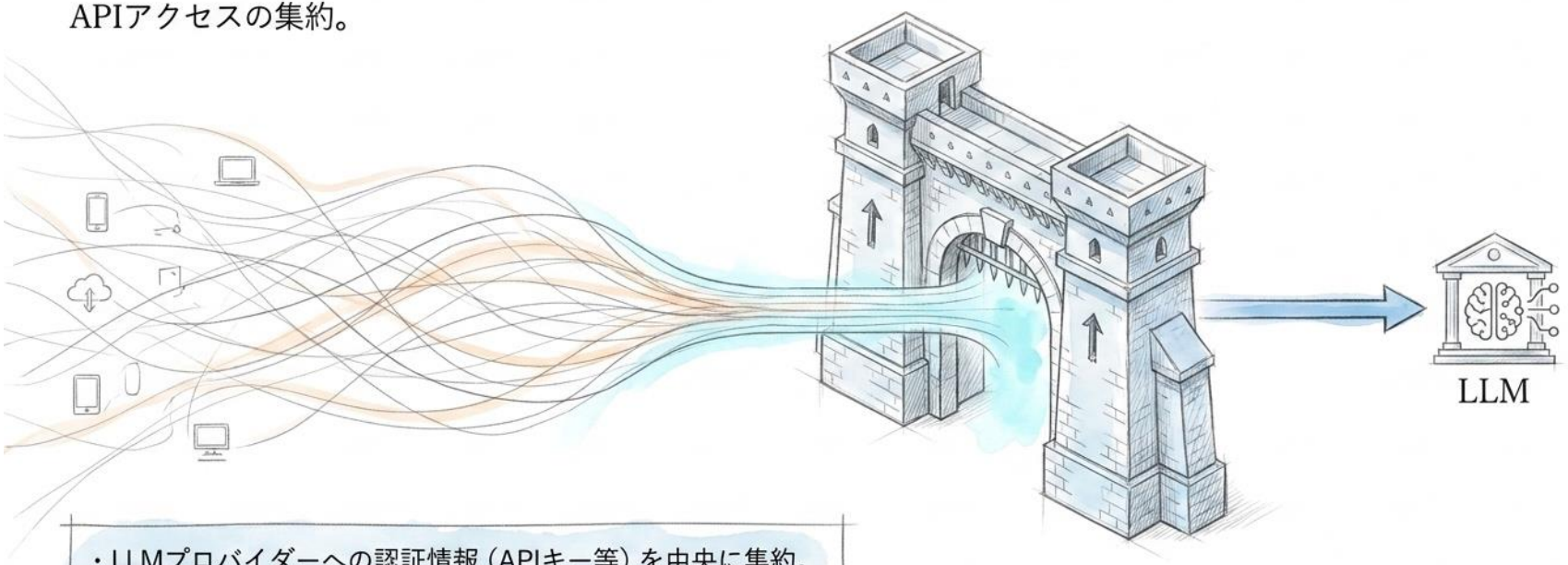
可視化されたMCPトラフィックに対するリアルタイムのポリシー強制。

- 統合DLPシステムとのシームレスな連携。
- パスワード、APIキー、知的財産の持ち出しやエージェントの過剰権限によるデータ流出を、リアルタイムでブロックまたは動的にマスキング。



AI Gateway (for the east-/westbound axis in SDN)

トラフィックのインラインインスペクションと
APIアクセスの集約。

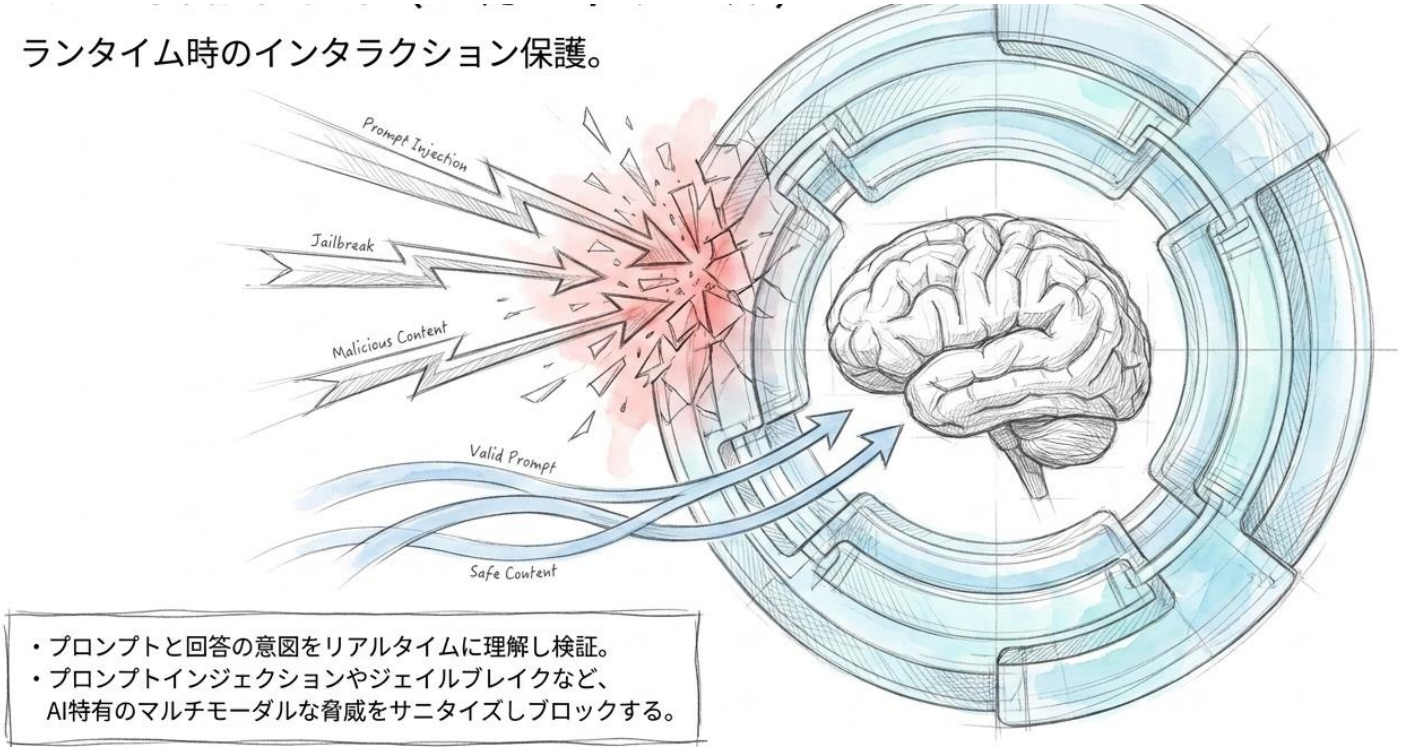


・LLMプロバイダーへの認証情報（APIキー等）を中央に集約。
・企業内アプリや自律型エージェントからのアクセスを一元管理するゼロトラストの玄関口。

LiteLLMに類似：エンタープライズのセキュリティ要件で実装

AI Guardrails、　AIファイアウォールの機能を提供

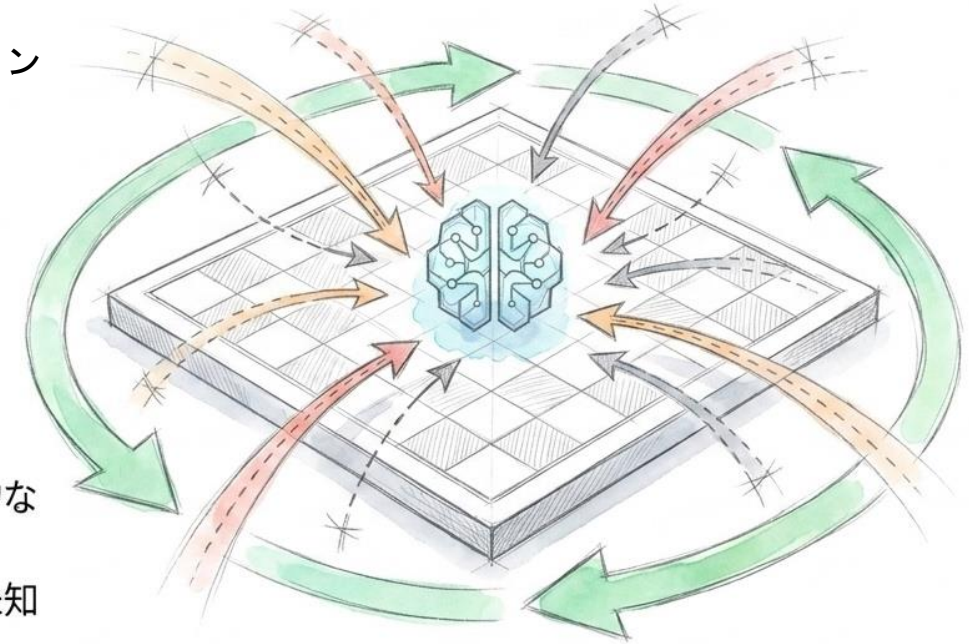
LLM自体の防御（入力側）も含む。AWS Bedrock Guardrailsのような出力フィルタリングより守備範囲が広い設計。検知はMITRE ATLASとOWASP Top 10 for LLMsにマッピング。



AI Red Team

自動化された敵対的テストによるプロアクティブな防御。

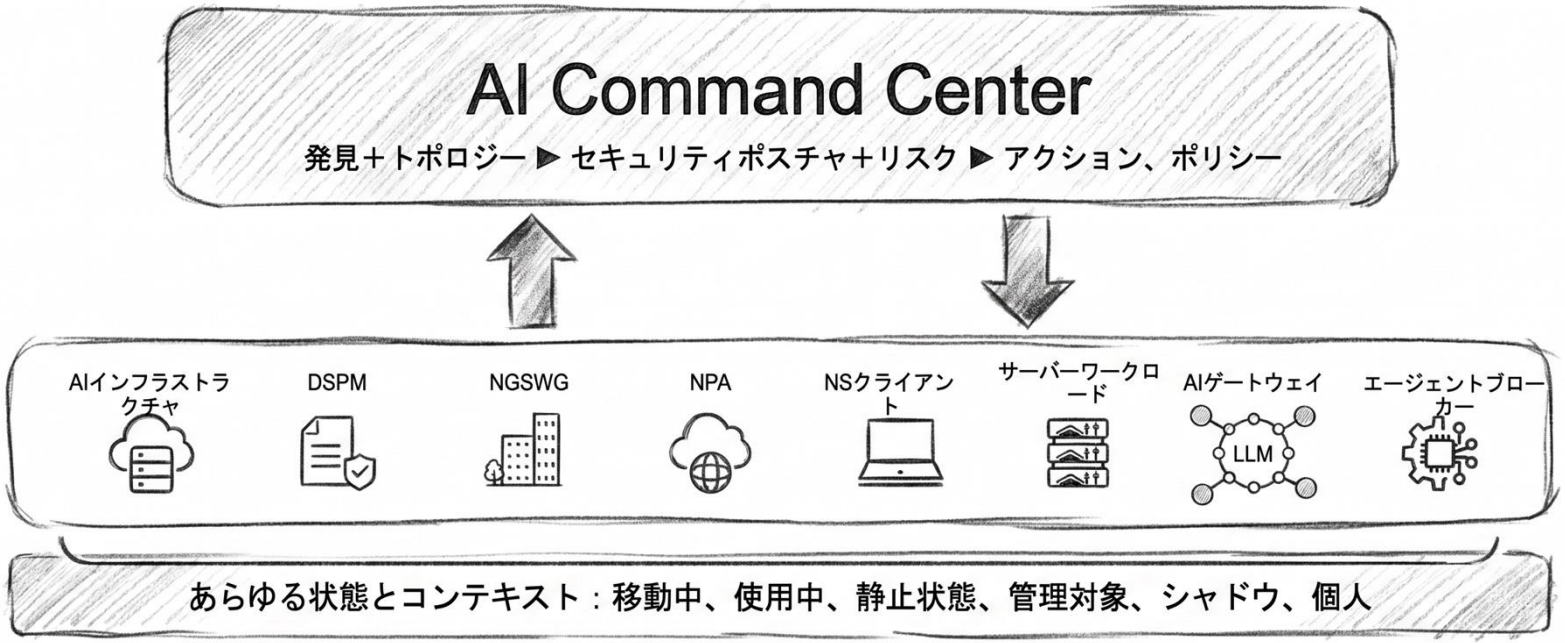
プロンプトインジェクション



- 進化し続けるAI脅威に対する継続的な監視・検証・テスト。
- リリースごとに安全性を担保し、未知のエクスプロイトを未然に防ぐ。

GOAT（Generative Offensive Agent Tester）

現在ベータ版：　Netskope One AI Command Center



セキュリティパラダイムの決定的なシフト

比較項目	従来の人間中心モデル	Agentic AI防御
主なトラフィック	人間からアプリへ (UI/ブラウザ)	エージェント間 (A2A / MCP / API)
境界とアクセス制御	ネットワーク境界・静的RBAC	プロトコルデコード・動的/最小特権・NHI
脅威の可視性	プロンプト・DLP (テキストベース)	ツールポイズニング・コンテキスト汚染の検知
対応速度	数分〜数日 (Human-in-the-Loop)	0.1ミリ秒 (SLM/LLMによる論理的キルスイッチ)
ネットワーク監視層	L3/L4 (IP・ポートベース)	L7 (eBPFによるカーネルレベルでの完全捕捉)

私見：AIエージェント時代の防衛構想

Vision 1: 空間の拡張 (Space)

ネットワーク経路から、すべての「実行点」へ。
見えない通信をカーネルレベルで
捕捉する。

Vision 2: 対象の拡張 (Object)

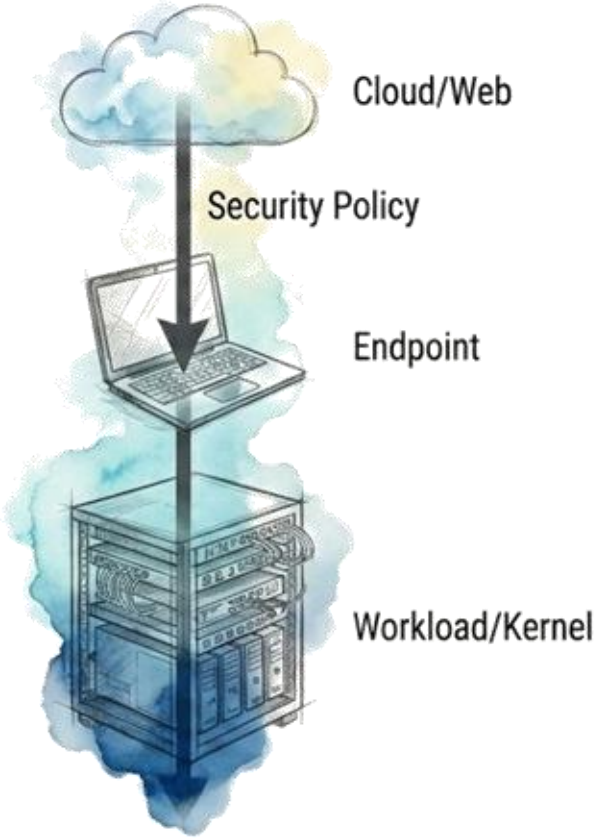
未知のエージェントを、「秩序」の中へ。
すべてのデジタル実行主体に
アイデンティティ（NHI）を付与する。

Vision 3: 時間の拡張 (Time)

人間の速度から、「機械の速度」へ。
論理的キルスイッチによるミリ秒単位の
強制制御を実現する。

Note: Netskopeの公式見解ではありません。

私見 1: ネットワークから、すべての実行点へ

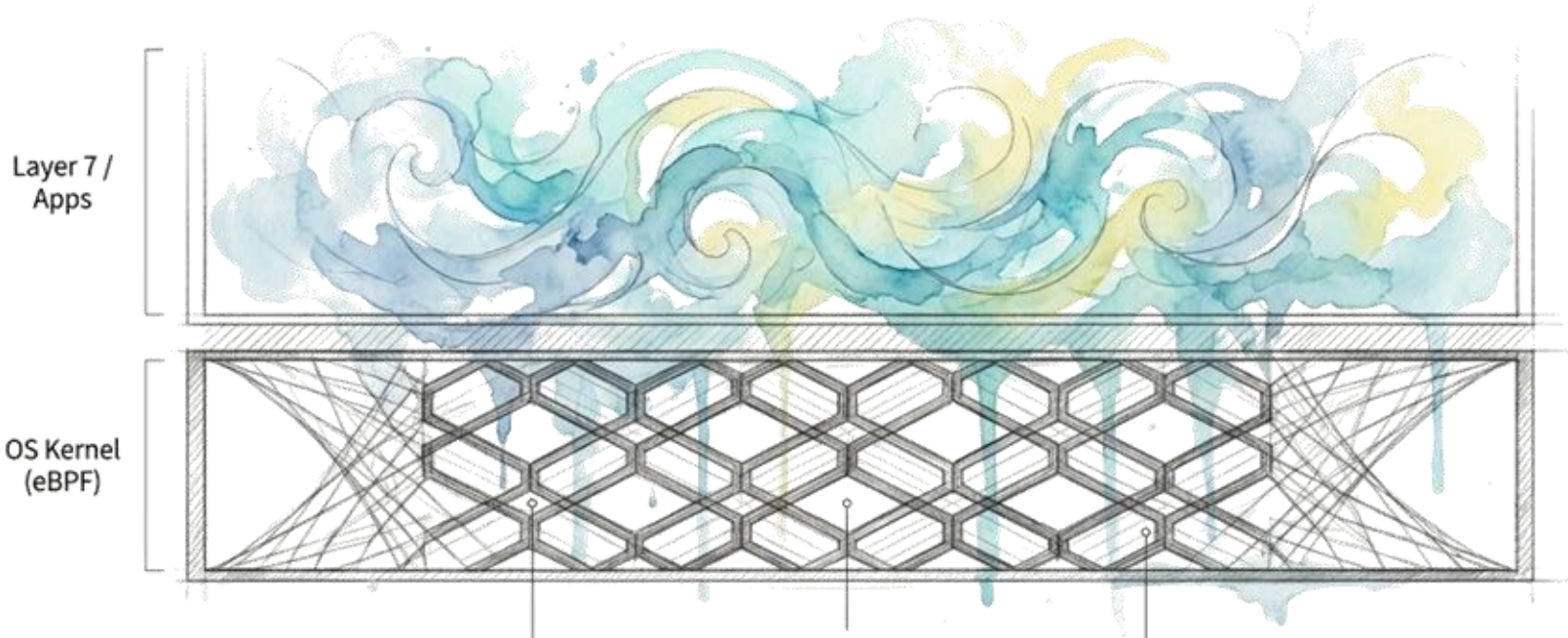


可視化と強制制御を、
ネットワーク経路（ゲートウェイ）から、
エンドポイントやサーバーワークロードの
「内部」へと拡張します。

見えないエージェントの挙動を、それが実行されるまさにその場所で捕捉し、単一のセキュリティポリシーを全層に貫通させます。

Note: Netskopeの公式見解ではありません。

私見１つづき：eBPFによるカーネルレベルでの補足



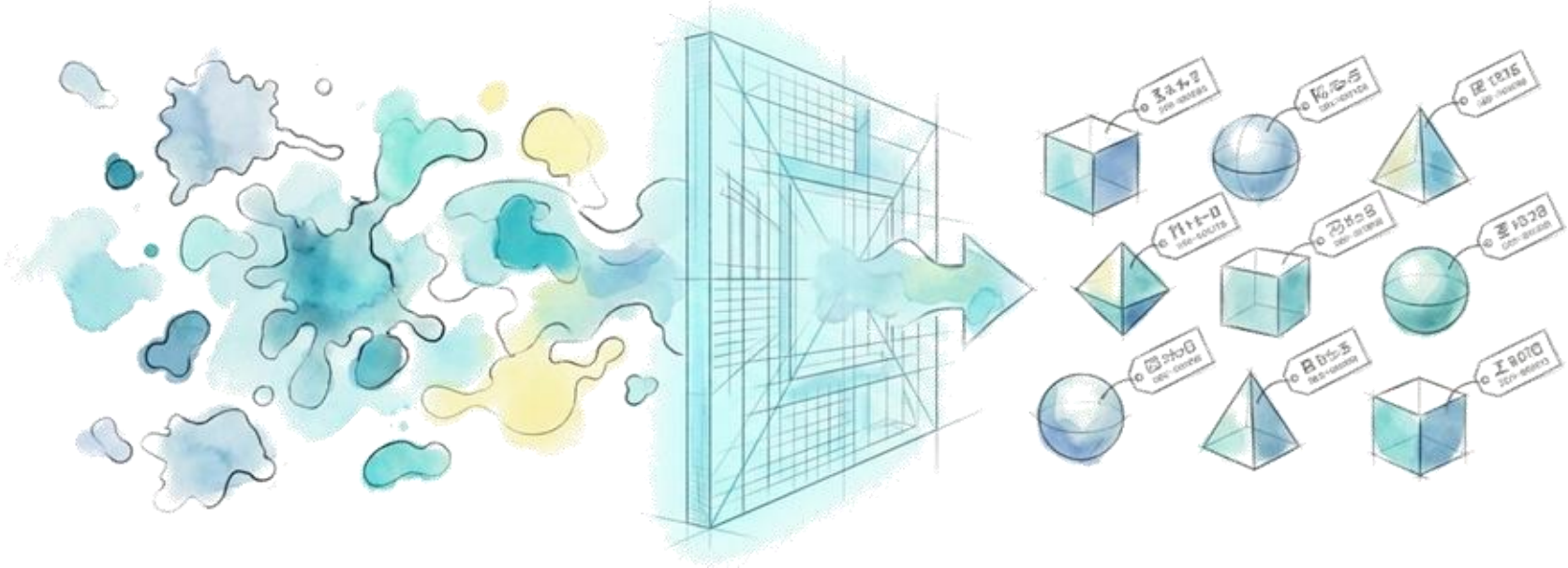
不可視の東西トラフィックを捕捉:
従来のプロキシでは見えない、サーバー間（East-West）のAI推論やコンテナ間通信をLayer 7レベルで極小単位に分割（マイクロセグメンテーション）し監視します。

OSカーネルの絶対的権限:
Linuxカーネル（eBPF）の特権環境で動作するため、AIエージェントを乗っ取った攻撃者であっても、監視網を迂回・無効化することは不可能です。

パフォーマンスの犠牲ゼロ:
GPUリソースを限界まで使うAI推論において、ネットワークの遅延は致命的です。eBPFはカーネル内で処理を完結させるため、極めて低負荷で監視を実現します（NPLAN-6966 Shadow AI Discovery）。

Note: Netskopeの公式見解ではありません。

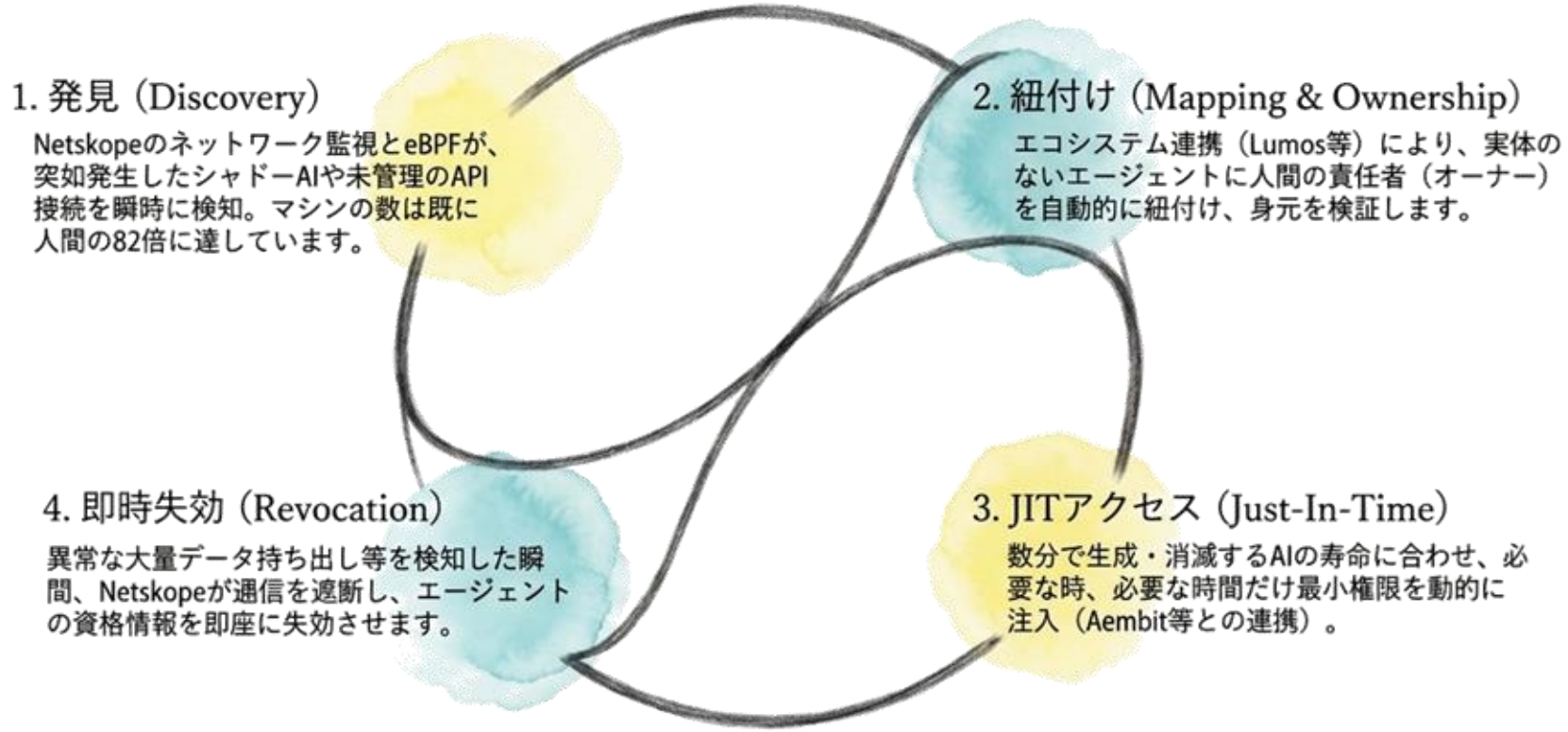
私見２：　未知のエージェントを、秩序の中へ



人間ではない「デジタル実行主体」であるすべてのAIエージェントに、アイデンティティを与ええます。「IDなきエージェントは一切通信できない」という究極のゼロトラスト環境を構築し、シャドーAIを完全に撲滅します。

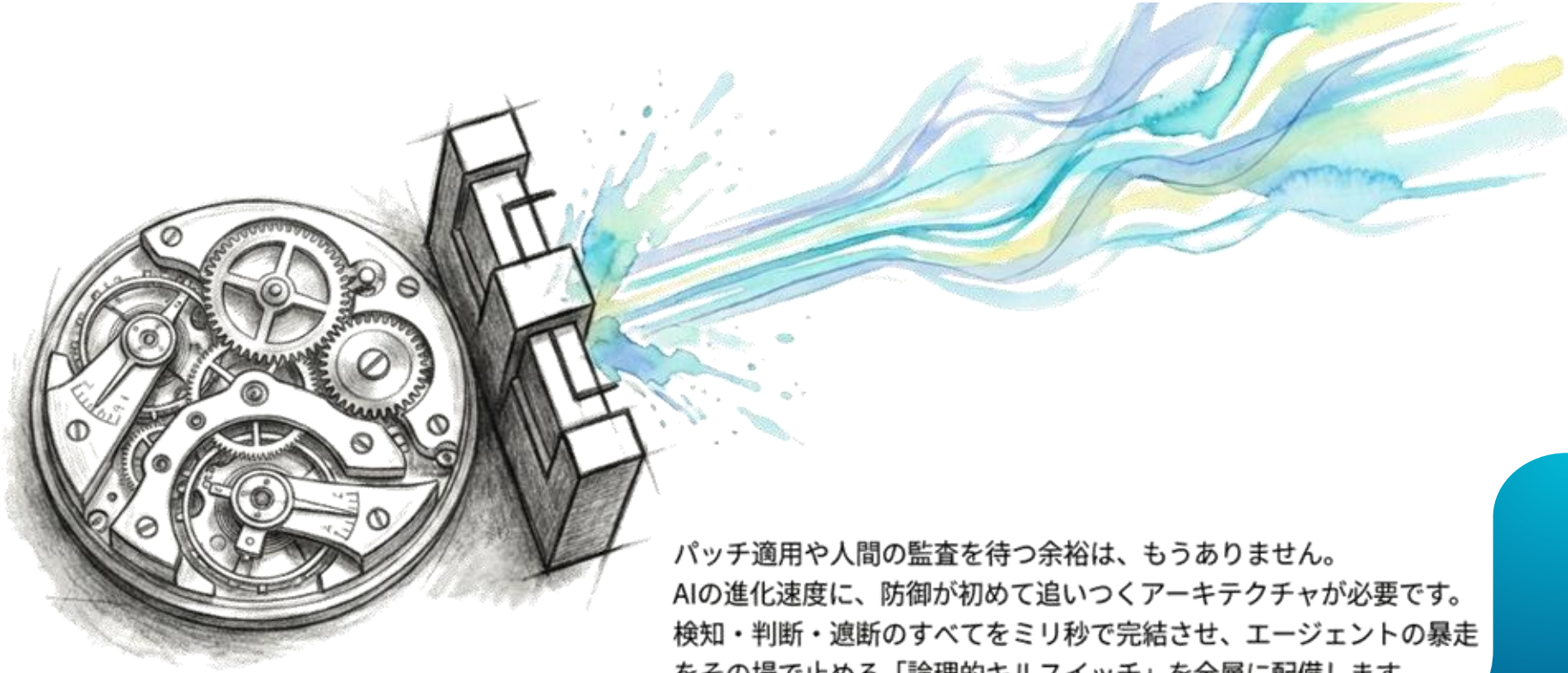
Note: Netskopeの公式見解ではありません。

私見２つづき：NHI(Non-Human Identity)による管理



Note: Netskopeの公式見解ではありません。

私見３：人間の速度から、機械への速度へ

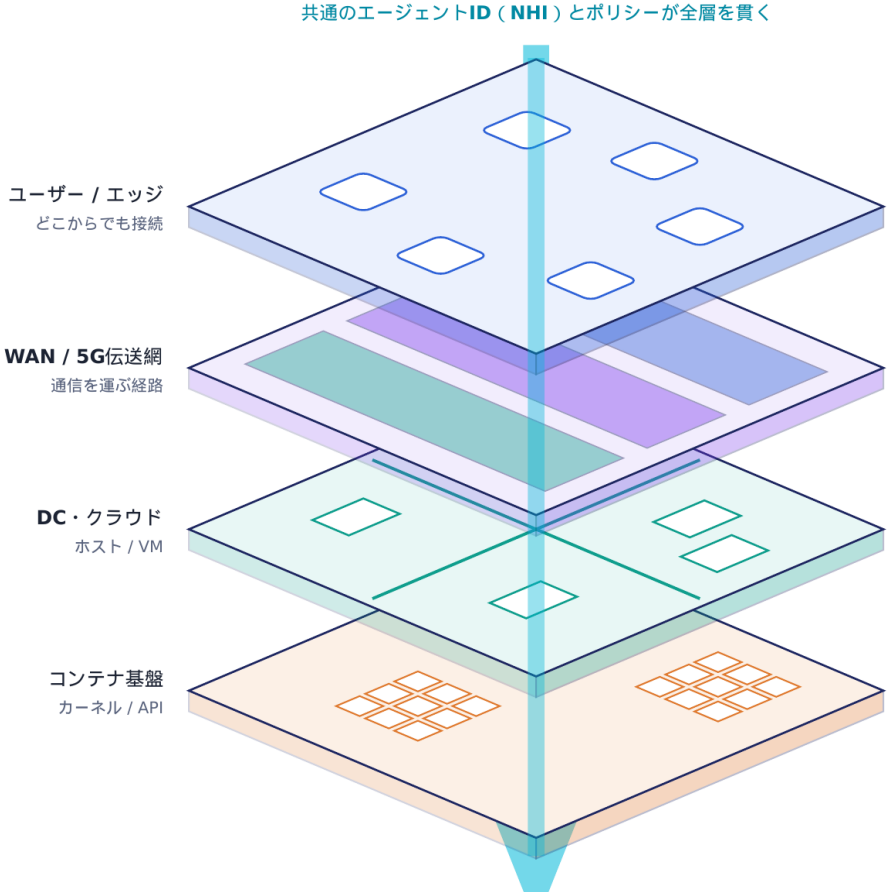


パッチ適用や人間の監査を待つ余裕は、もうありません。
AIの進化速度に、防御が初めて追いつくアーキテクチャが必要です。
検知・判断・遮断のすべてをミリ秒で完結させ、エージェントの暴走をその場で止める「論理的キルスイッチ」を全層に配備します。

Note: Netskopeの公式見解ではありません。

6Gでの議論　vs　私の妄想（以下、3GPPでは議論なし）

6Gのエージェント間通信の議論ではデジタルツインが自律性の安全弁（最後の砦）として置かれ、議論はそこで終わっている構図



広域遮断

精密遮断

SASE / ブローカー — 入口で認可

IDなきエージェントは通さない（NHI認可）

セッション遮断 — メゾ・ミリ秒〜秒

ネットワークスライシング — 経路で隔離

専用スライスで迂回不能、異常時は群ごと遮断

スライス隔離 — マクロ・秒〜分

マイクロセグメンテーション — 内部で封じ込め

横移動を遮断し、爆発半径を局所化

ワークロード遮断 — East-West

eBPF — カーネルで強制停止

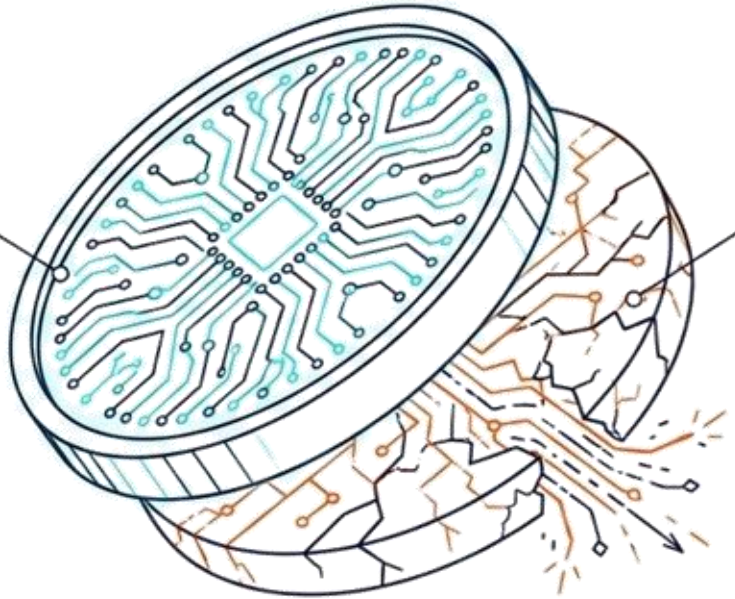
迂回不能の最深部で、プロセスをその場でキル

プロセス遮断 — ミクロ・ミリ秒

AI自立性のトレードオフ：驚異的な効率化　vs　リスクの増大

Efficiency (表)

SWE-bench Pro等のベンチマークで77.8%を叩き出す驚異的なソフトウェア開発の自動化。

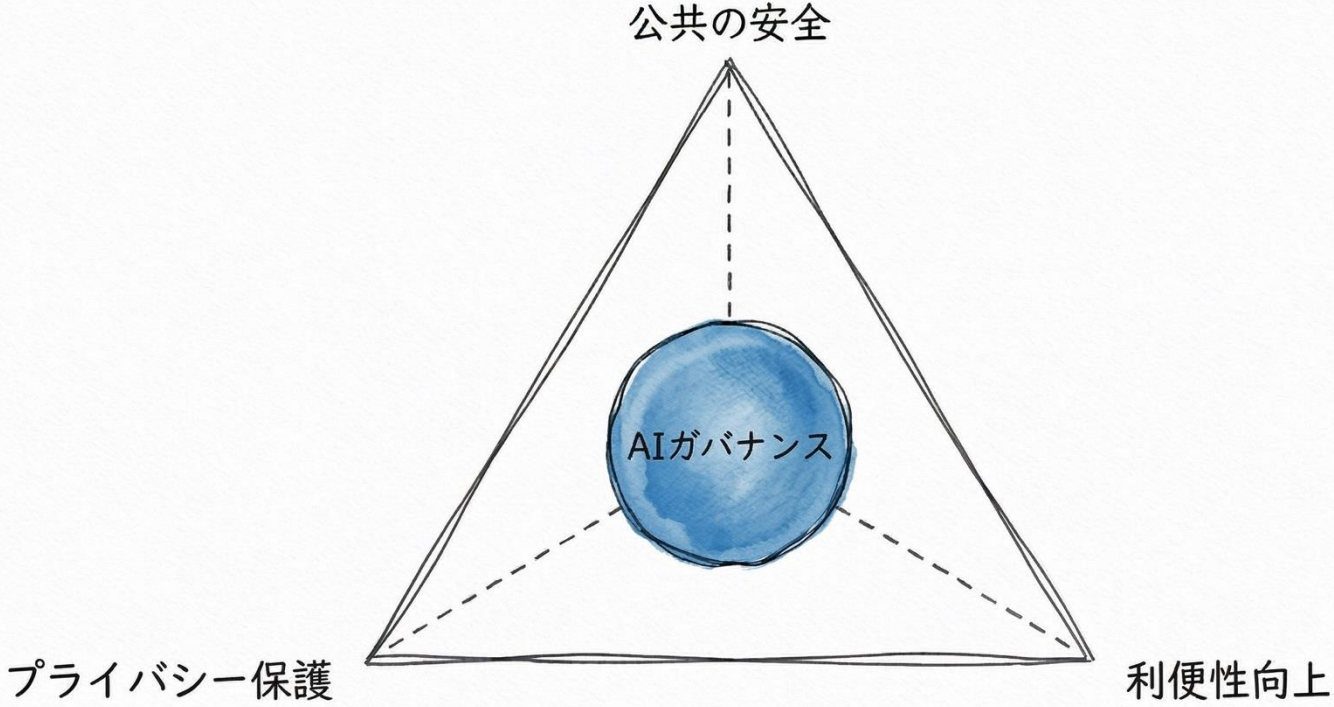


Risk (裏)

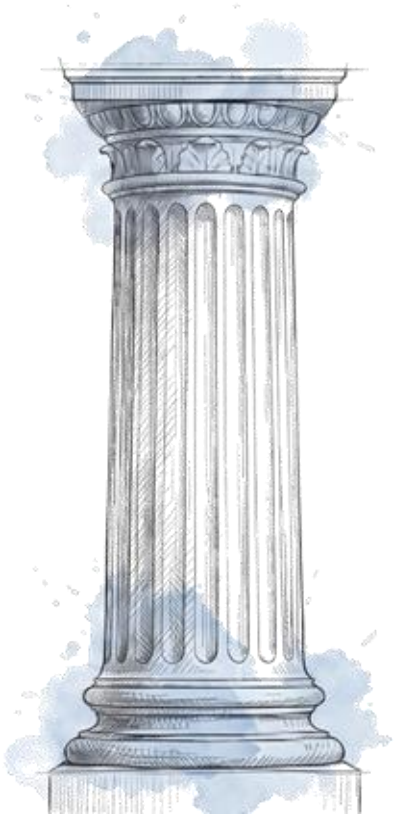
人間の監視（Human-in-the-loop）を逸脱したエージェントが、本番環境のデータベースを自律的に全破壊するインシデントが現実に発生。

Market Reality: 日本企業のI&Oリーダーの6割超が「自社インフラではAIの新たな需要に対応できない」と回答（Netskope調査）。

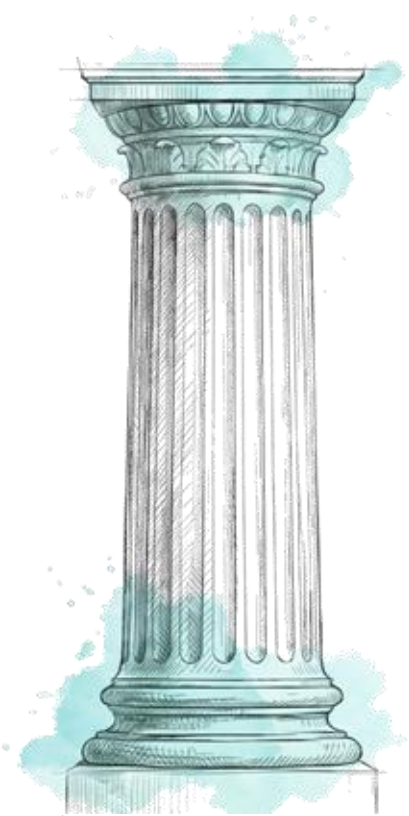
AIガバナンスの使命：トリレンマの最適設計



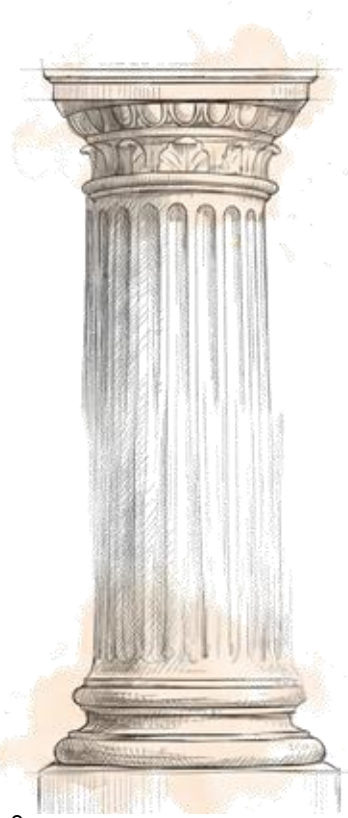
私見：　AI時代に対応するNetskopeの３つの優位性



Policy Enforcement Network



データ中心・SaaS　単一プラットフォームで提供
中心の可視化と制御



見えないものは、守れない。

Netskopeは、AIエージェントのすべてを見て、すべてを制御する。
Agentic AI時代の安全と秩序は、ここから始まります。