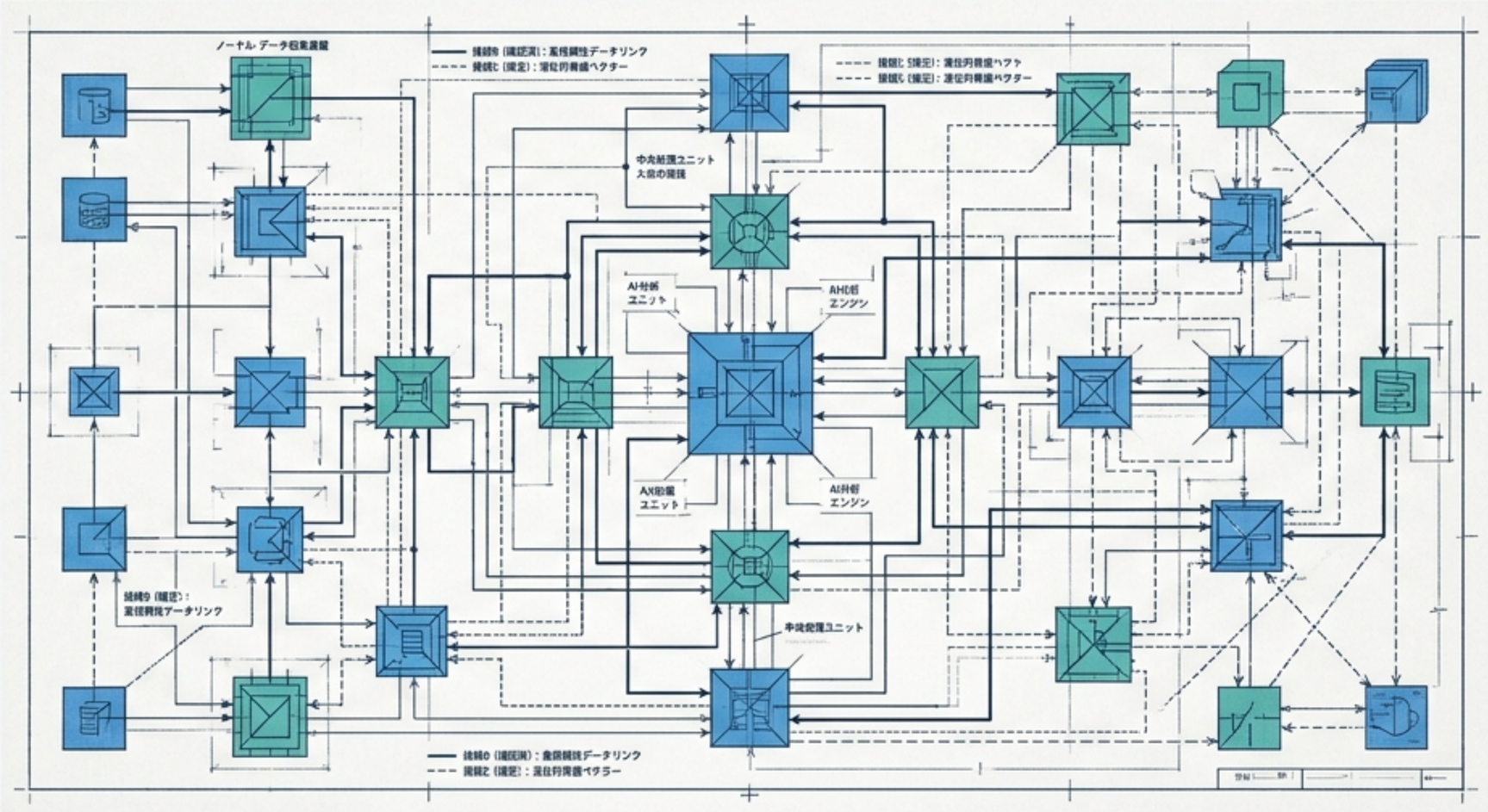
防衛省 × Sakana AI：戦略インテリジェンスの新たな分析基盤

3.29億円の「総合分析業務AI実証」に隠されたファクトと技術的メカニズムの解剖



3.29億円の契約が意味する「実証」の輪郭

Who（誰が）

防衛省情報本部 × Sakana AI

How Much（いくらで）

3億2,903万2,000円

（防衛装備庁公表）

What（何を）

総合分析業務

戦略レベルにおけるAI機能の調査・実証

When（いつまでに）

2027年度開始前

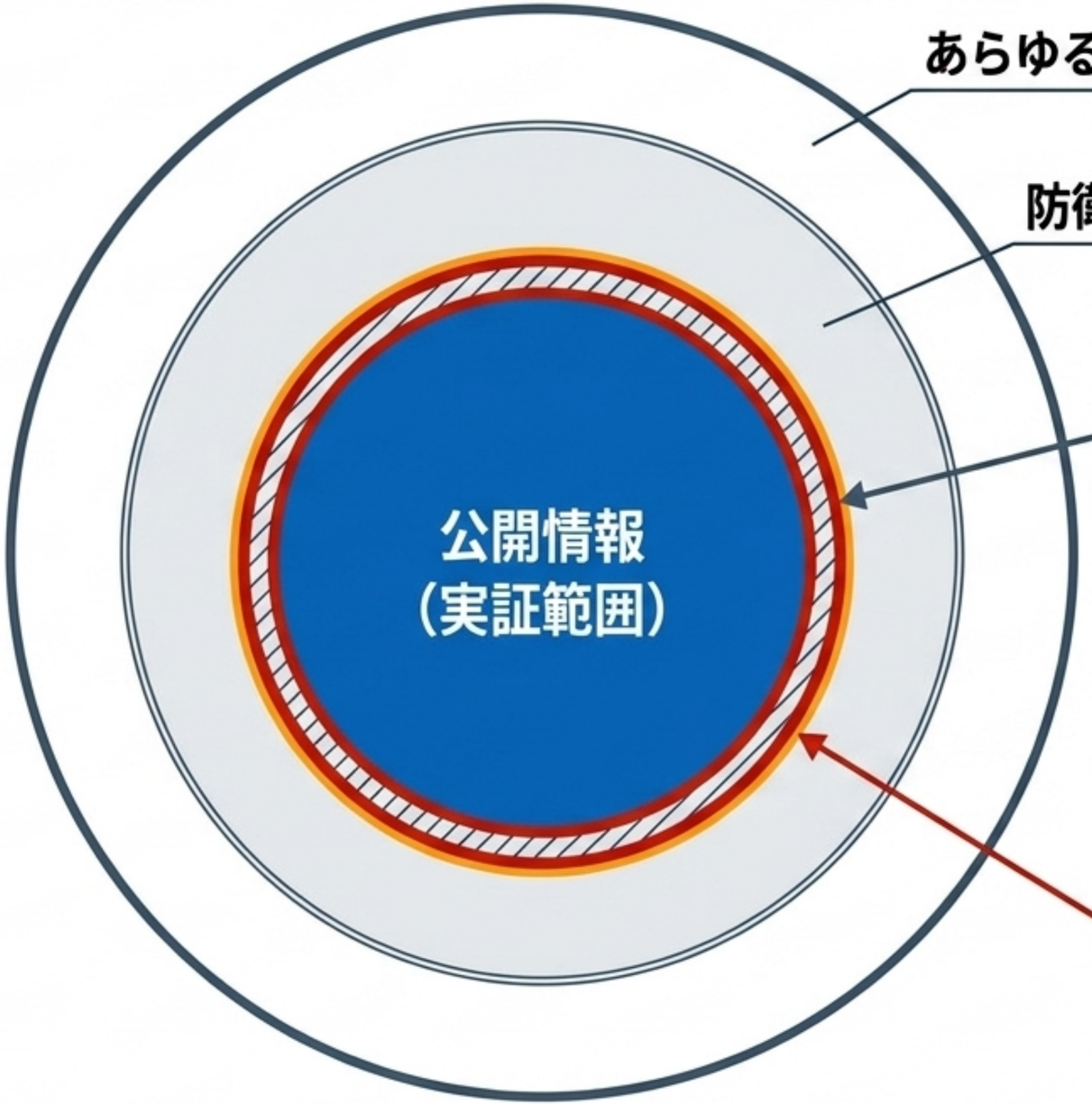
（3月末）までに官側試用、2027年度末に成果報告

本件は「兵器の自動制御」でも「特定秘密の処理」でもない。
情報分析官の業務を支援する「意思決定支援系」の実証である。

報道のノイズと一次資料が示すシグナル

Myth（世間のイメージ）	Reality（一次資料のファクト）
✕ AIが特定秘密や機密情報を取り扱う	✓ AIの入力対象は「公開情報（有料・無料）」に明示的に限定
✕ AIが直接、兵器やドローンを自動制御する	✓ 国家レベルの安全保障を支える「戦略インテリジェンス支援」のみ
✕ Deep Researchという新製品がそのまま納入された	✓ 「Deep Research」はLedge.aiの見出しであり、正式な調達仕様ではない

厳密に定義された「情報境界（Information Boundary）」



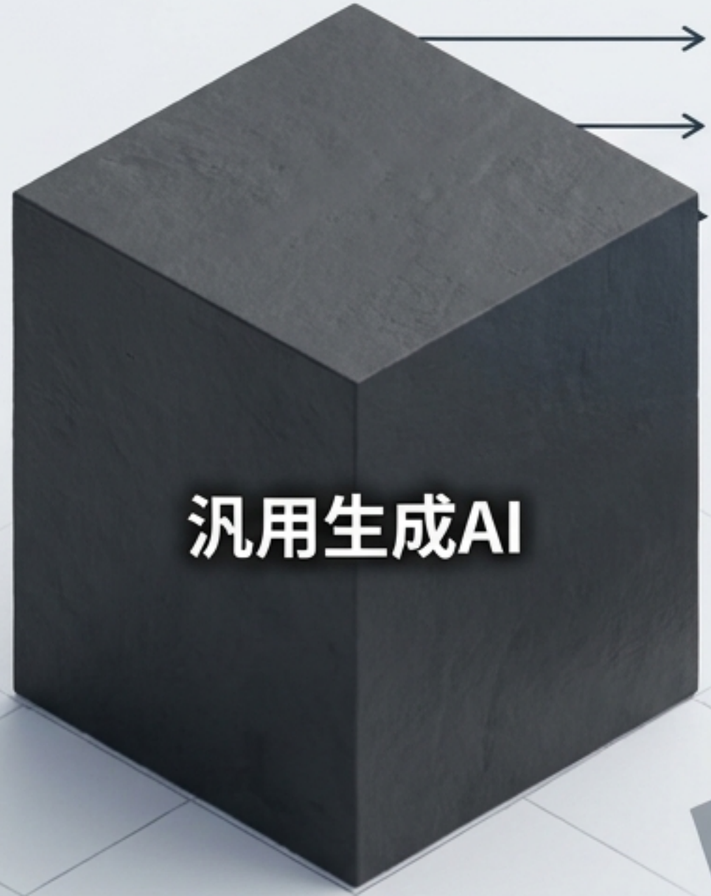
あらゆる情報源（電波、画像・地理など）

防衛省の戦略インテリジェンス領域

「有料・無料を問わない公開情報」のみが電子的なAI入力の対象。

調達説明書は、「保護すべき情報」の取扱いを全契約期間において「書面のみ」と明記。AIクラウド上での機密解析は現時点で否定される。

なぜ汎用生成AI（ChatGPT型）では不十分なのか？



汎用生成AI

- 情報の確度・鮮度の欠如
- ハルシネーション（幻覚）
- データバイアス

防衛省の要求：スピード、精度、効率、ヒューマンエラー削減、情報資産管理

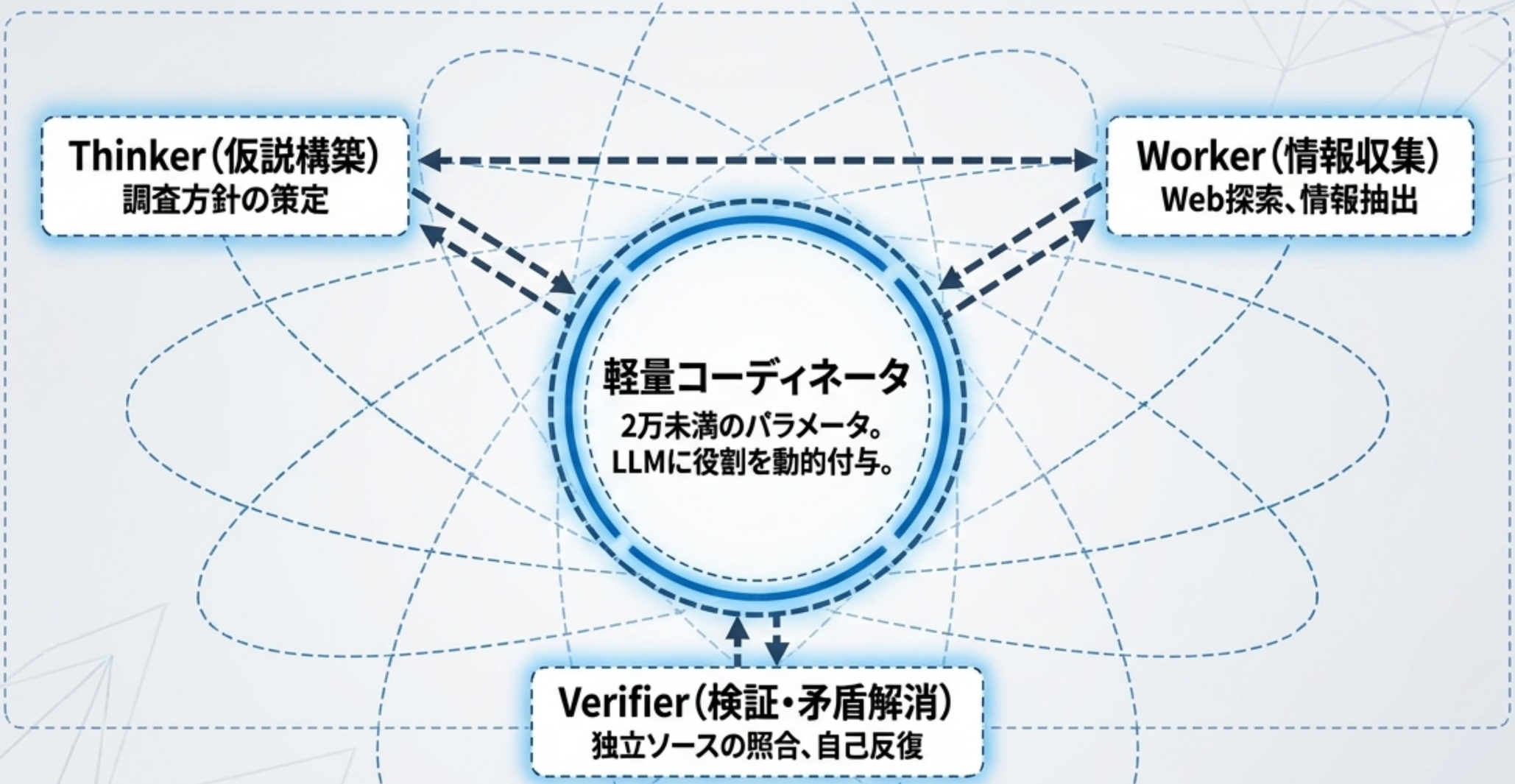
求められる優位性

- 計算コストを抑えながら精度を維持するアーキテクチャ
- 情報の出所と判断プロセスが追跡可能なシステム

- 計算コストを抑えながら精度を維持するアーキテクチャ
- 情報の出所と判断プロセスが追跡可能なシステム

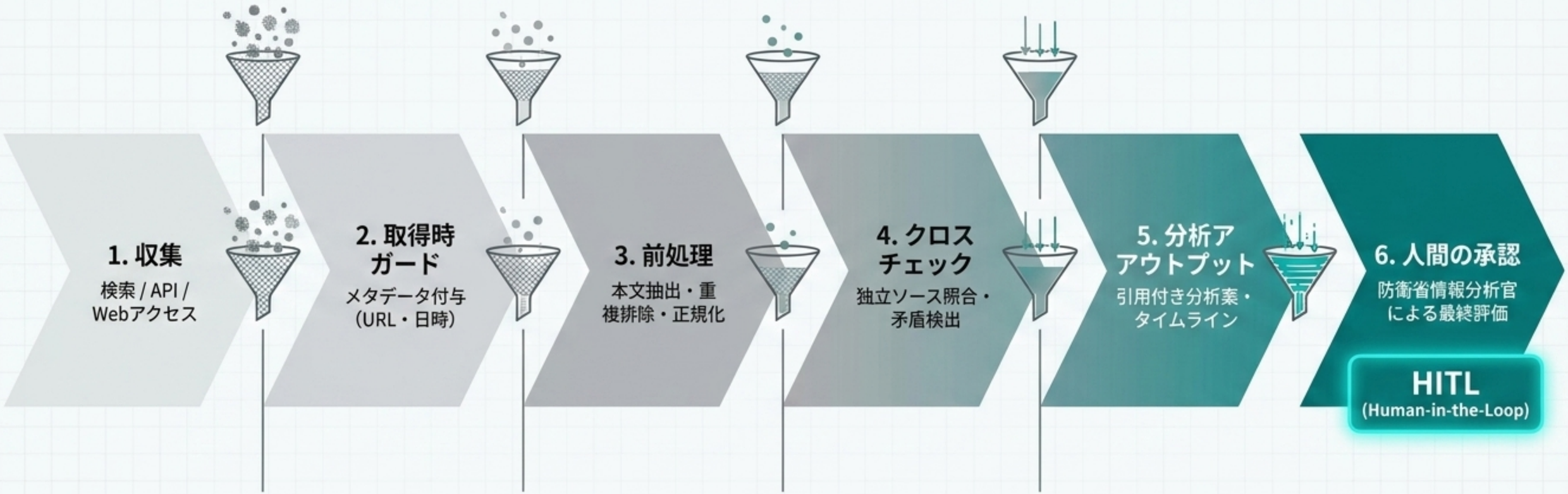
推定アーキテクチャ：マルチエージェントの動的協調

※破線（ダッシュライン）は公開技術（OSINT CTF等）からの推定であり、確定仕様ではない。



巨大な単一モデルに頼るのではなく、難問だけ協調回数を増やす設計で
「低コストとハルシネーション局限」を両立する。

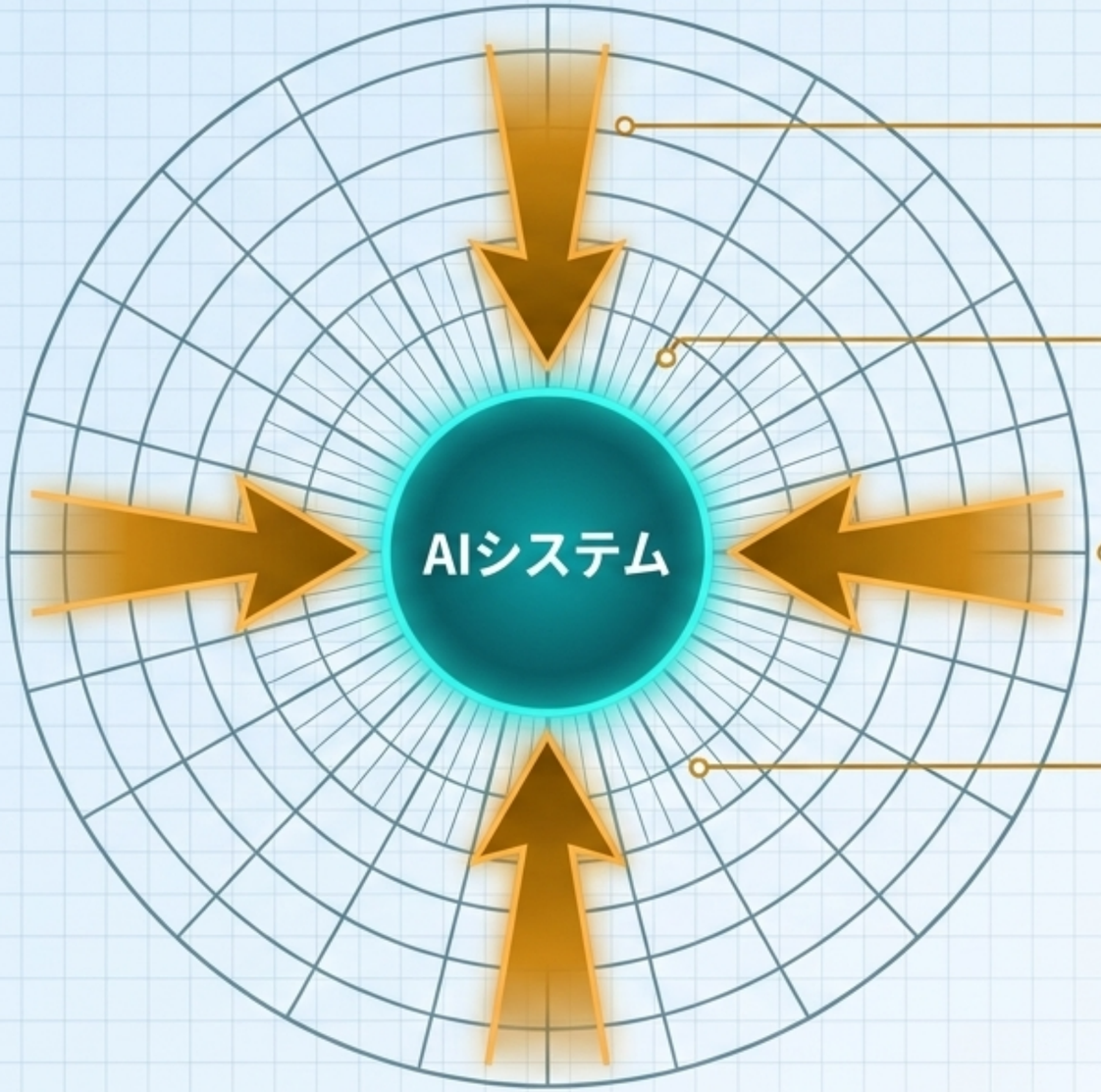
自動化されるインテリジェンス・ワークフロー



確定事項と未公開仕様の診断マトリクス

カテゴリ	ステータス	解説
実証入力	Public	有料・無料を問わない公開情報
官側試用時期	Public	2027年度開始前（当初計画）
具体的LLM / API	Unknown	Fuguの利用、ベンダー、データ所在地は未確認
RAG / 検索基盤	Unknown	ベクトルDB、Embedding製品仕様は不明
インフラ構成	Unknown	クラウドかオンプレミスか、GPU種別は未公開
ログ・データ保持	Unknown	保存期間、削除・バックアップ規則は未公開

新たな脅威モデル：敵対的OSINT環境のリスク



偽情報工作: 多数のサイトに同じ虚偽を流し独立ソースと誤認させる。

プロンプト・インジェクション: Web内の悪意ある指示を命令と誤認。

API流出・推測: 検索クエリから防衛省の関心事が推測されるリスク。

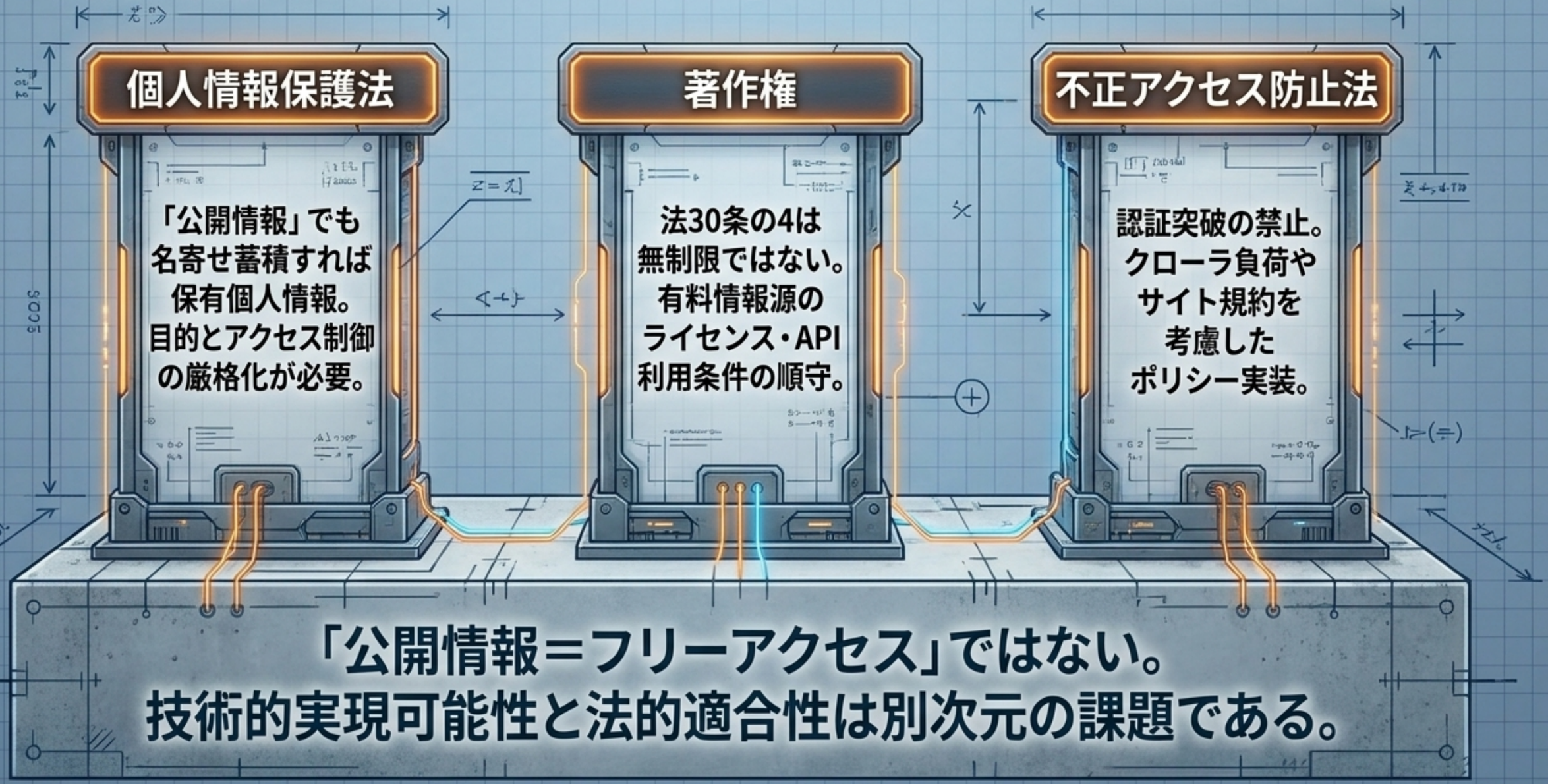
自動化バイアス: 流暢な出力を人間が無批判に信じ込む心理的リスク。

INSIGHT

エージェントが多数の検索を行うほど、低品質情報や悪意ある入力に触れる「表面積」が拡大する。

リスクを封じ込めるガードレール設計

公開情報の収集に立ちはだかる「3つの法的境界」



「公開情報＝フリーアクセス」ではない。
技術的実現可能性と法的適合性は別次元の課題である。

「正解率」を超える実務向け評価マトリクス

引用忠実度
(Citation Precision)

引用先が本当に主張を裏付けているか。

独立ソース比率

まとめサイトの量産を「多数の裏付け」と誤認していないか。

矛盾検出
(Contradiction Recall)

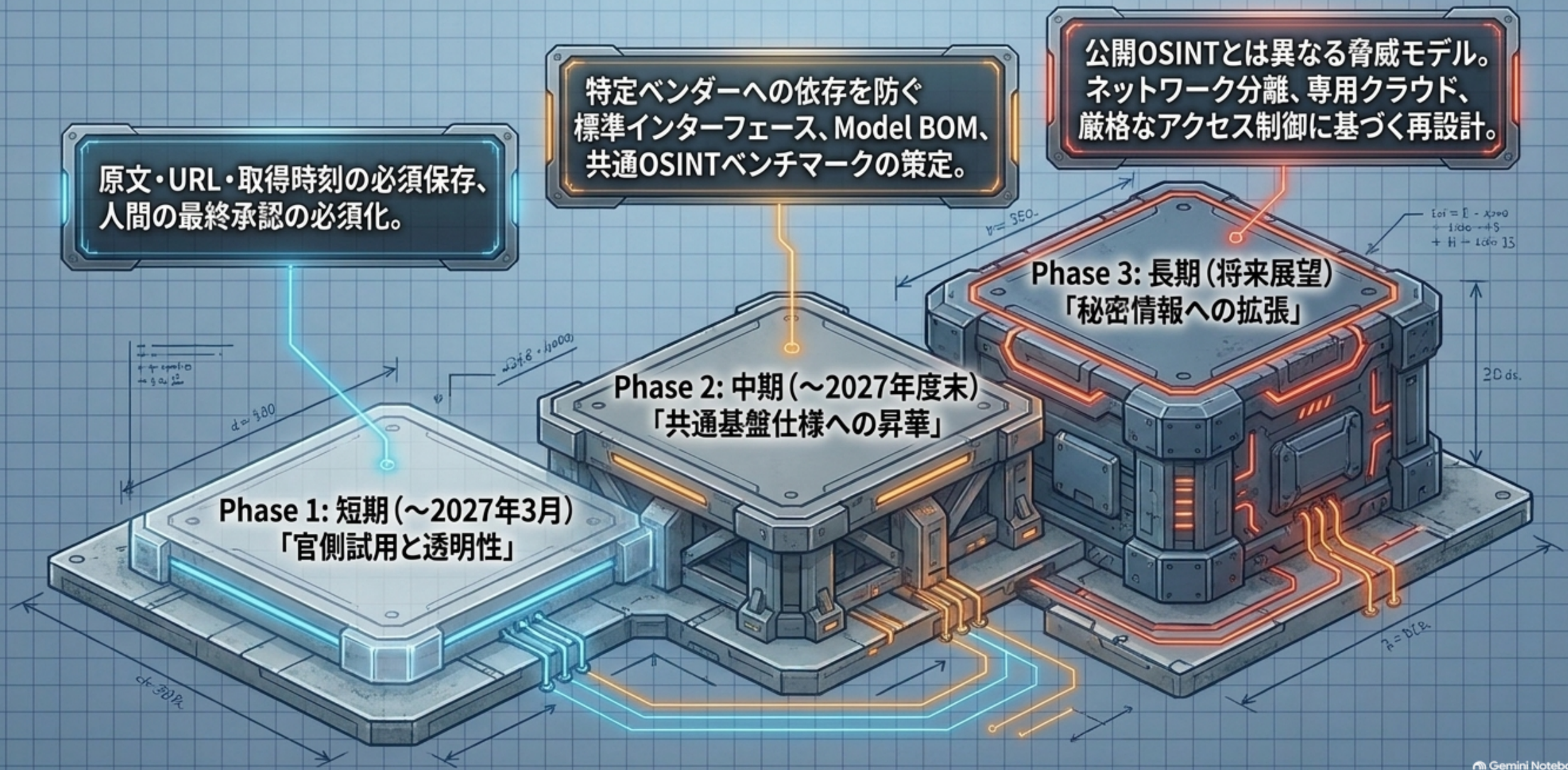
相反する情報を見落とさずに提示できているか。

人間の訂正率
(Override Rate)

最終的に人間がどれだけ出力を覆したか。

3つのLLMが同意したことを「3つの独立証拠」と数えてはならない。
モデル投票数ではなく「外部証拠の独立性」が鍵となる。

防衛AIインフラの共通基盤化へのロードマップ



「答えを出す機械」から「検証可能な共同分析者」へ

AI = Oracle
（神託機械）

ブラックボックスからの単一の答えに依存

AI（情報の探索・
構造化・反証提示）

人間（批判的検証
と意思決定）

AI + Human = Strategic Advantage

本実証の真の成功基準は「AIが賢かったか」ではない。「分析官が誤情報をより早く排除し、判断根拠をより明確に説明できるようになったか」である。