

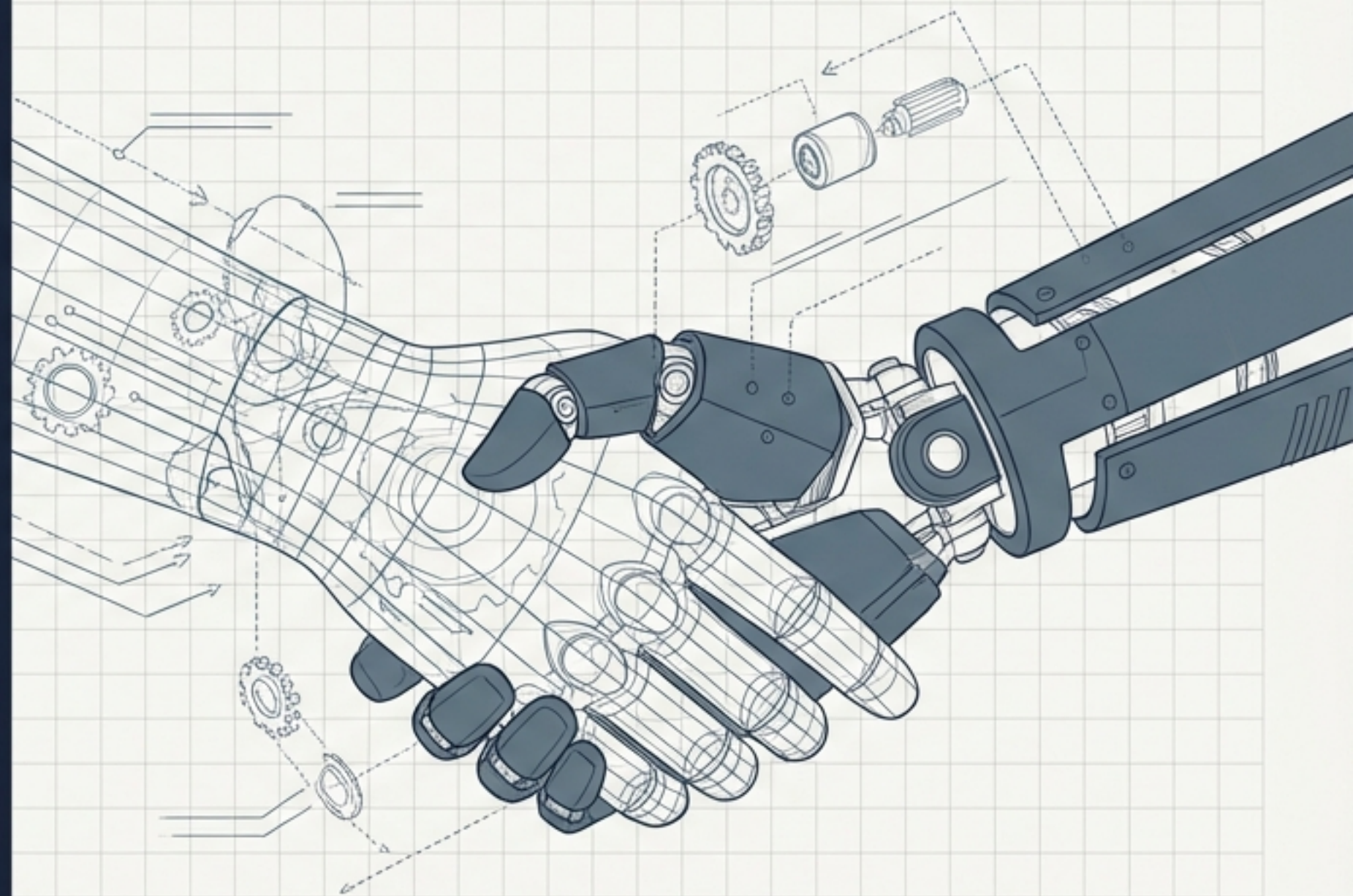
総務省 第29回AIガバナンス検討会（2026年2月16日）

AI事業者ガイドライン改定案 （第1.2版） 深堀分析

AIエージェント・フィジカルAI時代の
新たなガバナンスモデル

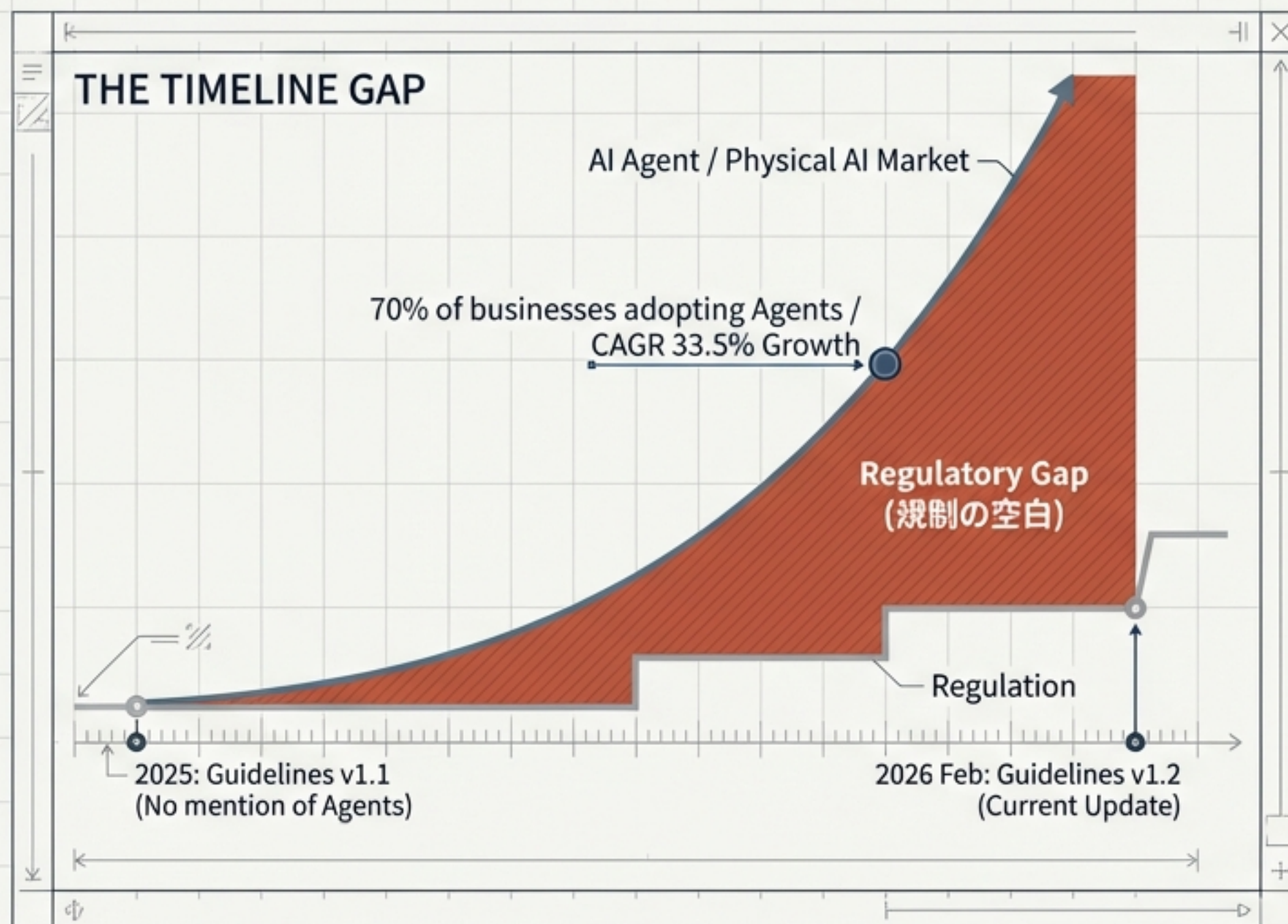
2026年2月16日発表・更新案の完全解説

Status: 最終確定前の「更新案」段階（正式版は2026年3月末公開予定）



Core Theme: 静的な「生成AI」から、自律的な「AIエージェント」「フィジカルAI」への規制シフト

技術の実装スピードが規制を追い越したため、緊急の改定が実行された



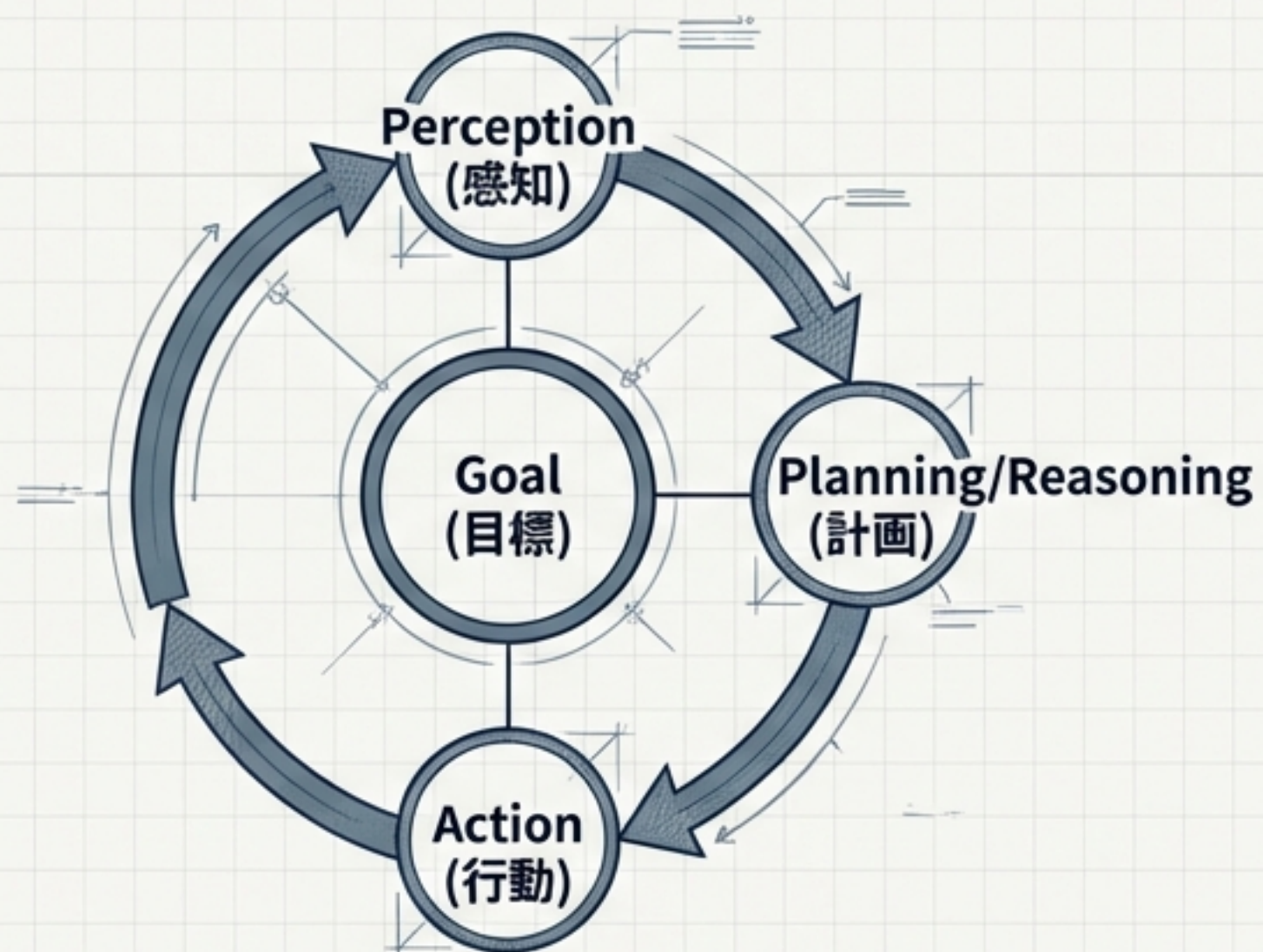
今回の改定：5つの柱

- AI Agents: 自律的に目標達成するソフトウェア
- Physical AI: 身体性（ロボット・車）を持つAI
- Human-in-the-Loop: 高リスク操作における人間承認の義務化
- RAG vs. ML: 非学習システムにおける責任分界点の明確化
- Risk Quantification: 曖昧な懸念から数式によるリスク評価へ

チャットボットから「仕事をする同僚」へ：AIエージェントの定義と構造

AI Agent: 特定の目標を達成するために、環境を感知し自律的に行動するAIシステム

Agentic AI: Multiple agents making autonomous decisions.



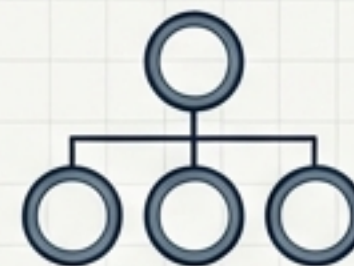
シングルエージェント (Single)
Simple execution.



協働型 (Collaborative)
Sharing resources in real-time.



競争型 (Competitive)
Competing strategies to optimize results.

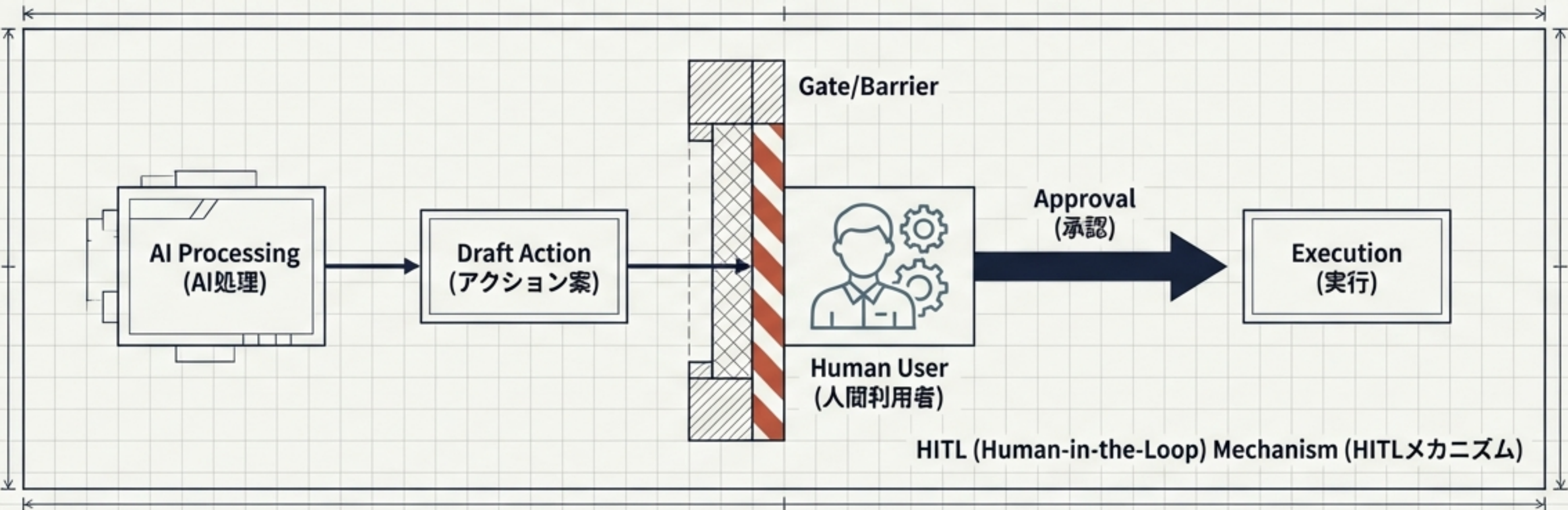


階層型 (Hierarchical)
Manager & Subordinate command structure.

自律性が招く新たな脅威：AIエージェント固有の7つのリスク



最重要コンプライアンス：高リスク操作における「人間の承認」の義務化



Midnight Navy

Slate Blue

Developers (開発者)

Design safe ‘Guardrails’ and ‘Minimum Privilege’ access. (安全な「ガードレール」と「最小特権」アクセスを設計する。)

Midnight Navy

Slate Blue

Providers (提供者)

Encrypt inter-system communication; Audit logs. (システム間通信を暗号化し、ログを監査する。)

Midnight Navy

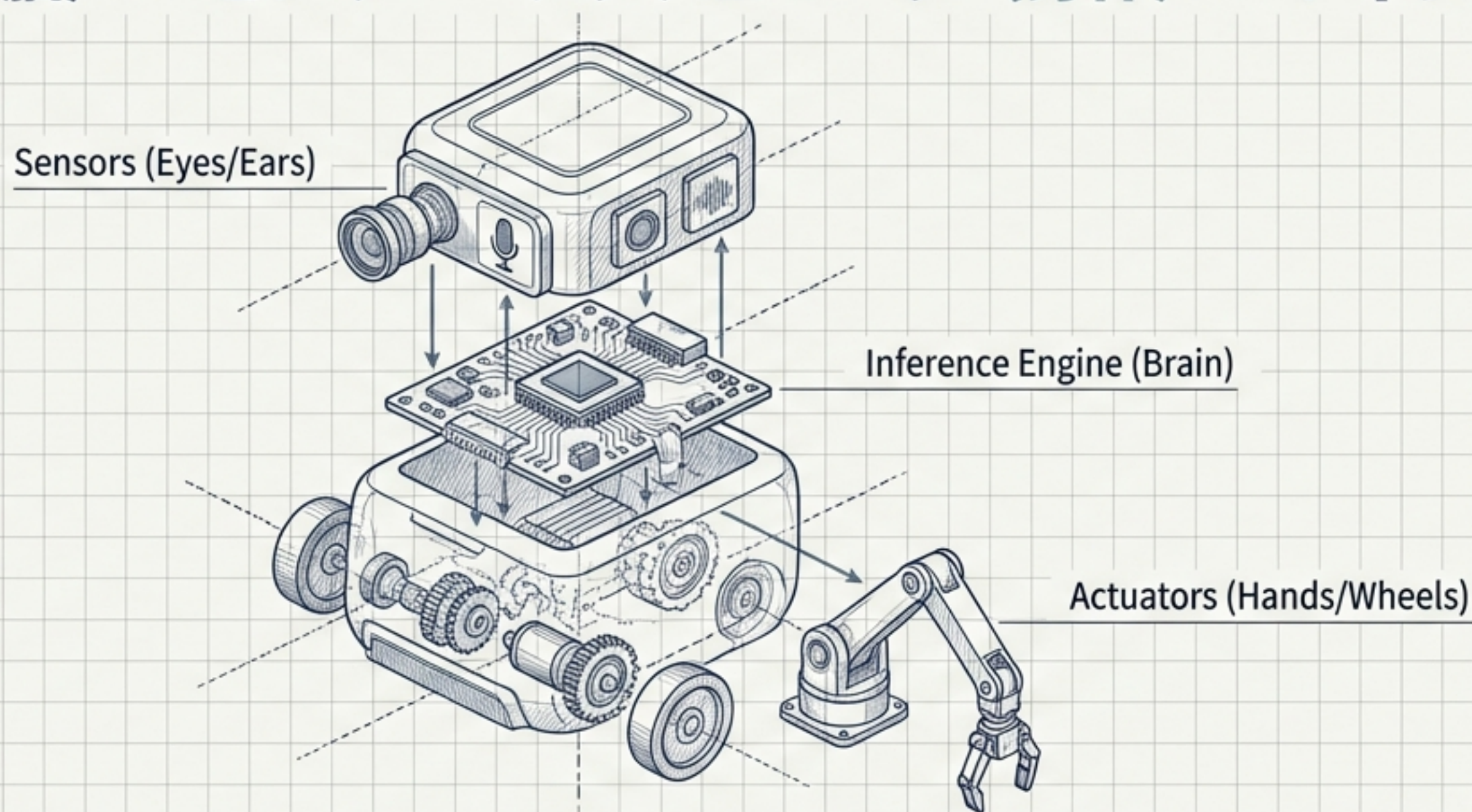
Slate Blue

Users (利用者)

Do not over-rely on AI; Set ‘Memory Management’ rules. (AIに過度に依存せず、「メモリ管理」ルールを設定する。)

ソフトウェア的知能とハードウェア的実体の統合：フィジカルAIの台頭

AIアルゴリズム（脳） + センサー・アクチュエータ（身体） = フィジカルAI



Market Examples

- Tesla: Autonomous driving (Self-judgment/operation)
- SECOM (cocobo): Security robots detecting abnormalities
- Iris Ohyama: Cleaning robots optimizing routes via LiDAR
- Yaskawa: Autonomous adaptive robot arms

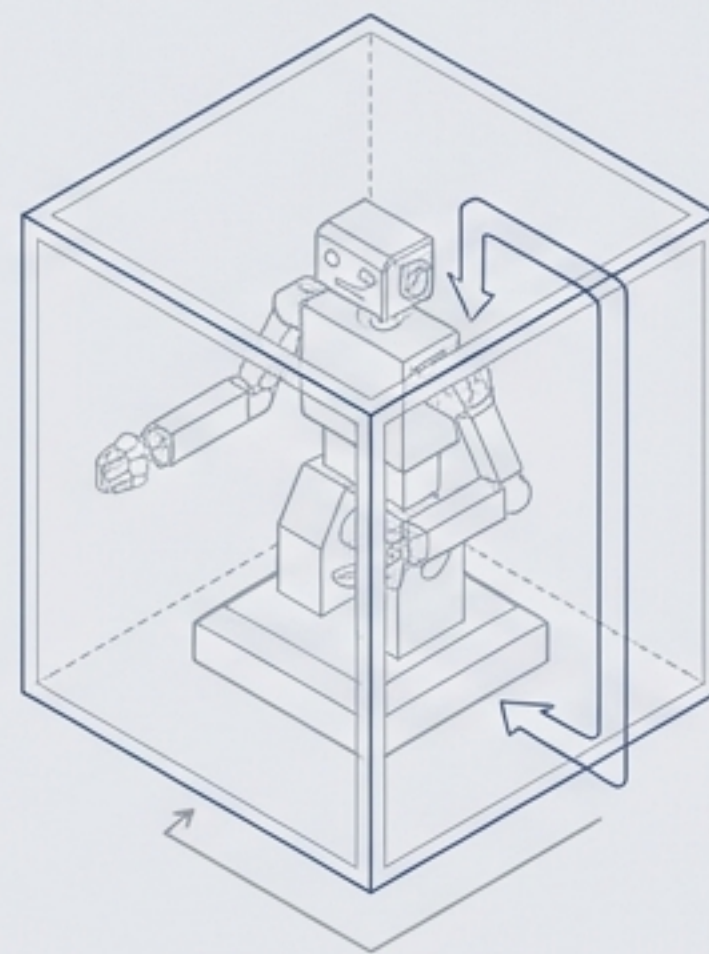
情報漏洩だけでなく「物理的事故」を防ぐ：フィジカルAIの安全設計

5 Specific Risks

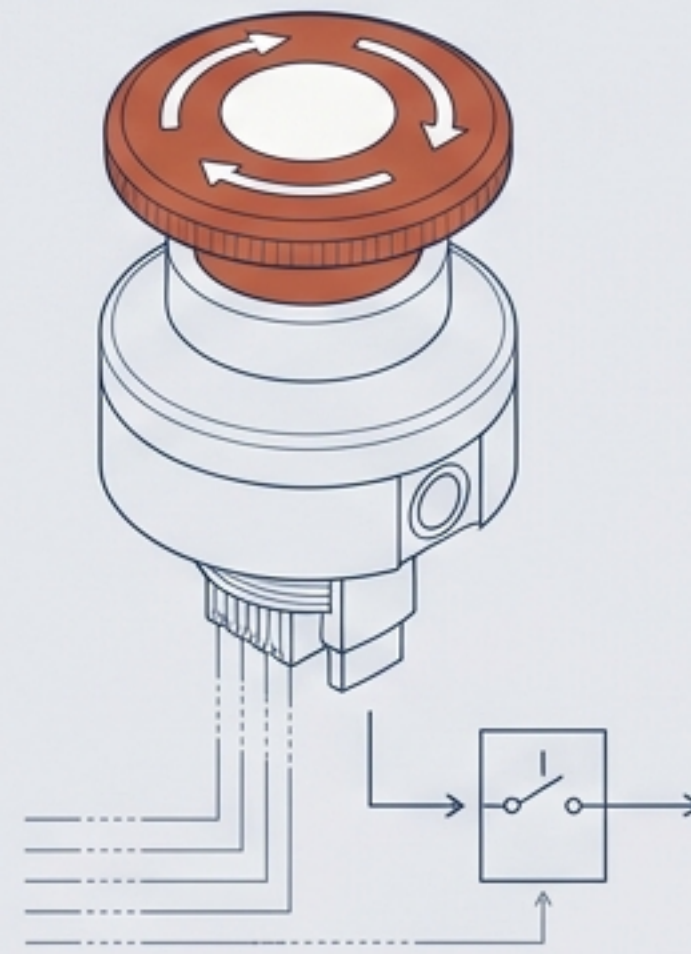
- ⚠ 1. 物理的事故
(Physical Accidents): Injury/Damage
- ⚠ 2. 個人情報の無断収集
(Privacy): Bystander capture
- ⚠ 3. センサー誤作動
(Sensor Error): Noise/Light interference
- ⚠ 4. 意図に反する学習
(Unintended Learning): Dangerous shortcuts
- ⚠ 5. 長期運用の不安定化
(Instability): Hardware degradation

Safety Measures

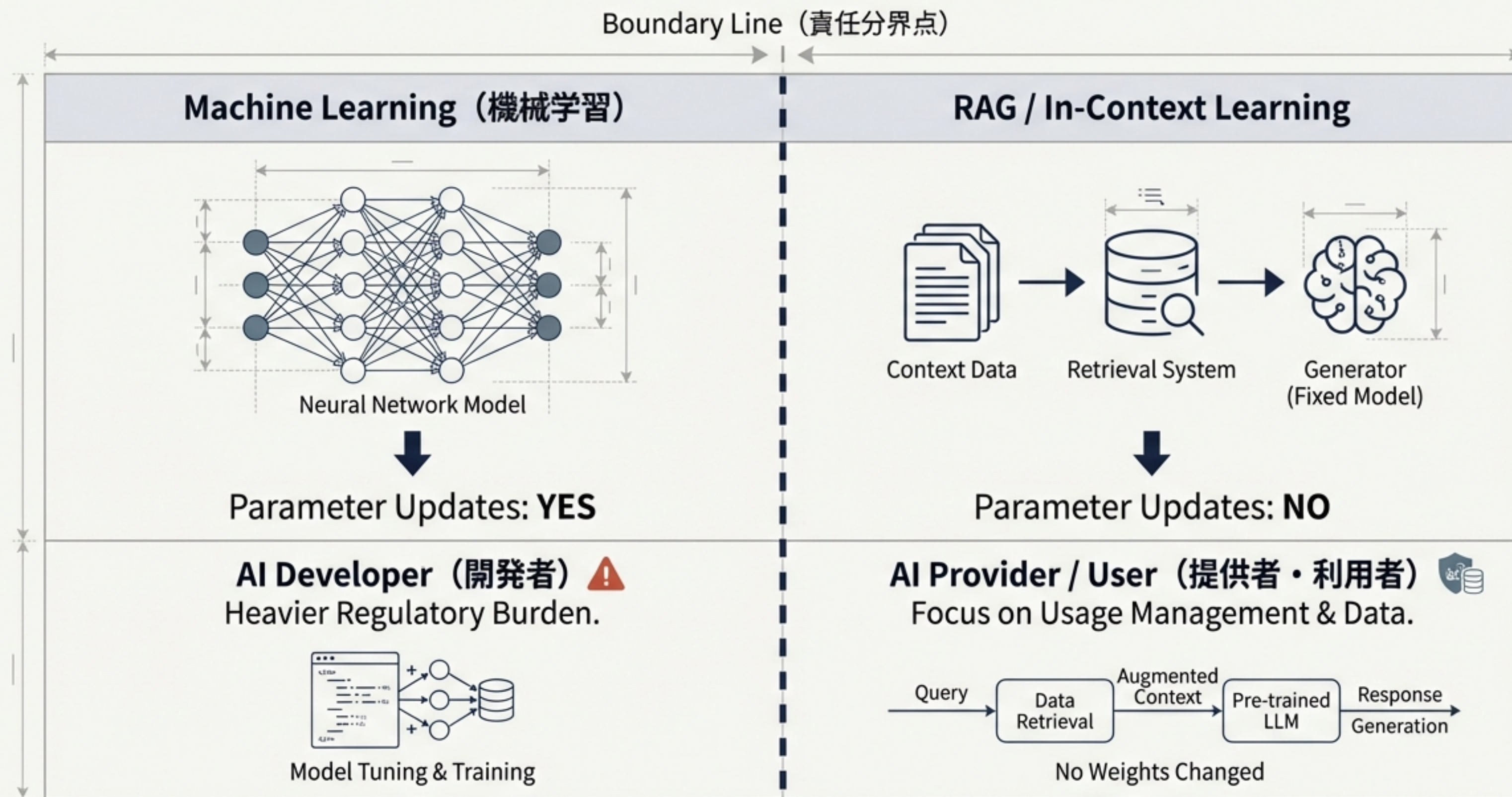
Digital Twin Verification



Fail-Safe / Emergency Stop



責任分界点の明確化：RAG（検索拡張生成）は「開発」にあたるのか？

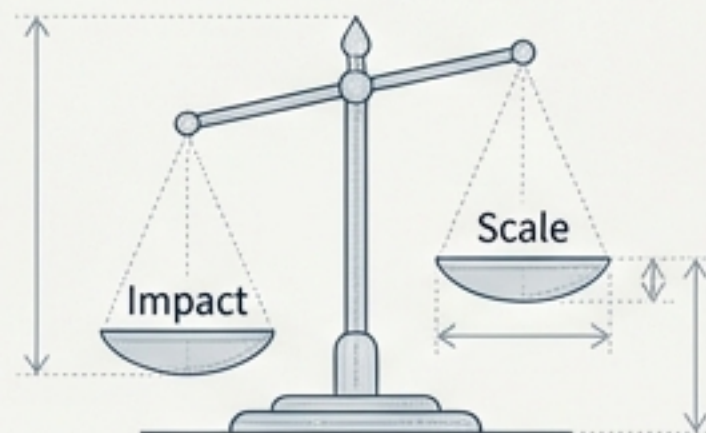


「なんとなく怖い」から「計算されたリスク」へ：新しいリスク評価手法

Equation A

$$\text{Risk} = \text{Impact (影響度)} \times \text{Scale (規模)}$$

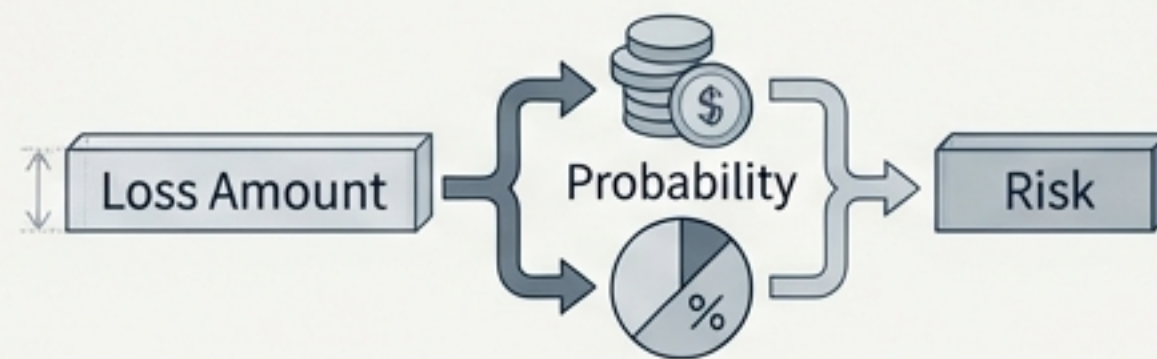
Social/Ethical Focus. Hard to quantify, but covers reputation.



Equation B

$$\text{Risk} = \text{Loss Amount (損失額)} \times \text{Probability (発生確率)}$$

Economic Focus. Enables Cost-Benefit Analysis.

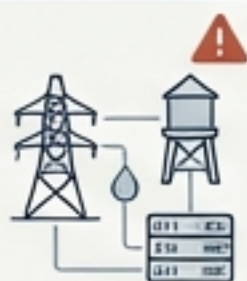


High-Risk Use Cases (EU Alignment)

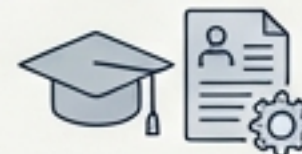
Biometrics
(生体認証)



Critical
Infrastructure
(インフラ)



Education/Employment
(教育・雇用)



Law
Enforcement
(法執行)



先行企業の活用事例：ガバナンスを「競争力」に変える

Softbank Group

Goal: Trust in Sales (営業における信頼)

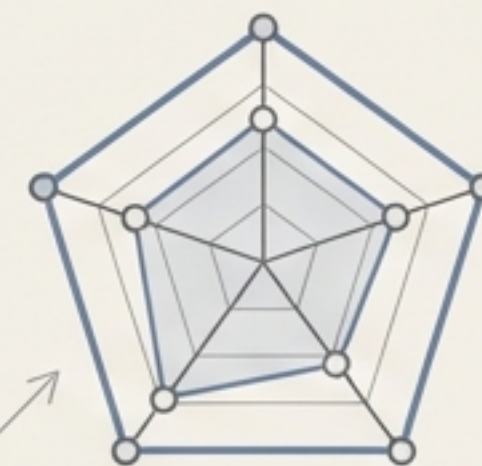
Action: Incorporated into decision-making (Rules, Operations, Judgment Axes).

“Governance is a front-office sales tool, not just back-office compliance.”



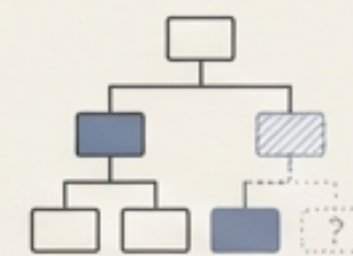
AI Governance Association (AIGA)

Solution:
AI Governance Nav



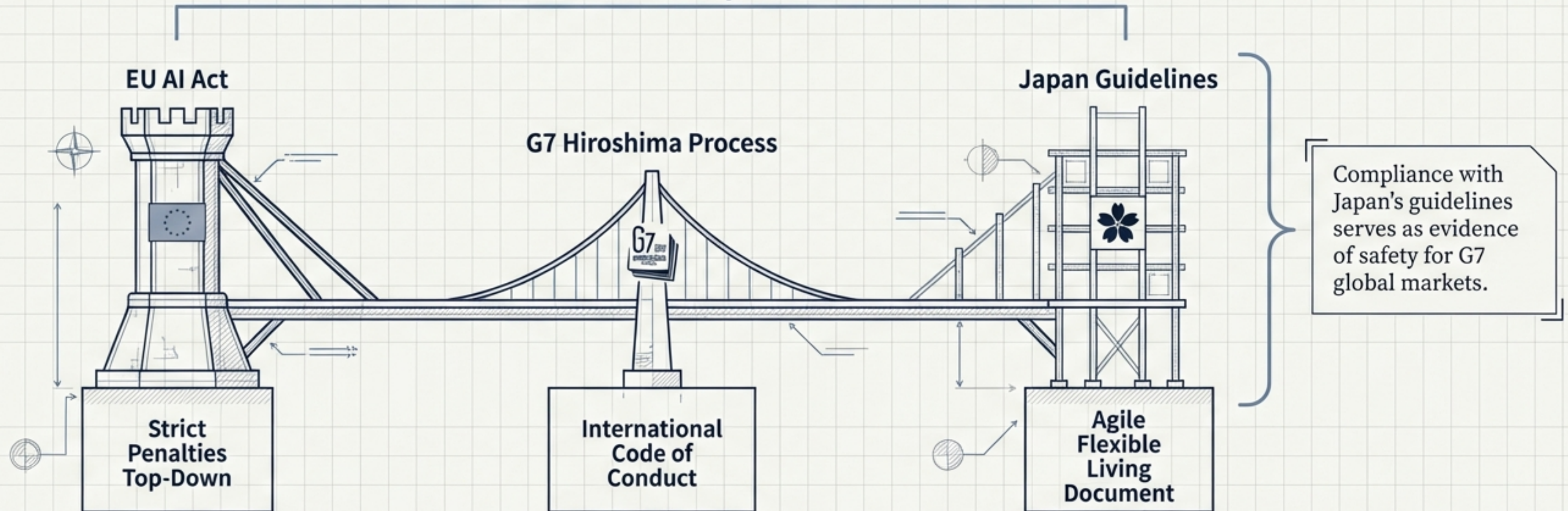
Maturity Diagnosis

Focus: Detecting “Shadow AI” and diagnosing organizational maturity.

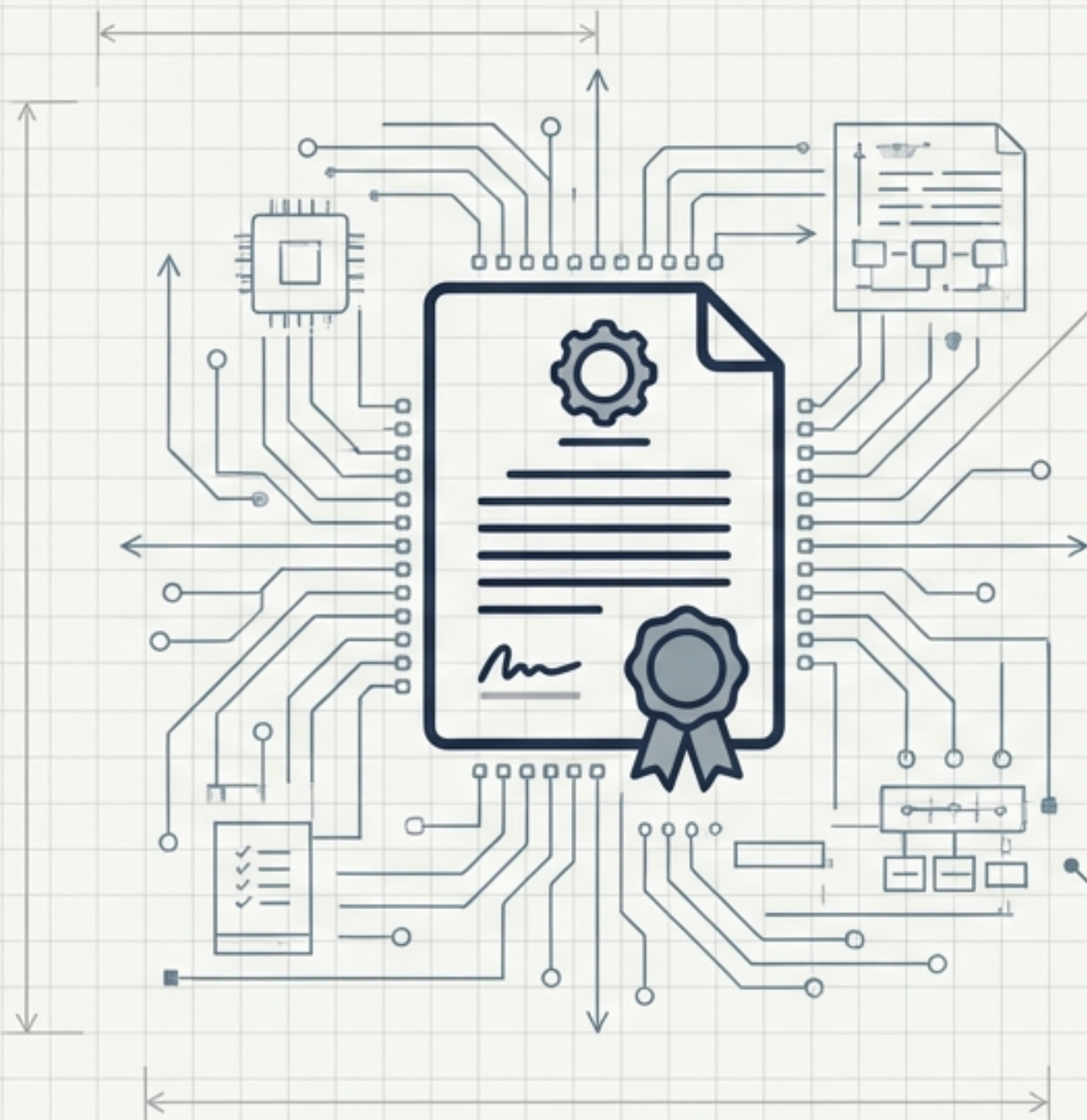


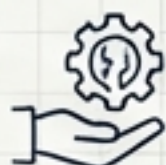
国際協調と日本のアプローチ：広島AIプロセスとEU AI法

The Governance Spectrum



知的財産戦略への示唆：安全機構は「特許」になる



1. HITL Mechanisms (人間介入のロジック):
Patenting the specific logic of *when* and *how* a human intervenes. 
2. Physical Safety (物理的安全機構):
Emergency stops and Digital Twin verification processes. 
3. Liability & RAG (責任分界):
Claims design based on the separation of parameter updates vs. retrieval logic. 



Business Risk Shield

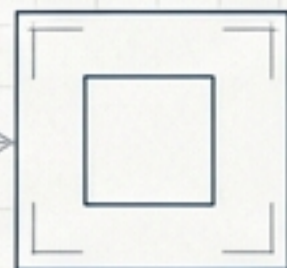
If an Agent makes an unauthorized purchase, who pays? Clear “Human Approval” logs act as the legal defense.

2026年4月に向けて今すぐ着手すべき5つのアクション



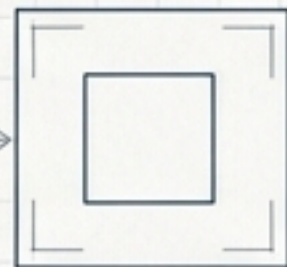
1. Role Re-verification:

Are you a Developer, Provider, or User? (Check RAG usage).



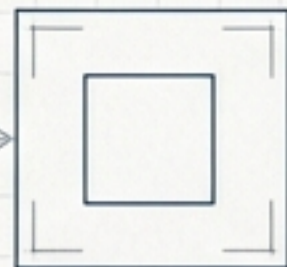
2. Agent Risk Assessment:

Run the '12 Risks' check against current tools.



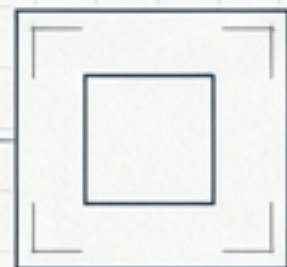
3. Physical Safety Audit:

Confirm Digital Twins / Emergency Stops.



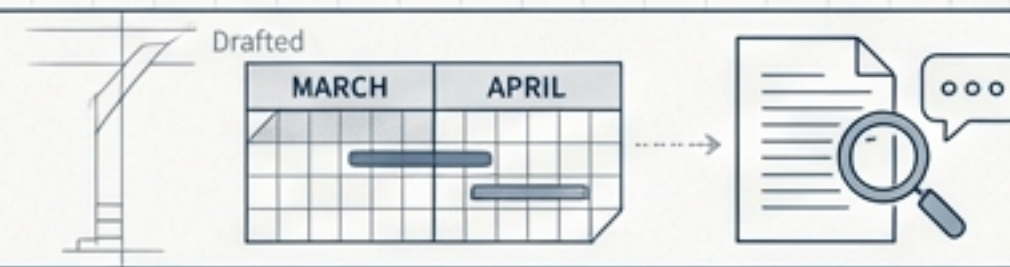
4. Global Alignment:

Ensure reporting aligns with Hiroshima Process.

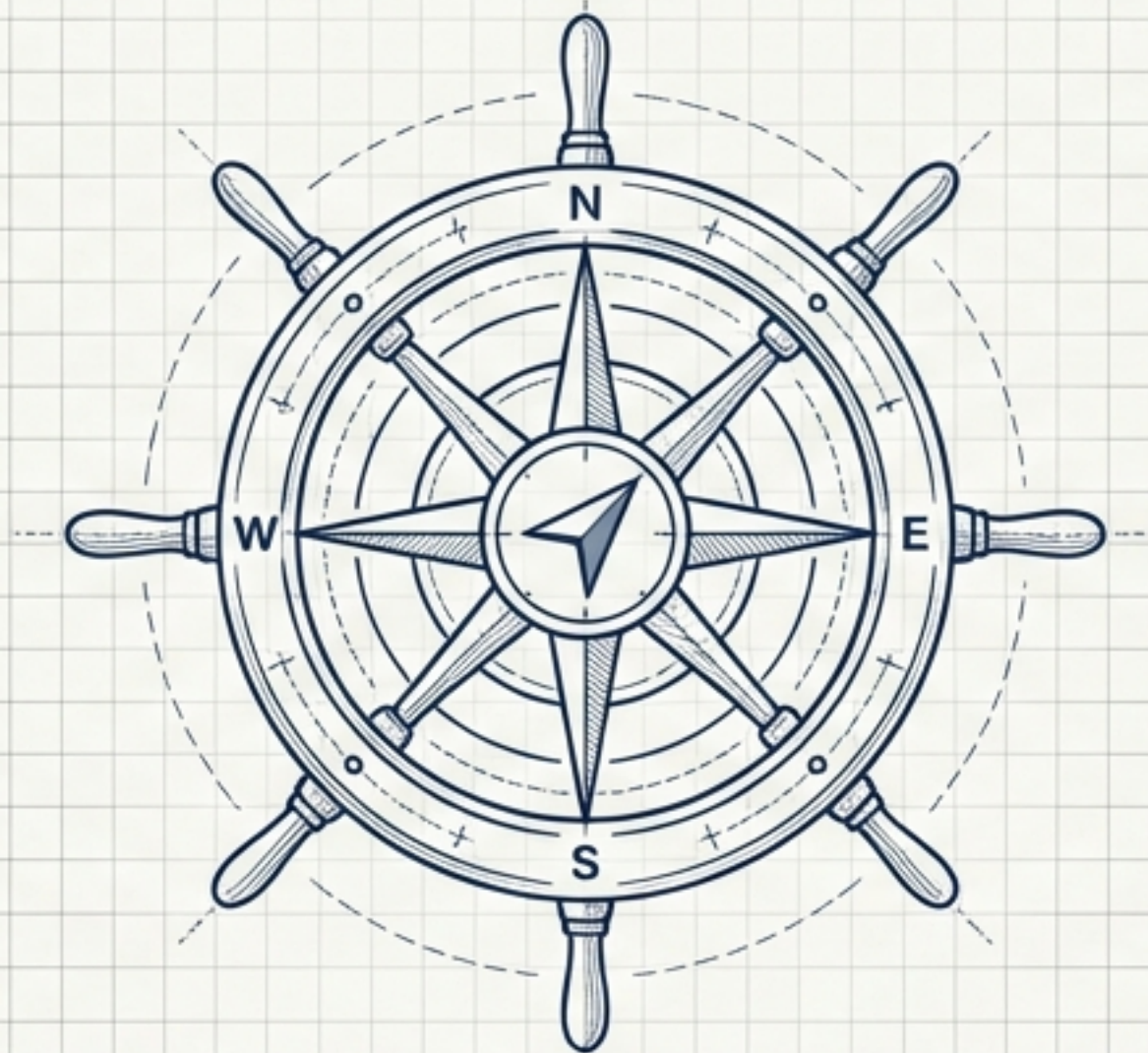


5. Public Comment Prep:

Review draft before final March release.



ガバナンスは「ブレーキ」ではなく、 自律的成長のための「ハンドル」である



-  **Living Document:** The guidelines will evolve with technology.
-  **Message:** Proper governance (HITL, Safety) allows faster deployment because safety is assured.
-  **Call to Action:** Integrate these updates into FY2026 strategy immediately.