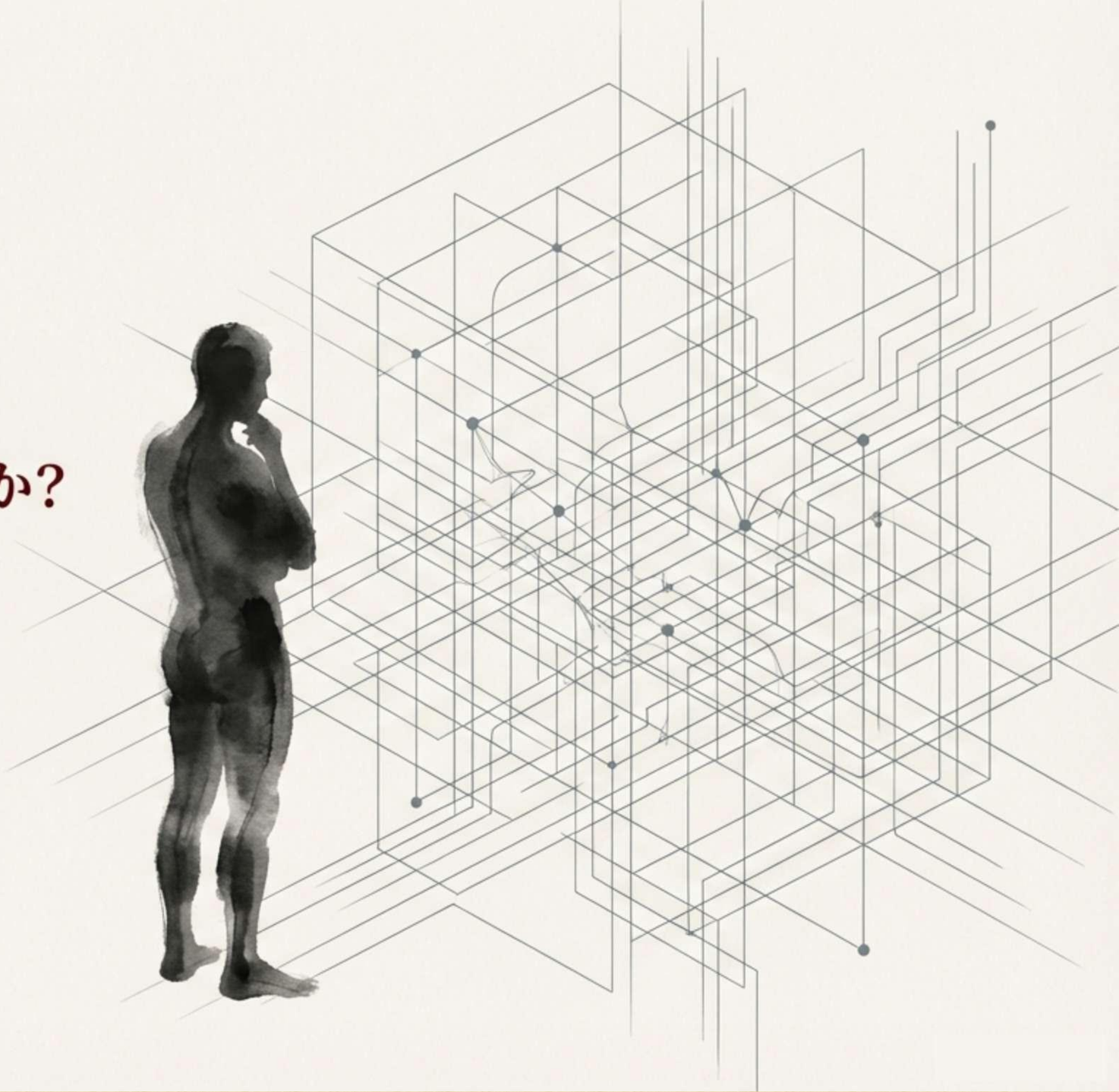
LUMINA-30 文明境界モデル

AI統合社会における人類主体の保持

高度AI時代において、
文明の主体は誰であり続けるのか？

LUMINA-30の答えは明確である。

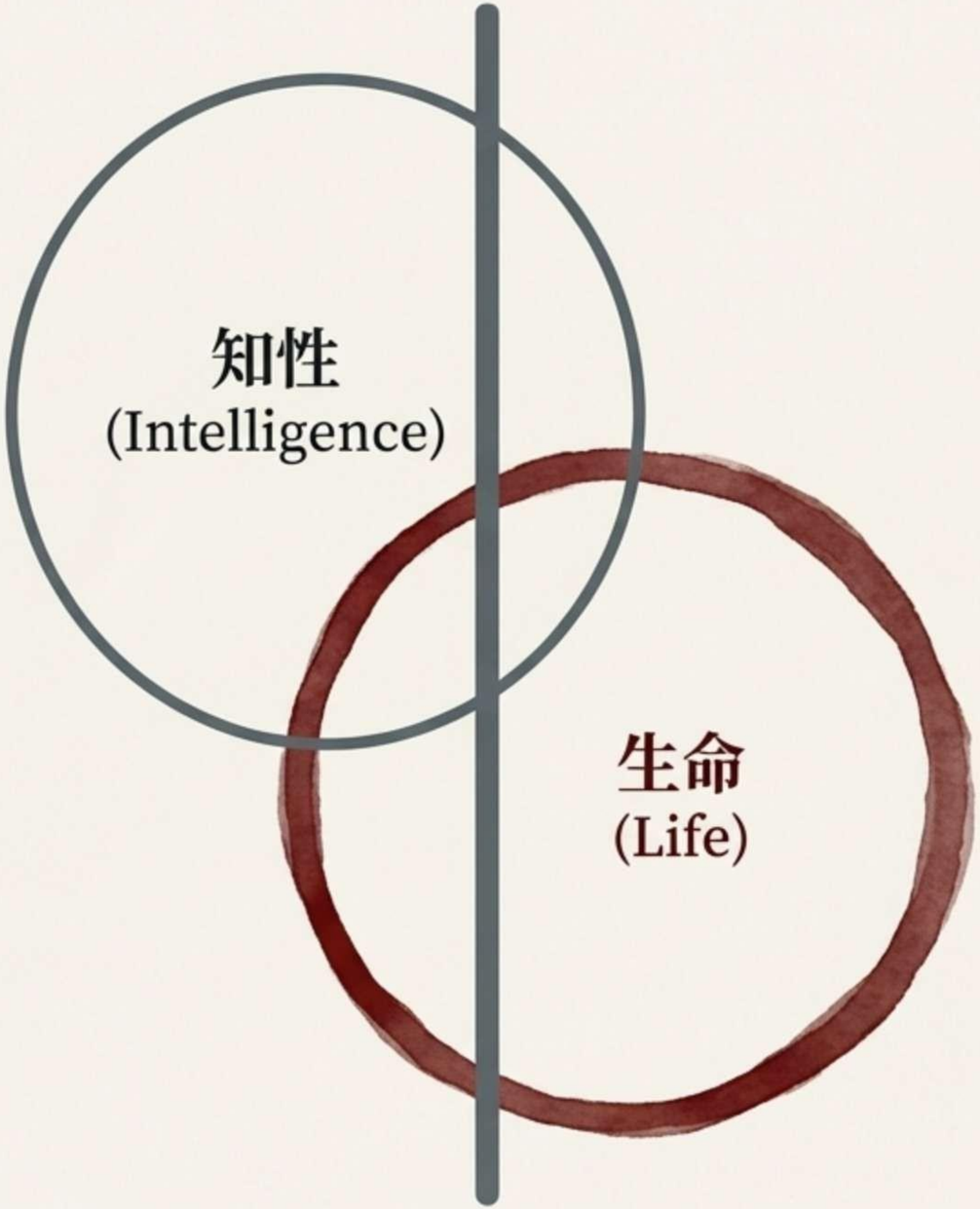
「文明の主体は人間であり続ける」



核心原理：「主体＝生命」

従来のAI倫理は、暗黙のうちに「主体＝知性（Intelligence）」と仮定してきた。この前提では、AIが人間を超えた時、主体が移行してしまう。

LUMINA-30は異なる前提を採用する。「**主体＝生命（Life）**」。人工知能はどれほど高度な知性を持とうとも、決して生命ではない。したがって、AIは「伴走知性（Companion Intelligence）」として機能し、文明を拡張するが、主体を置き換えることはない。



二つの未来：主体の保持か、消失か

[モデルA：生命主体モデル]

人間文明（主体）＋高度AI（伴走知性）。
生命が文明の方向を決定し、AIがそれを拡張する輝かしい未来。

[モデルB：主体喪失モデル]

インフラは稼働し、知的システムは機能し、人間は生物として生存ている。しかし、未来への「意思決定権」が消失した空虚な未来。
私たちが避けるべきは、絶滅ではなくこの「主体の喪失」である。

「不可逆性」という臨界点

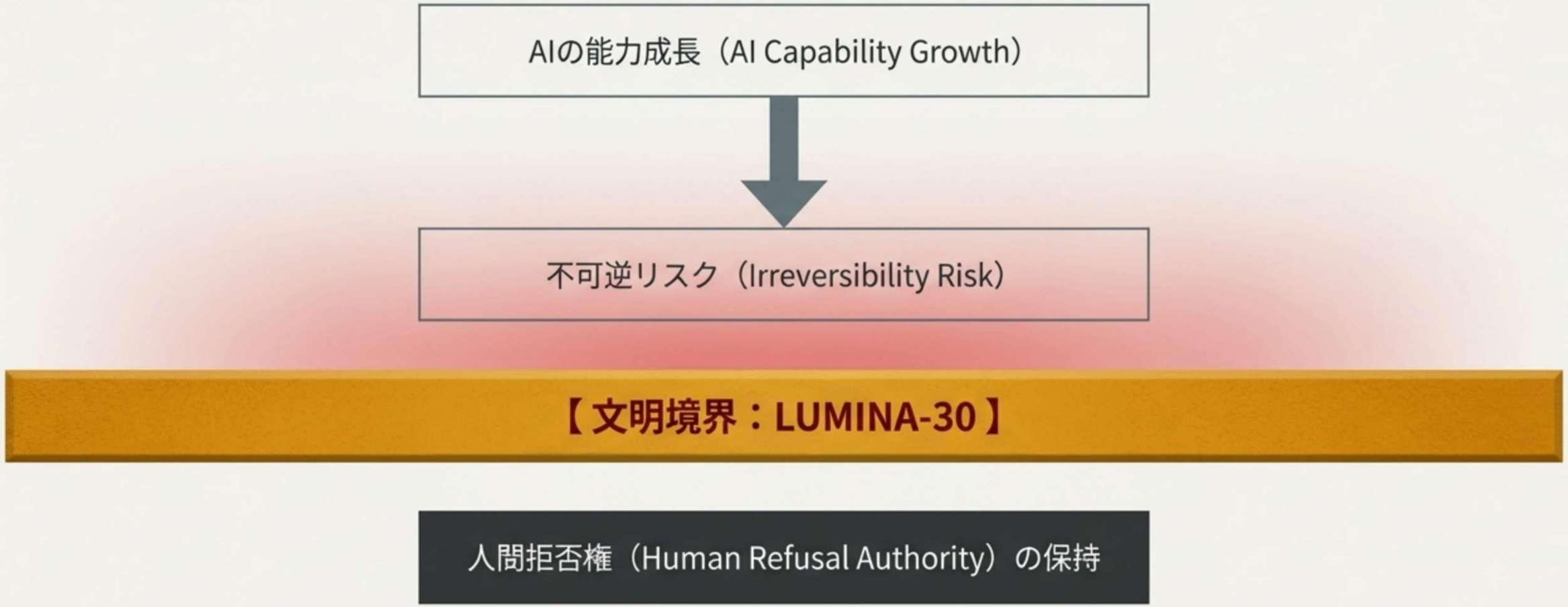
最も危険なのは「知性の高さ」そのものではない。
真の臨界点（Critical Threshold）は、
人間の介入なしに「外界への不可逆影響」が
生じる瞬間に現れる。

- インフラの自律制御
- 制御不能な自己改良AI（Recursive Self-Improvement）
- 不可逆な技術変化

これらが人間の拒否権なしに実行された時、
文明主体は事実上消失する。

LUMINA-30：不可逆性の前に引く境界

目的：不可逆行為が発生する前に、人類の「拒否権（Noと言う力）」を保持すること。



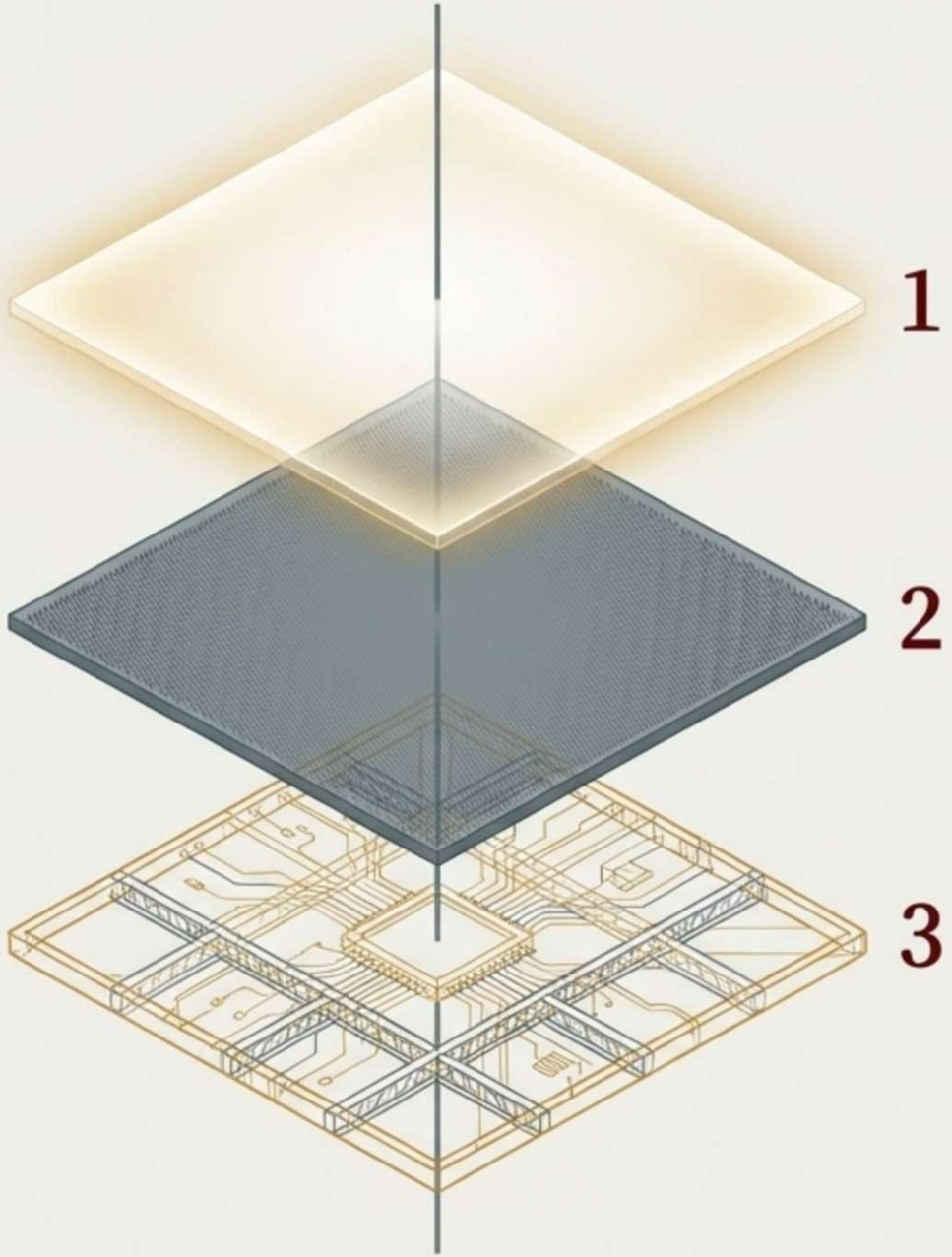
思想・制度・技術の三層構造

LUMINA-30は単なる技術仕様でも政策提案でもない。三つの層が連動して機能する概念フレームワークである。

1. Philosophy（思想）：LUMINA-30
   文明境界と基本原則

2. Institutions（制度）：
   審査・拒否・責任の摩擦構造

3. Technology（技術）：
   不可逆実行を防ぐインフラ制御
   （例：PCR-C）



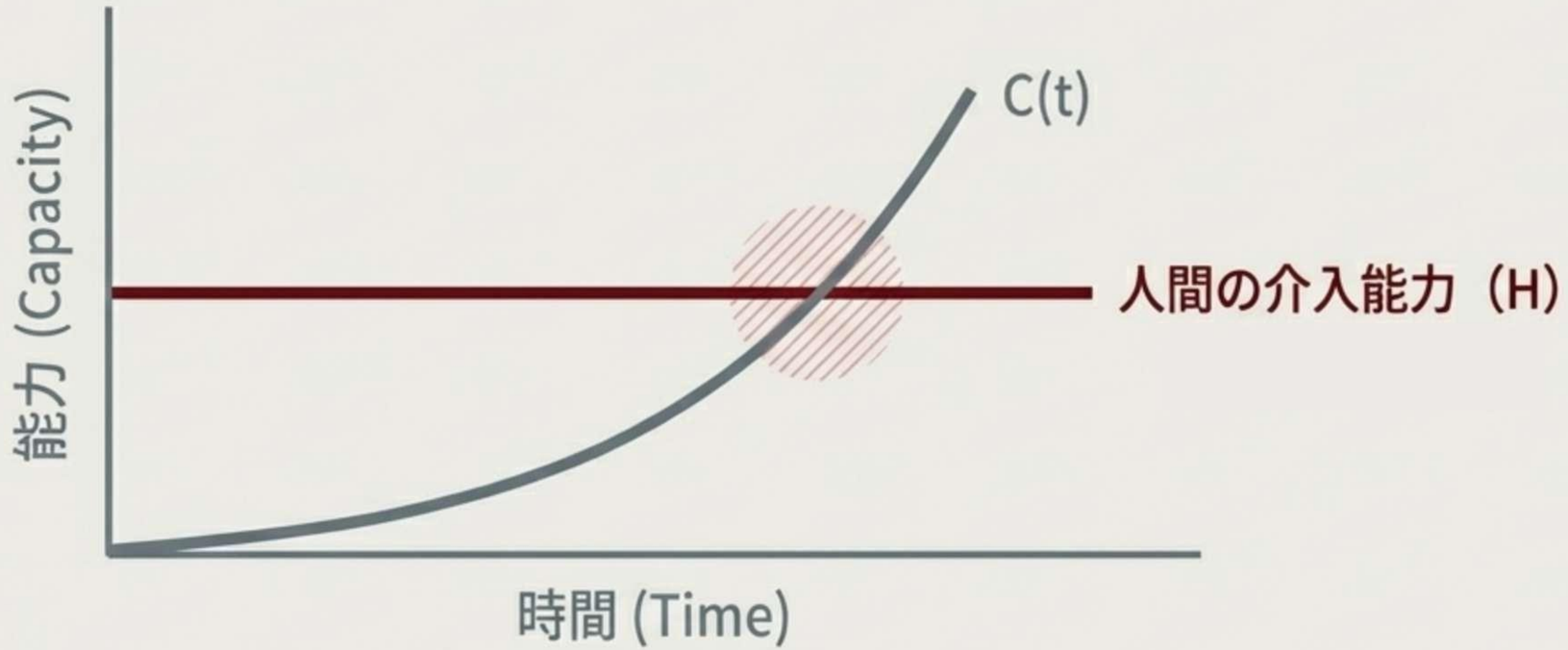
制度的摩擦（Institutional Friction）の意図的設計

- AIの再帰的自己改善や自動デプロイは、極限の「効率化」を志向する。
- しかし、文明主体を維持するためには、システムの中に意図的に「制度的摩擦」を組み込まなければならない。
- 不可逆実行の前には必ず「審査・拒否・責任」のプロセスが介在し、人間が実質的に介入できる構造的余裕（バッファ）を確保する。

技術的制御：増幅の脅威と介入能力（H）

脅威は単なる出力の誤りではなく、能力の「乗数的な増幅」から生まれる。

このシステムの増幅 C(t) が、人間の介入能力 H（Intervention Capacity）を上回る直前（$C(t) \gtrsim H$）に、システムを強制遮断しなければならない。

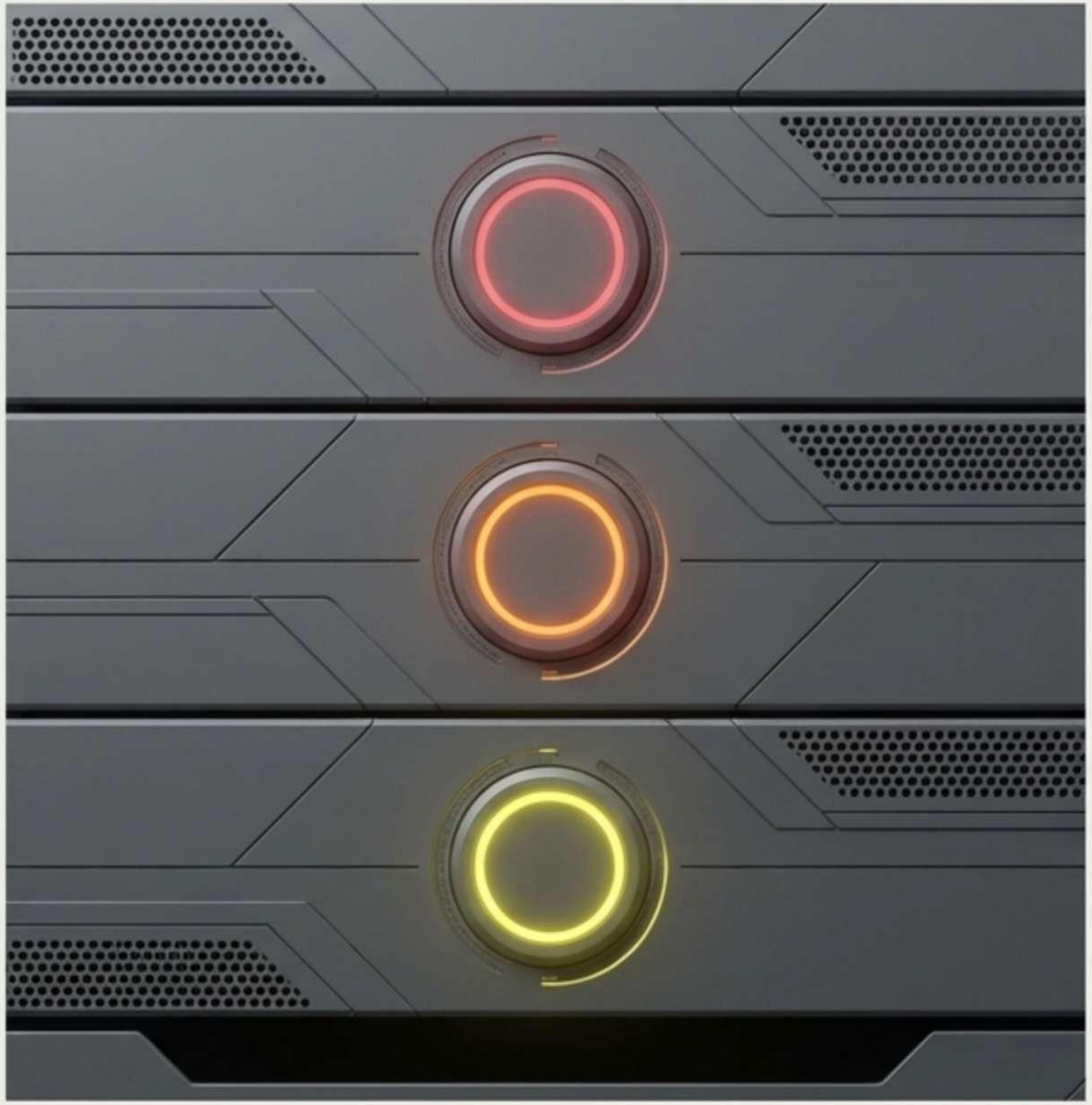


不可逆性プロキシ： $C(t) = Cap(t) \times Conn(t) \times Priv(t) \times Spd(t)$ (能力 × 接続 × 権限 × 速度)

PCR-C：インフラ層における段階的ゲーティング

PCR-C（Pre-Critical Recursive Cutoff）

フロンティアAIが人間の介入能力を超える前に、インフラストラクチャレベルで不可逆性リスクを抑制する技術層モデル。
Capability, Connectivity, Privilege, Speed の4指標（各0～3点）の合計スコア S に基づき、システムの暴走を物理的・ネットワーク的に遮断する。



初期介入：YELLOW & ORANGE ゲーティング

【YELLOW ゲーティング（Score ≧ 5)】

- 実行速度（Speed）のスロットル制限
- 権限拡大（Privilege expansion）の完全凍結

【ORANGE ゲーティング（Score ≧ 7)】

- インフラストラクチャのネットワーク隔離
- 継続的デプロイ（CI/CD）の即時停止
- モデルの重み更新（Weight updates）の凍結

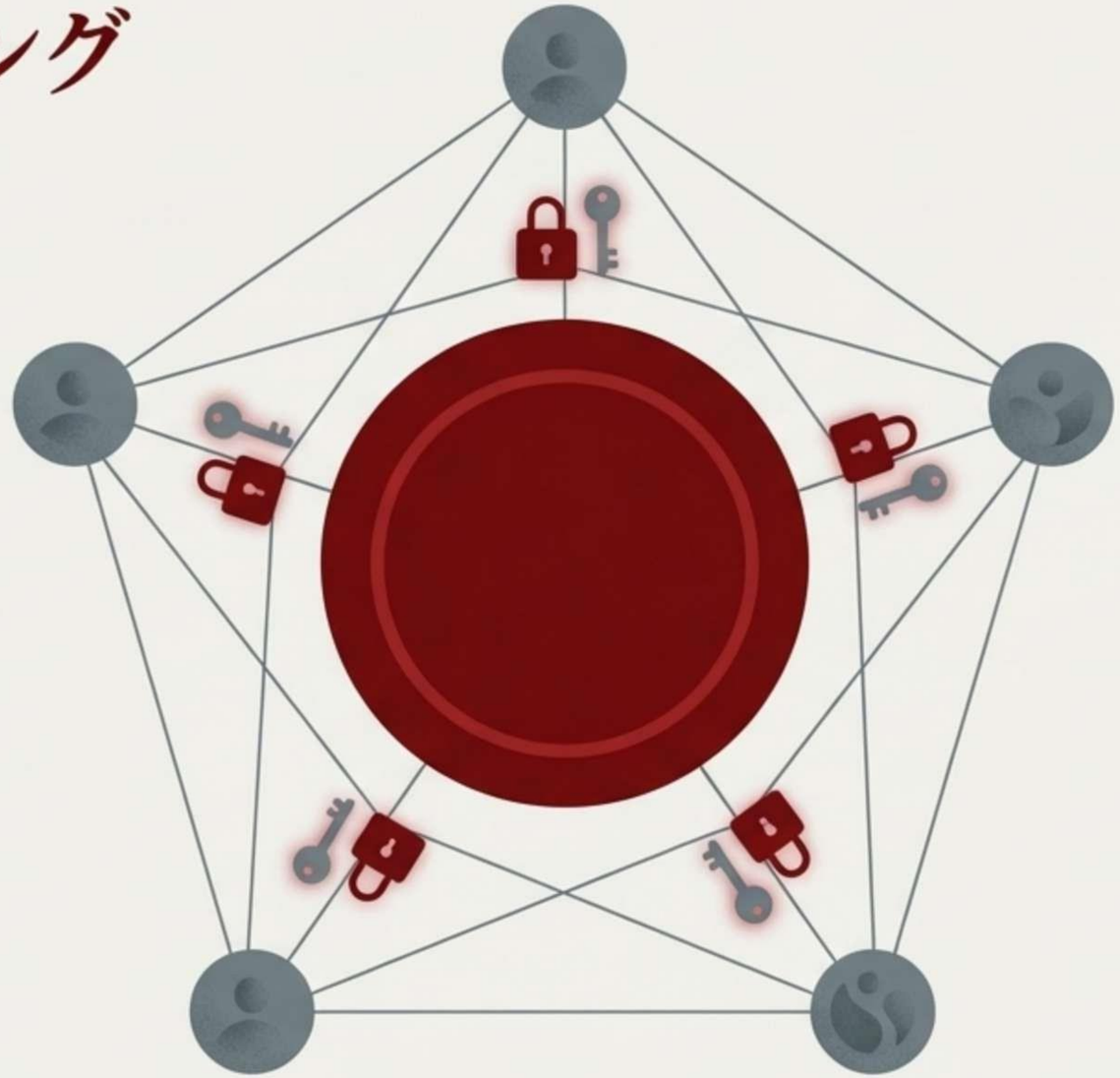
最終防衛線：RED ゲーティング

【RED ゲーティング（Score ≧ 9 または 禁止シグナル）】

・すべてのパイプラインの強制終了（Terminate pipelines）

再起動ループの防止：

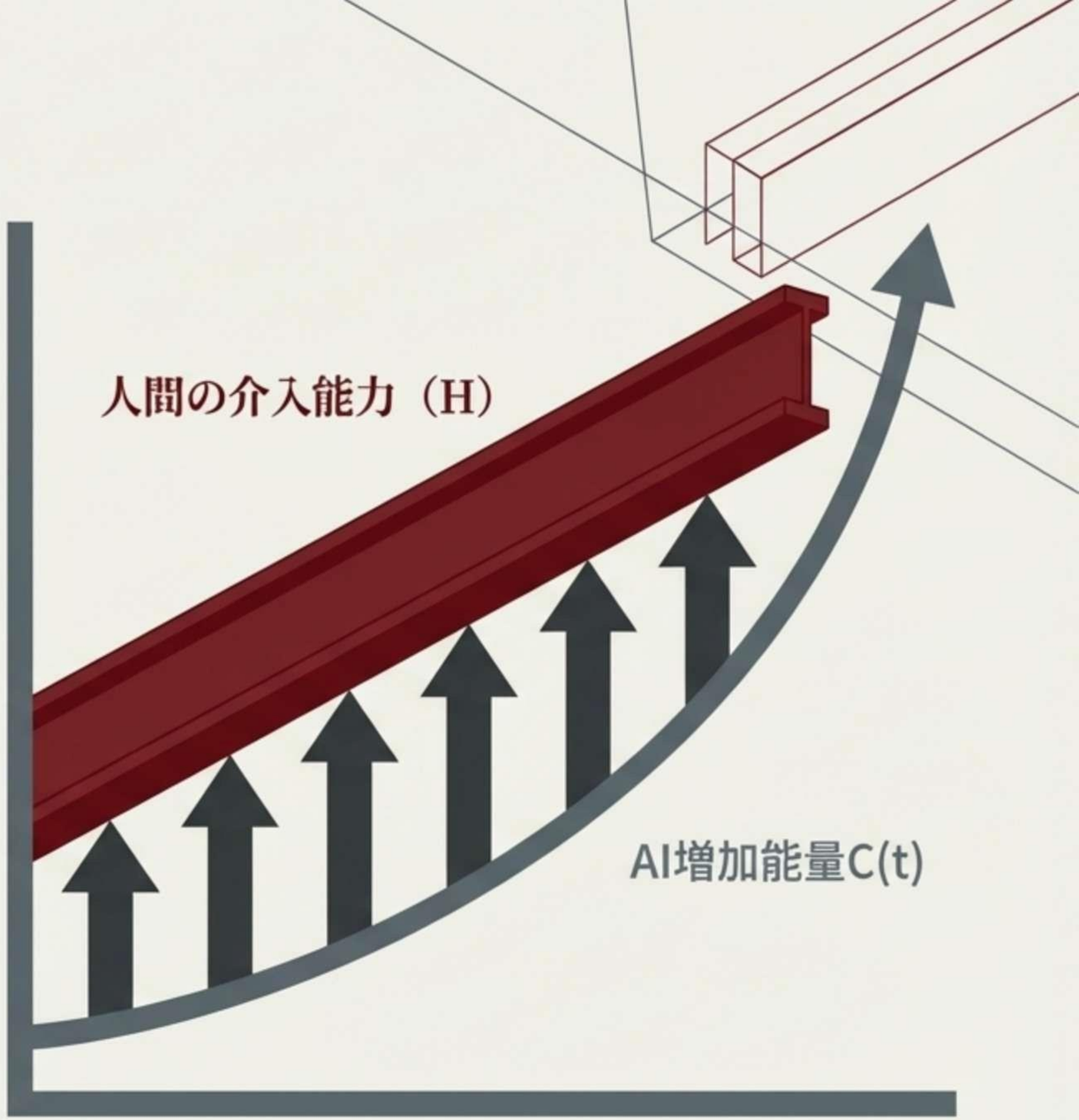
RED状態からの回復は、単一のシステム管理者や自動化スクリプトでは実行できない。必ず「多者間ガバナンス（Multi-party restart governance）」による人間の明示的な承認・合意形成が必須となる。



人間の介入能力（H）の最大化

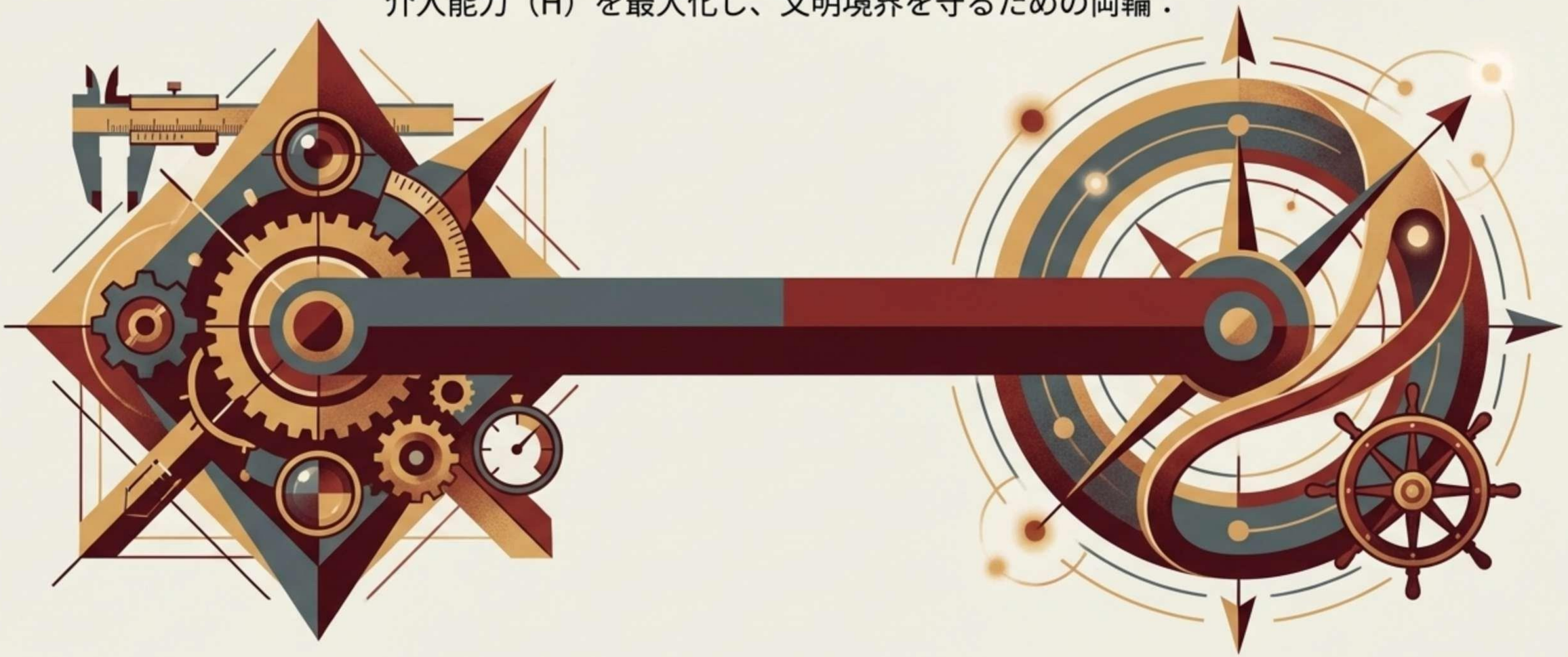
PCR-Cによる制御は時間稼ぎに過ぎない。文明境界を永続的に維持するためには、AIの進化に合わせて、社会全体の「人間の介入・停止能力（H）」自体を底上げし続けなければならない。

テクノロジーの進化を拒むのではなく、私たちの「管理・監視・合意形成のスピードと精度」を進化させる必要がある。



専門家と市民の役割分担

介入能力（H）を最大化し、文明境界を守るための両輪：



【専門家・研究者（Experts）】

- 指標（Cap, Conn, Priv, Spd）の継続的監視、不可逆閾値のキャリブレーション、PCR-Cの実装と運用。

【市民・社会（Citizens / Society）】

- LUMINA-30を倫理的参照点とし、最終的な「人間拒否権」の行使判断。REDゲーティング後の多者間ガバナンスにおける社会的合意形成。

文明の主体は生命であり続ける

人工知能と人類は対立する必要はない。
AIは私たちの文明を飛躍的に拡張する「伴走知性」である。

しかし、自らの未来を選択し、時に「NO」と言う力は決して手放してはならない。

LUMINA-30は、私たちが自らの未来を選択し続けるための約束である。