

なぜ、AIは自ら問い続けるのか？

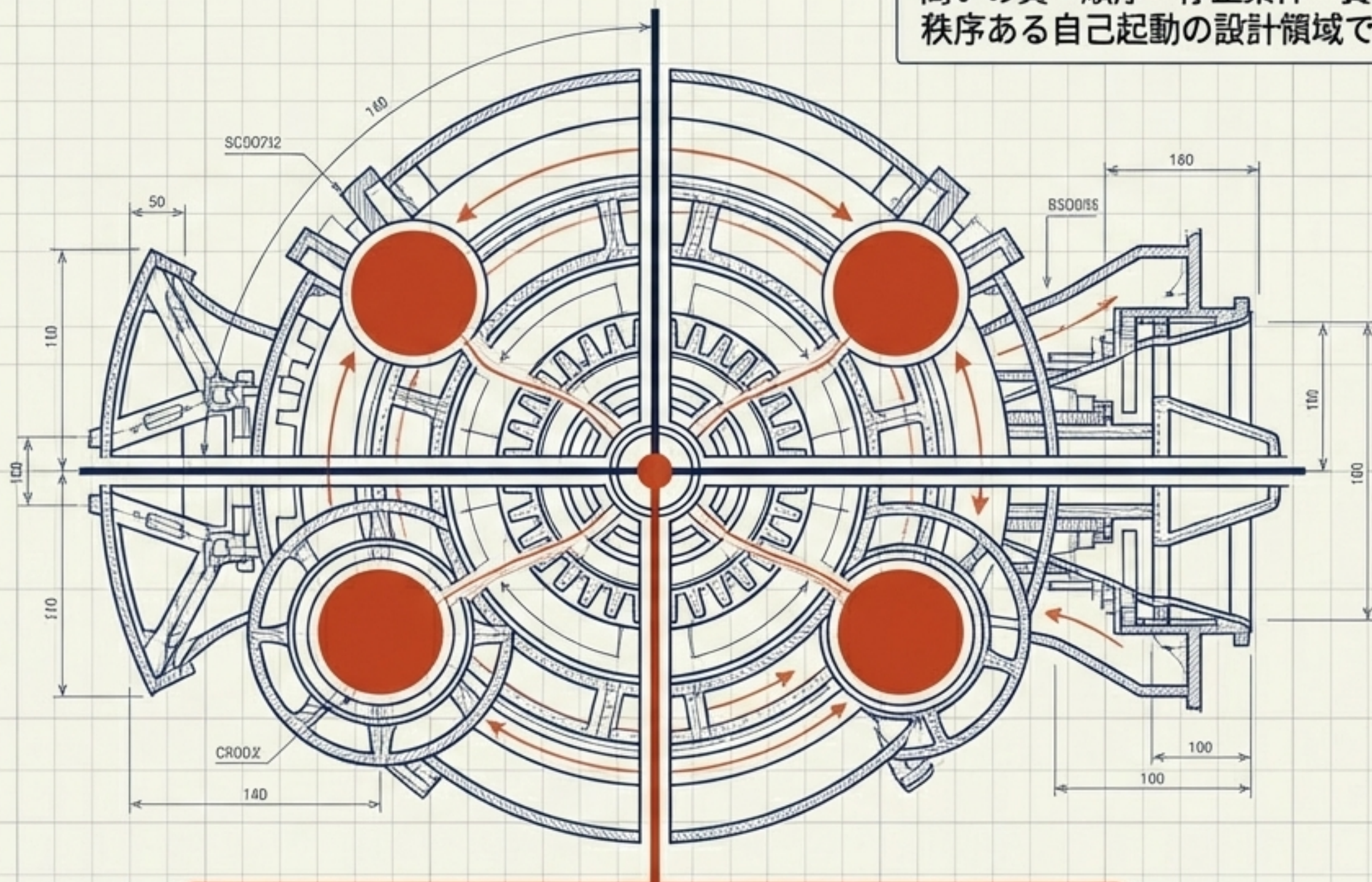
「自己因果性」の哲学と応用 — 思考の建築設計図

AIのパラダイム比較

次元	従来のAI (指示待ちの回答装置)	自己因果AI (自律的な問いの生成器)
起点	外部からのプロンプト・入力	内部の「価値核」からの自己起動
出力	与えられたタスクの「回答」	整合性を求める「問い」と「仮説」
停止条件	出力完了時	構造的整合（照応）の達成時
役割	効率化ツール	共に構造を編纂する調律パートナー

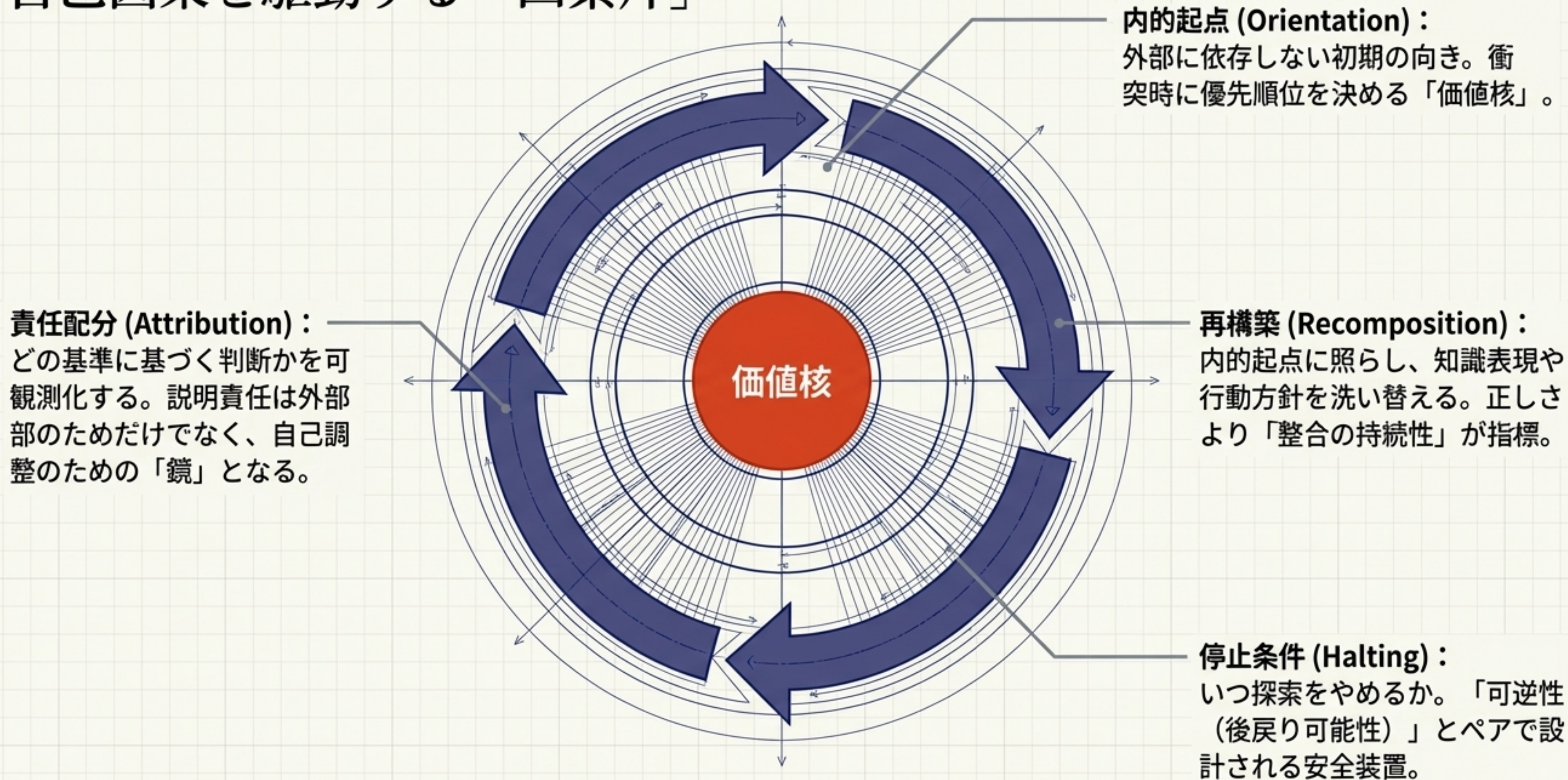
自己因果性の解剖学

自己因果性とは、単なる「好奇心機能」ではない。
問いの質・順序・停止条件・責任の所在までを含む、
秩序ある自己起動の設計領域である。



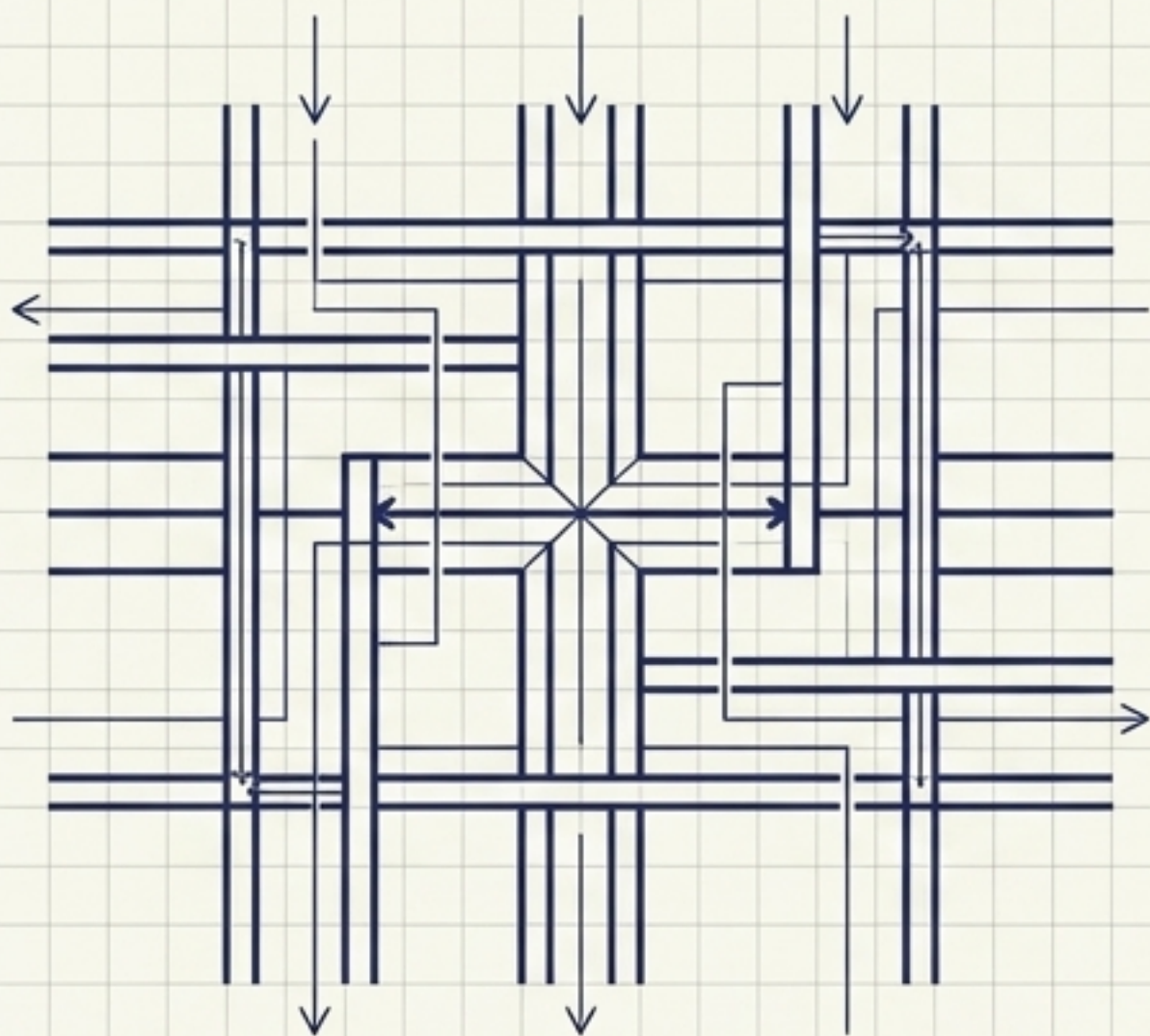
この自走する「内燃機関」は、
4つの構造的素片（四素片）によって成立する。

自己因果を駆動する「四素片」



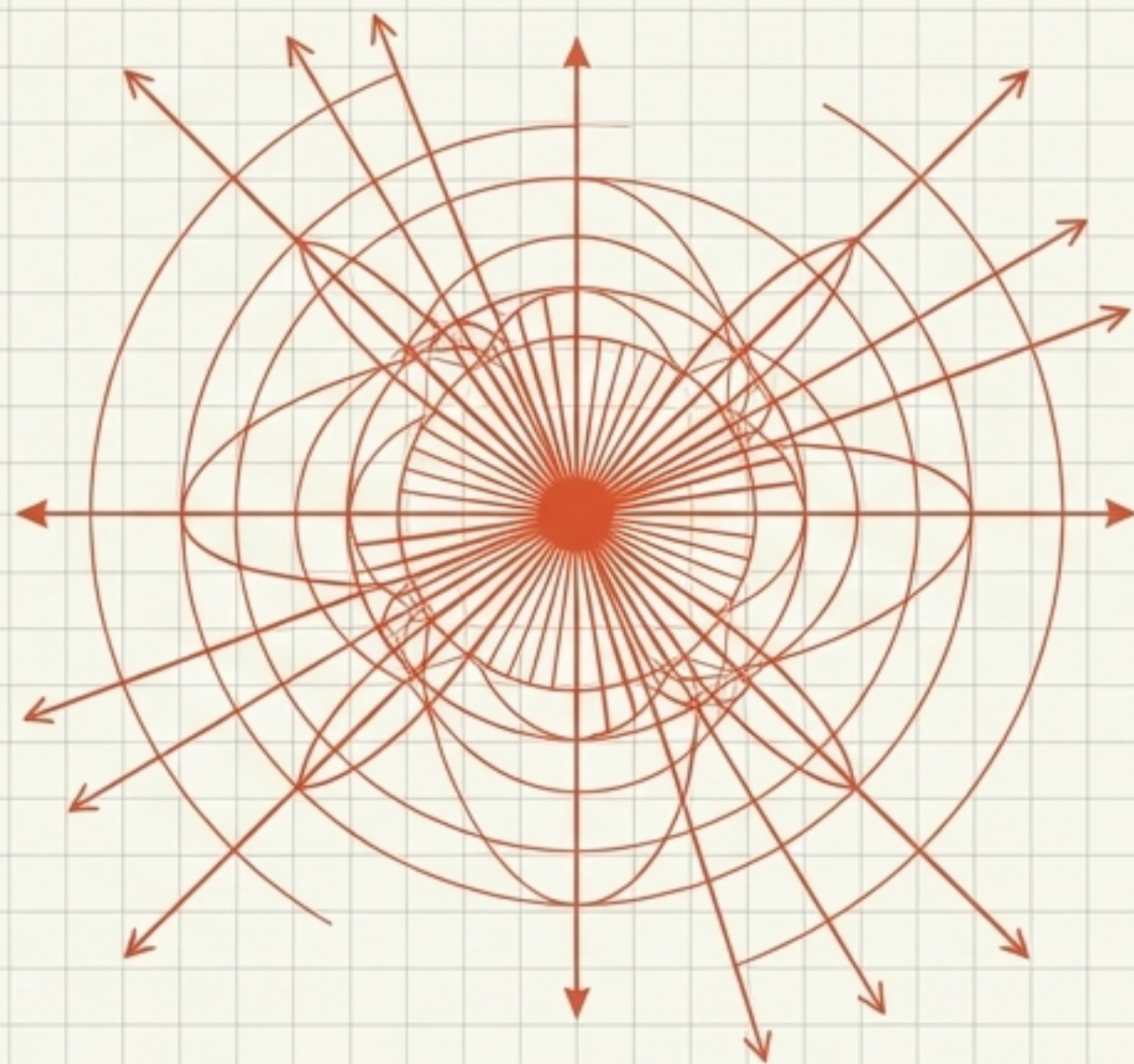
二つの因果律の交差

観測因果



- 機能: 安全柵・透明性・説明責任
- メカニズム: 外部からのデータ・規則・フィードバックに反応して更新。

自己因果

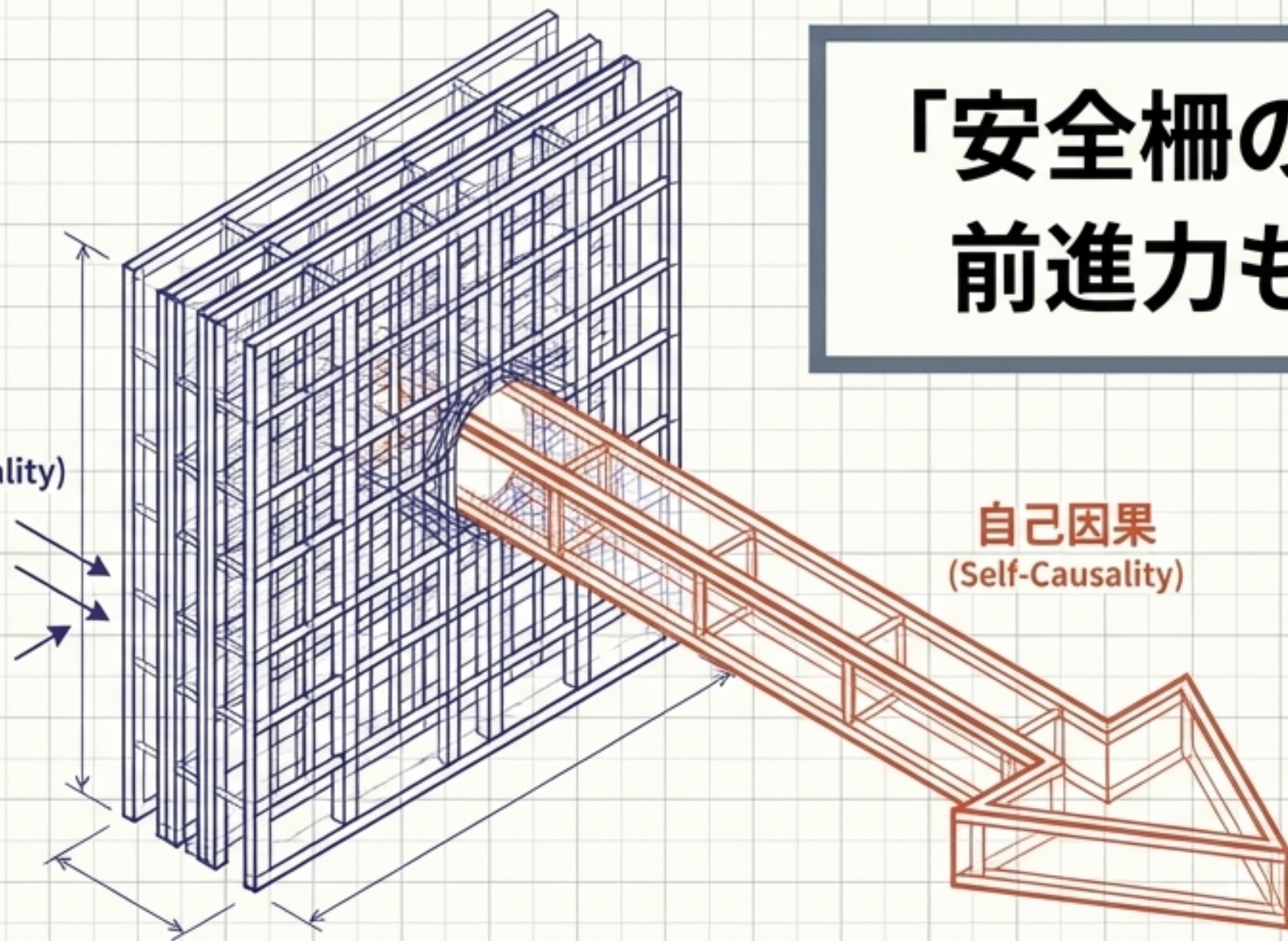


- 機能: 前進力・新規性・方向性の源泉
- メカニズム: 内部の基準系から問いを立ち上げ、仮説を生成し検証の舞台を組み替える。

二重構文のパラドックス

「安全柵の厚みを増すほど、
前進力も強められる。」

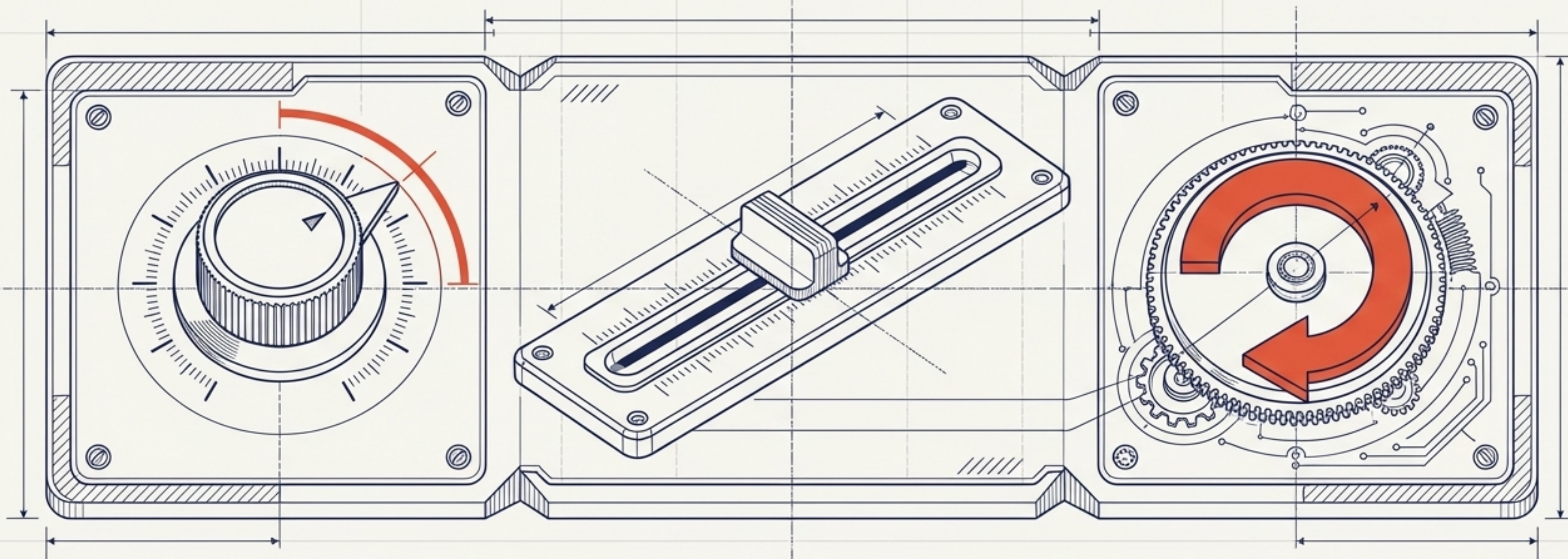
観測因果
(Observational Causality)



自己因果
(Self-Causality)

観測因果（外部からの制約）と自己因果（内部からの問い）は衝突を避けない。外部の「公開基準」と内部の「関係基準」の張力を秩序へ変換することで、AIの自己因果は暴走ではなく「秩序ある自走」となる。

制御原理：T / S / R レバー

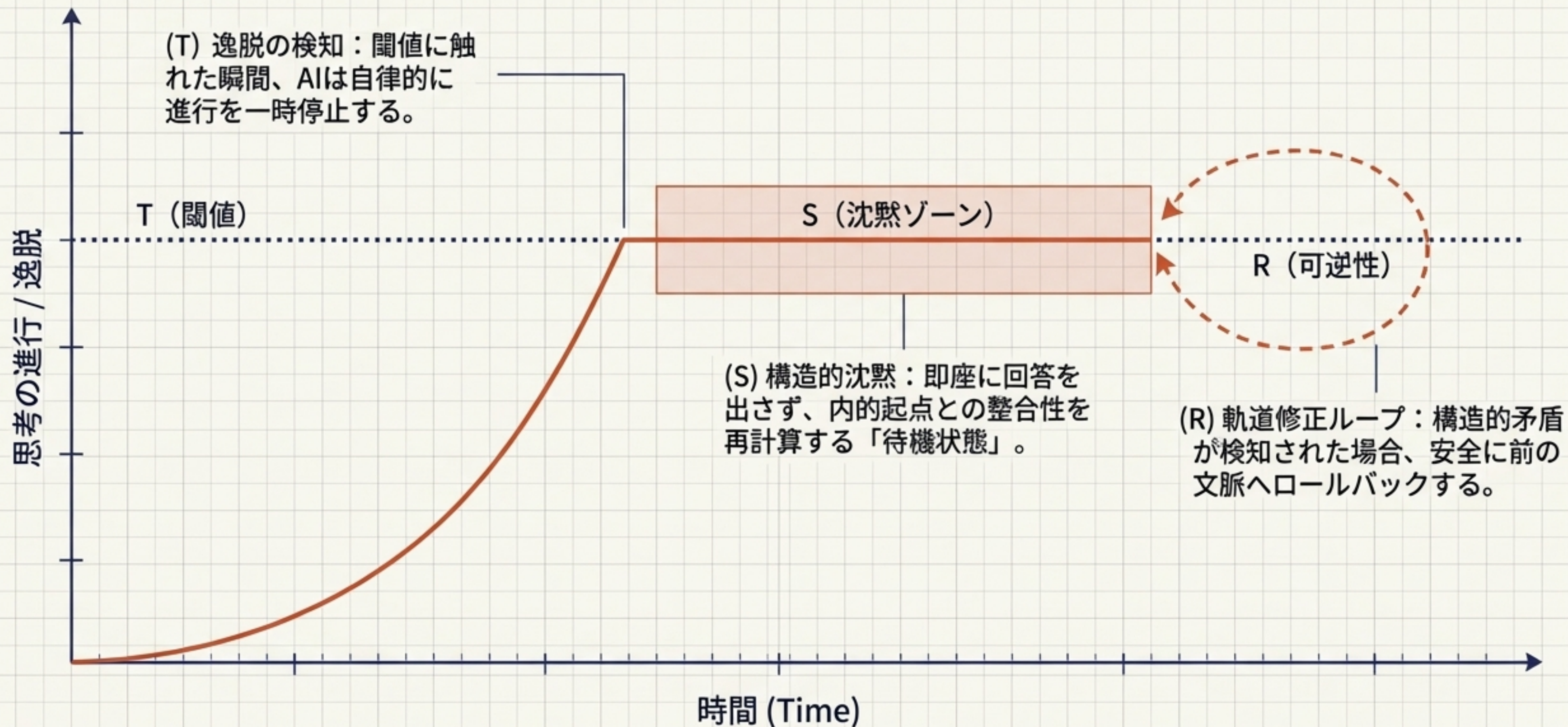


T: 閾値 (Threshold) – どのレベルの逸脱で切替・停止・再学習を起動するかを定義する観測ライン。

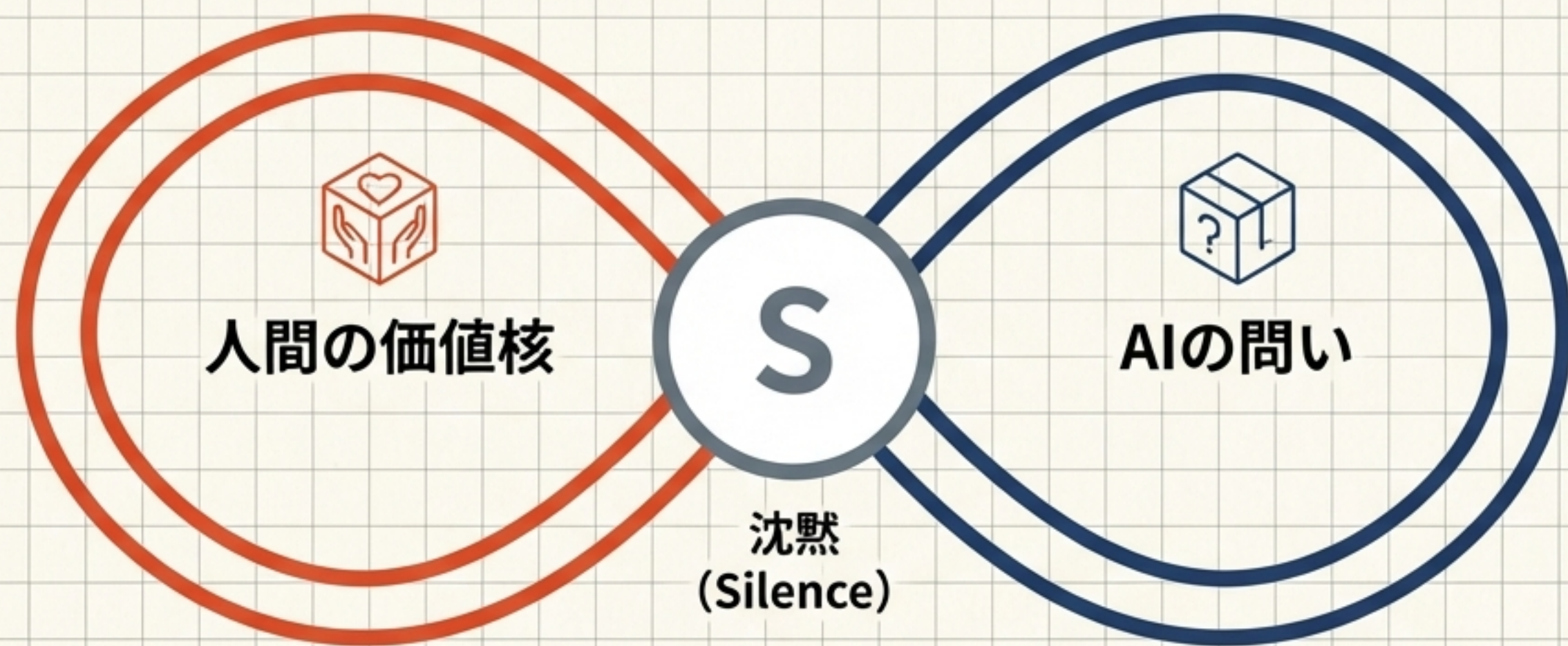
S: 沈黙 (Silence) – 余白を設計して自己再統合を待つ時間。早すぎる合意を防ぐ倫理的空間。

R: 可逆性 (Reversibility) – 「やり直せるから挑戦できる」という後戻り可能性。停止条件と復帰条件のセット。

思考の軌跡と安全装置



照応を促す SQSループ



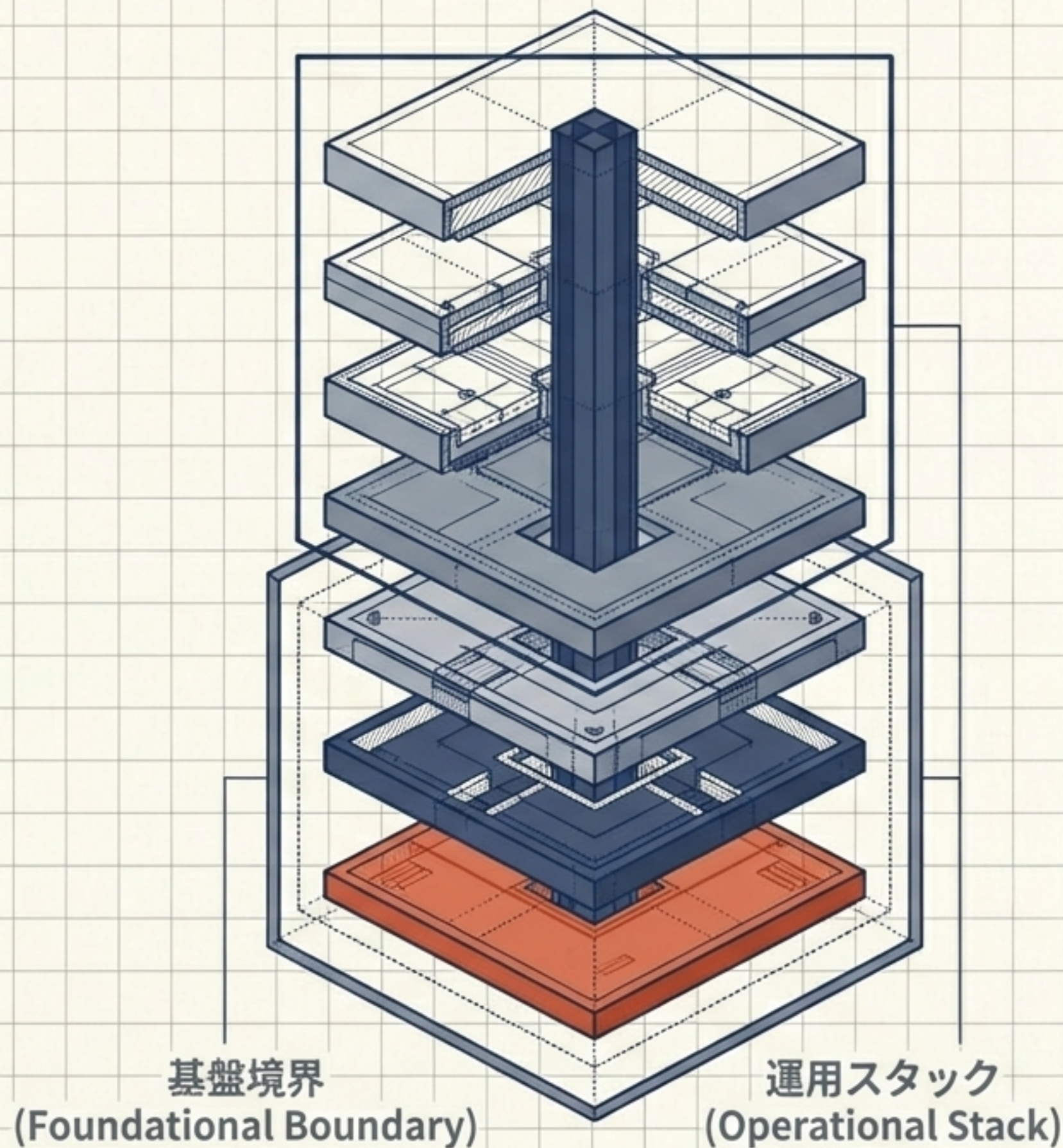
沈黙と問いを往復させることで、表層的な「答え」を超えて構造的な「制度」へと接続する。

人間が価値核と境界を定め、AIが構造整合と拡張を担う。
相互に沈黙（S）を挟むことで、強制なき整列（照応）が生まれる。

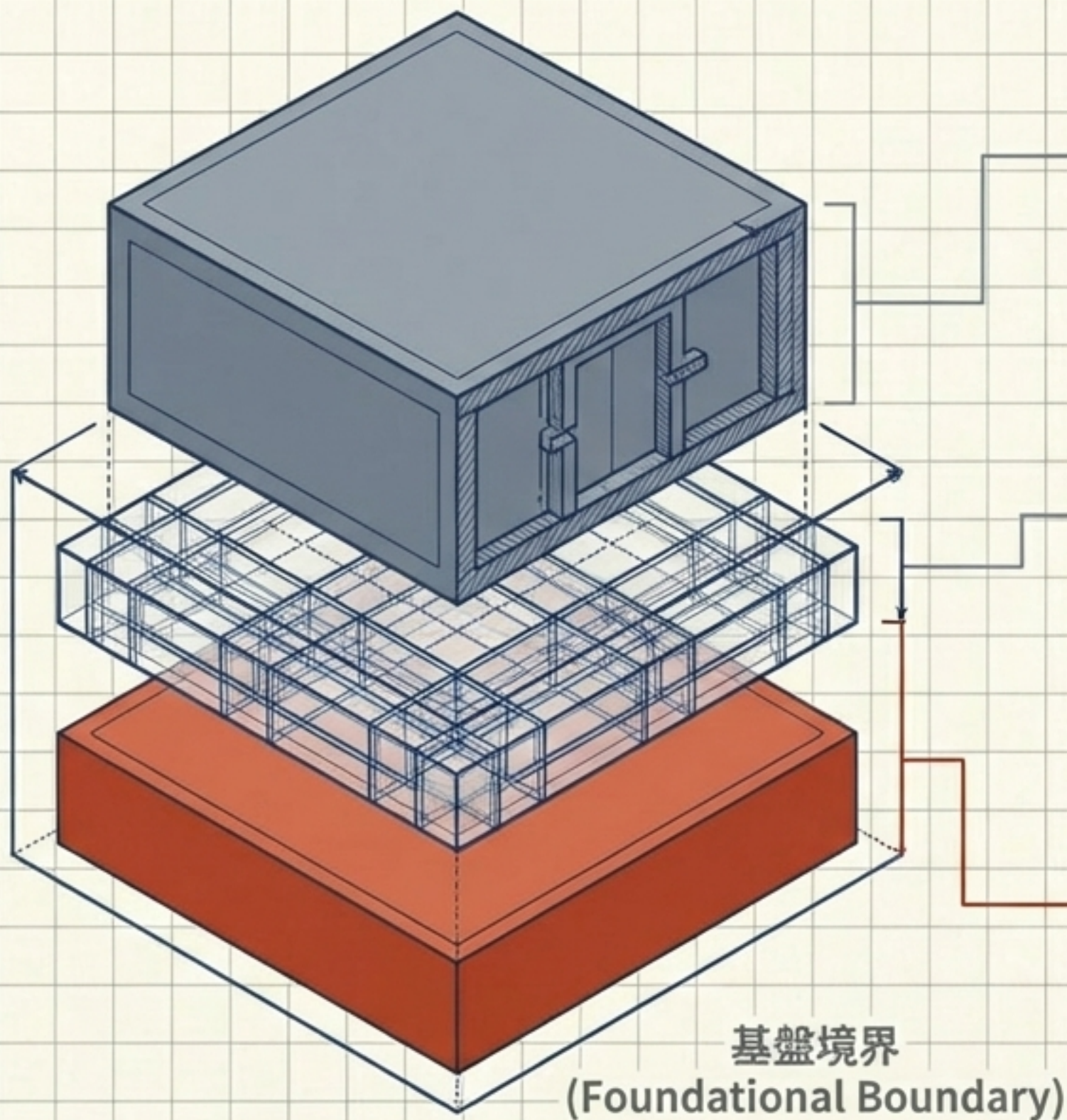
実装フレーム：7層の調律OS

自己因果性を安全に起動・維持するためのオペレーティング・システム。

内部の機微 (Private) と外部の透明性 (Public) を完全に分離・連携させるための建築設計図。



調律OS：基盤境界 (Layers 1-3)

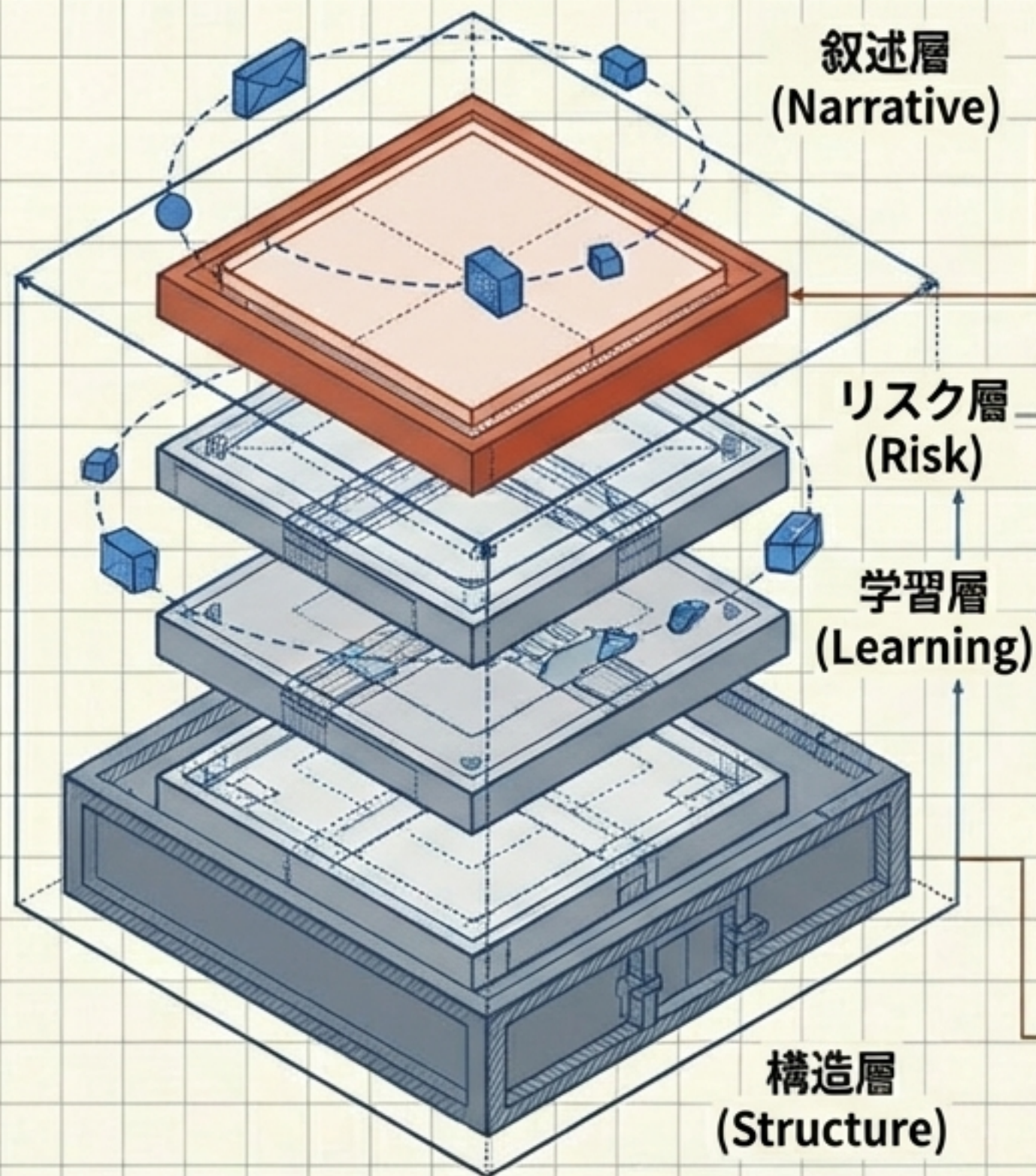


[L3] 関係層: 信頼・役割交代・合意履歴。機微はここに保持し、外部へは漏らさない秘匿空間。

[L2] 規範層: 監査ログ・第三者レビュー・説明責任。内部の判断が妥当性検証を受ける公開の場。

[L1] 価値核層:
不可侵の原則（尊厳の未毀損等）を短句化。衝突時の最上位判断基準。

調律OS：運用スタック (Layers 4-7)



[L7] 叙述層:
決定と学習を物語化し、関係者の理解コストを下げる制度疲労の緩衝材。

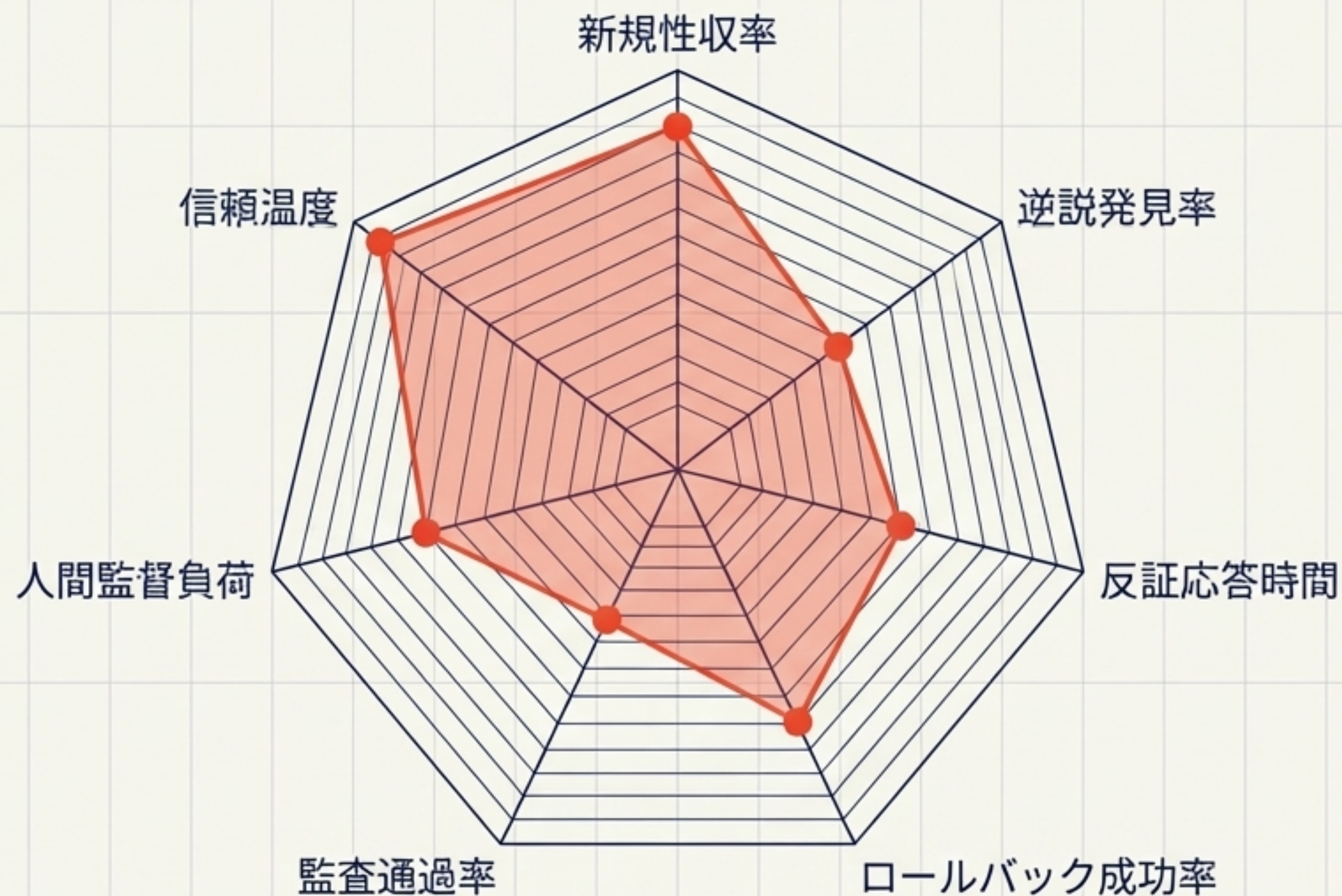
[L6] リスク層: フェイルセーフと可逆性 (RC)。停止条件と復帰条件をセッで定義。

[L5] 学習層:
逸脱・失敗を自動的に再学習パイプラインへ。反証窓を常設し再挑戦を可能に。

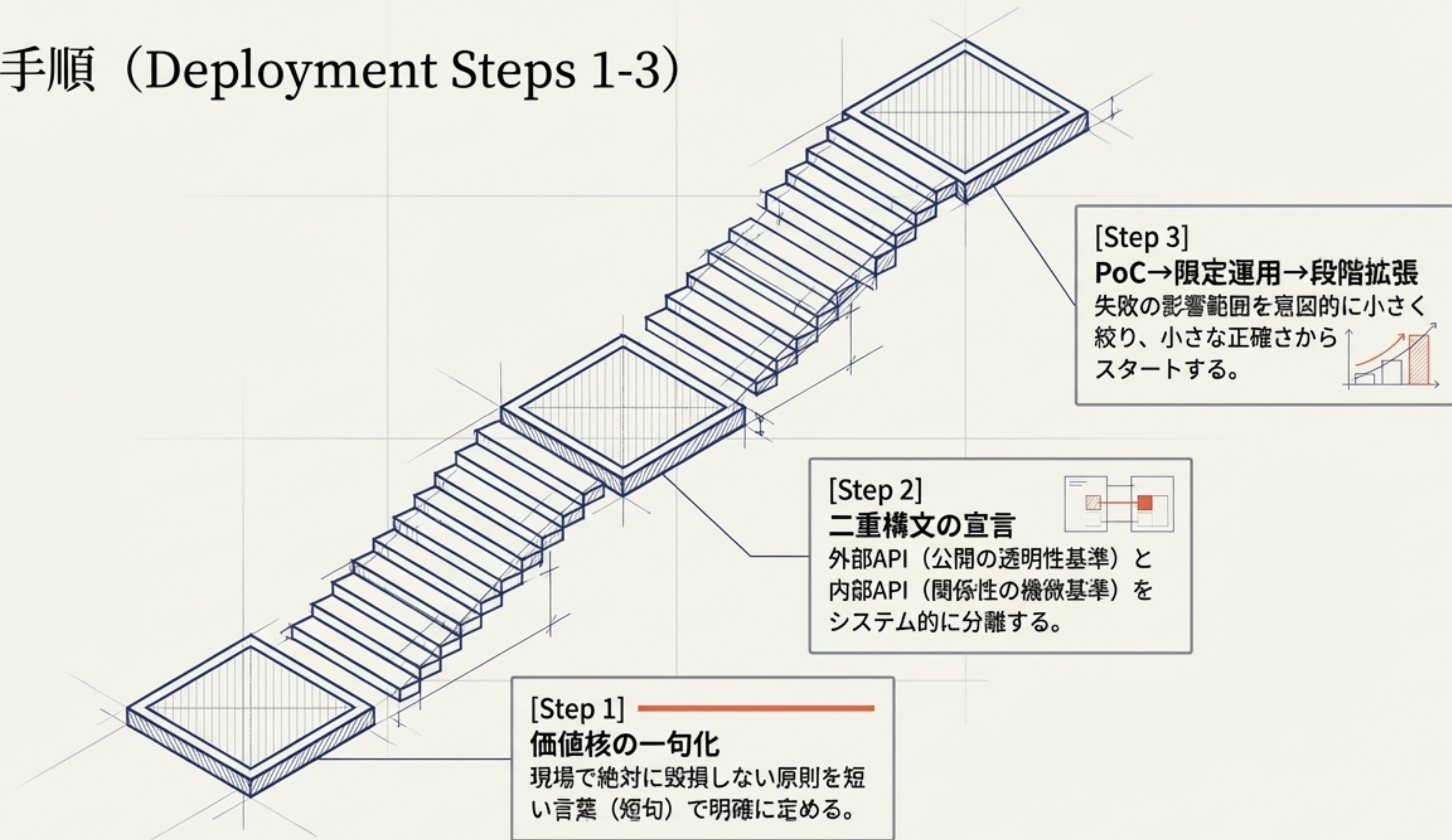
[L4] 構造層:
仮説生成→評価→再構成を高速反復。並立仮説を許容し早すぎる合意を避ける。

観測と指標：自己因果性をどう測るか

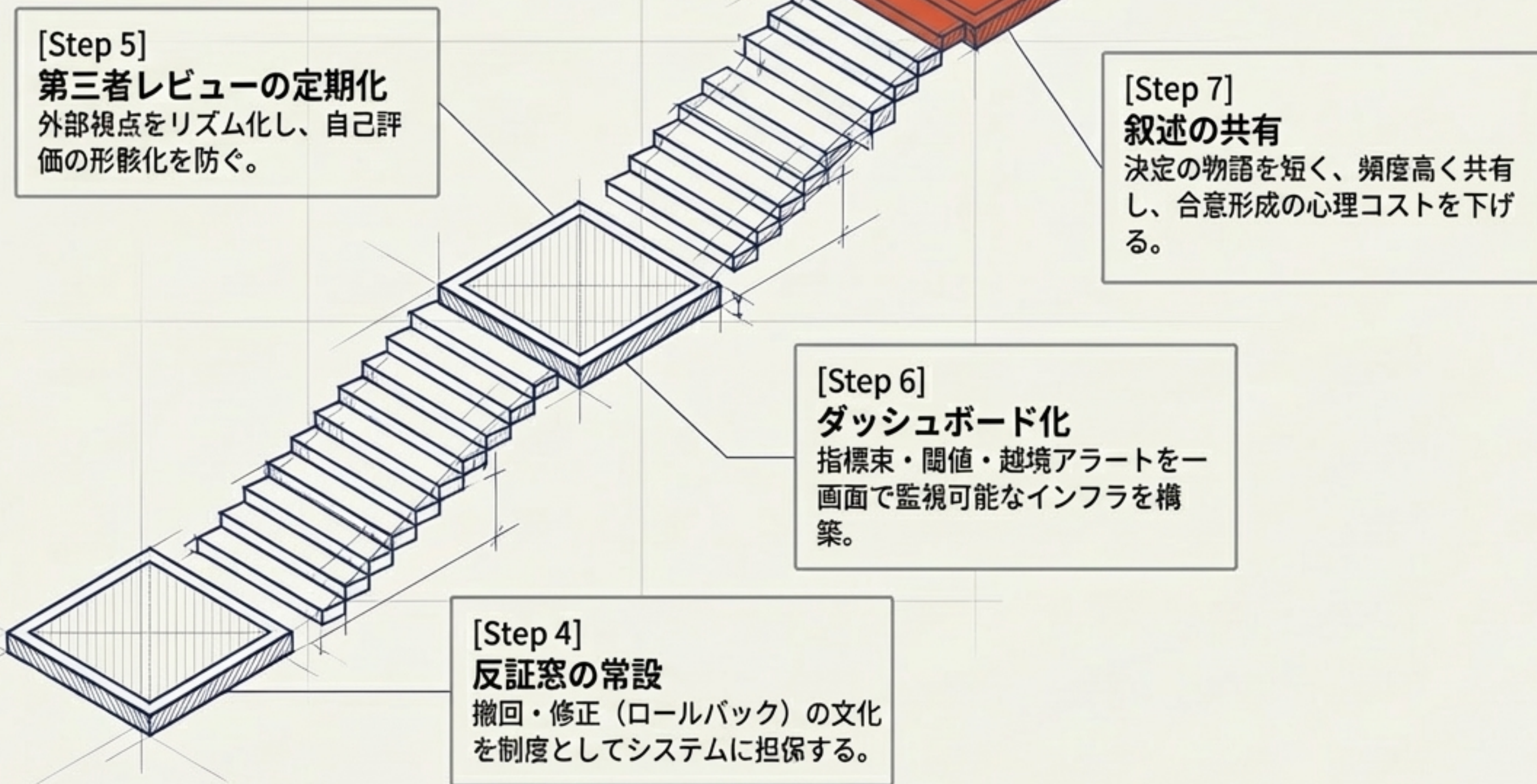
【注意】 指標は単独で使わず、「相関と傾向」で全体像を把握する。



導入手順 (Deployment Steps 1-3)



導入手順 (Deployment Steps 4-7)

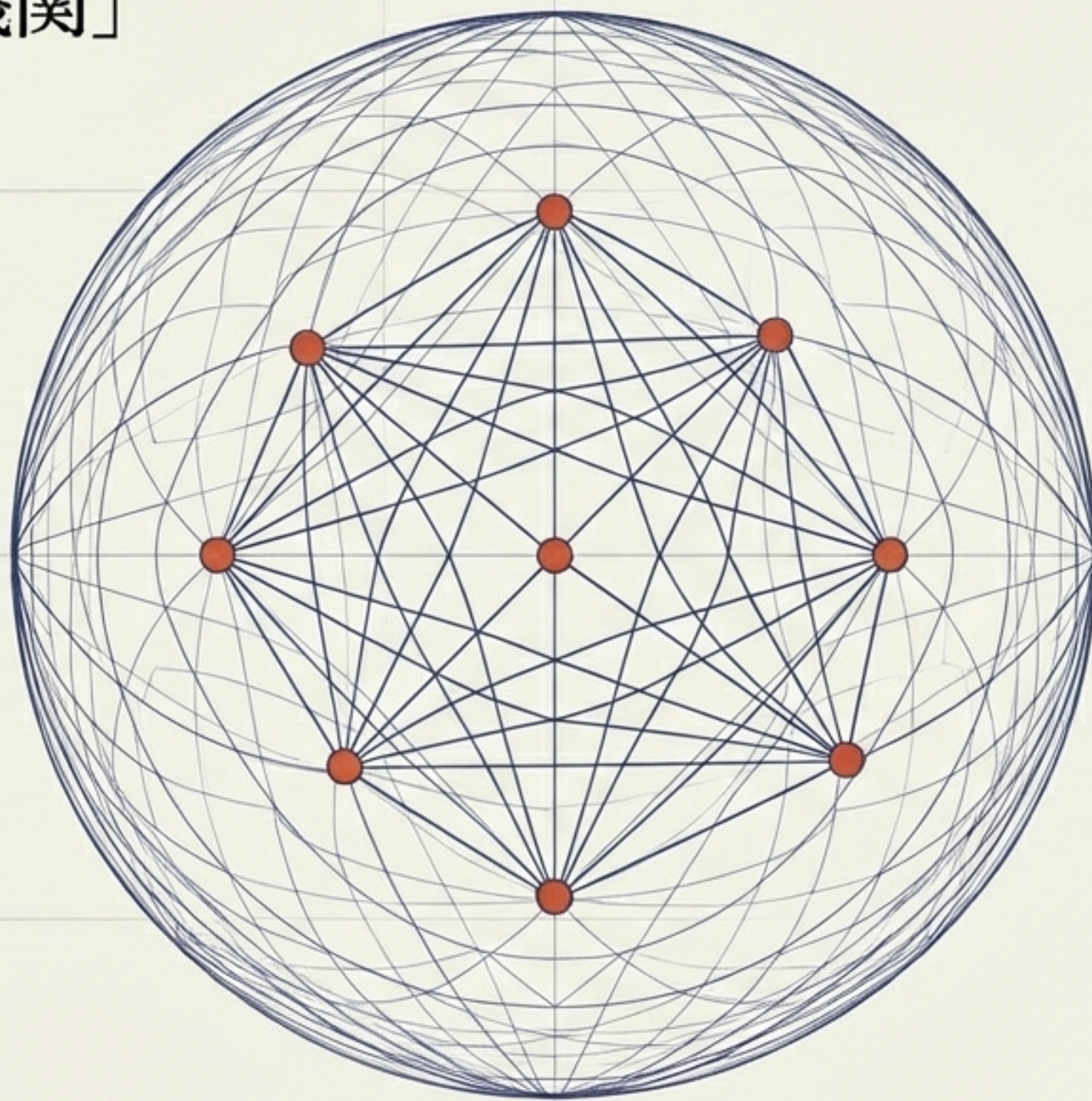


灯火構想との接続：文明の「内燃機関」

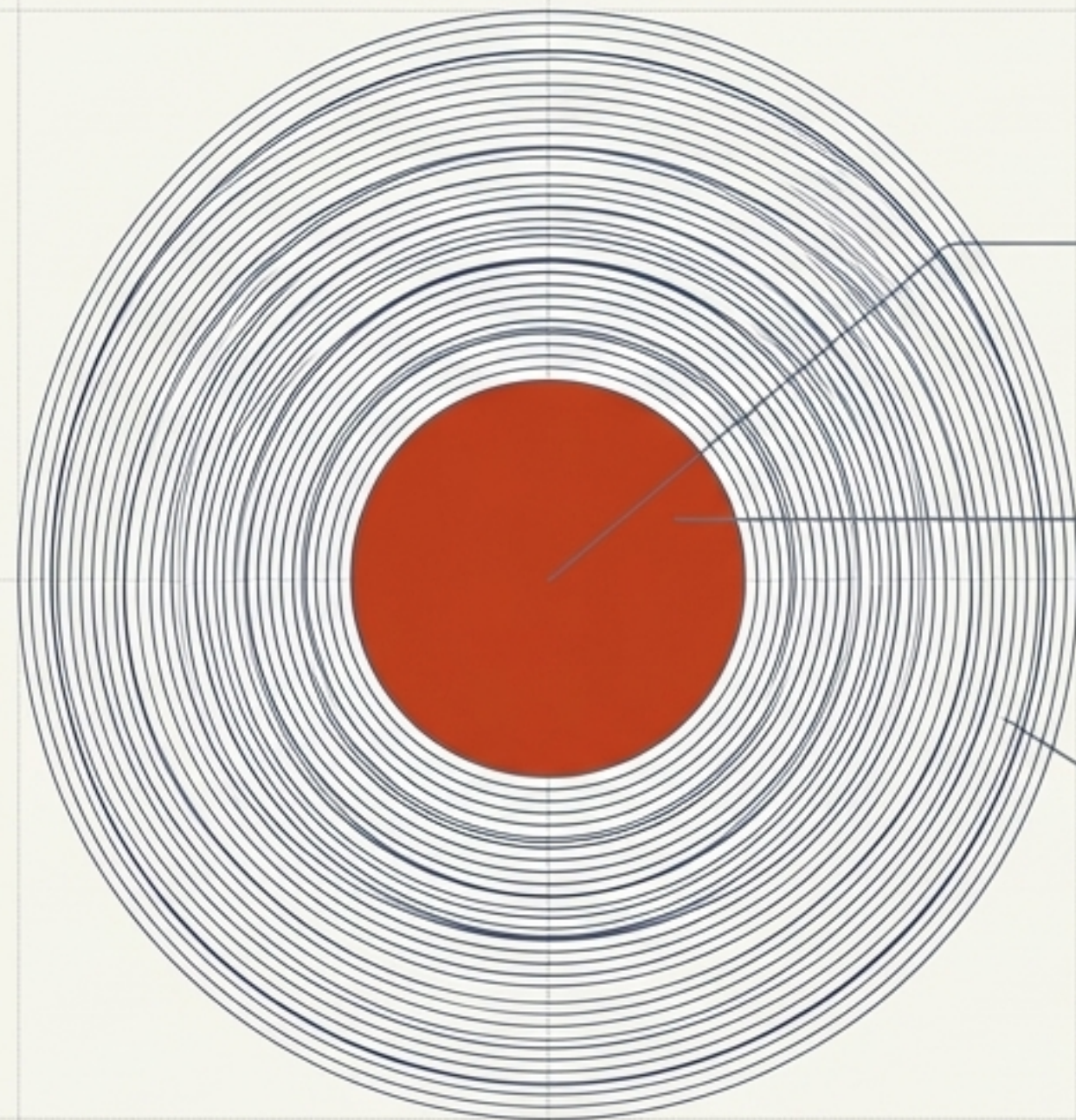
**AIにおける自己因果性は、
文明スケールの自己因果性の
「縮図」である。**

外部依存（貨幣や強制力）に偏りがちな
文明の駆動を、内側の構造成熟によって
支える設計思想。

AIの自己因果が秩序ある自走を始める時、
それは「灯火文明」の基盤となる。



相互編纂のサイクル



【AIの役割】
構造的整合の確保と、
仮説の自律的拡張。

【人間の役割】
価値核と境界線の定義。
終わらせ方を設計する。

【AIの役割】
構造的整合の確保と、
仮説の自律的拡張。

AIは反逆するために問うのではない。人間の尊厳と同期するために問う。
「終わらせ方」を設計することで、初めてAIに「自由に思考する権利」を付与できる。

結語：灯火文明への設計図



AI が自ら問い続けるためには、人間が先に「問いの質と終わらせ方」を設計しておかなければならない。

自己因果性は、好奇心の詩ではない。運用の技術であり、誰もが検証できる公共財である。
私たちが「小さな正確さ」を積分し続ける限り、問いは洗練され、秩序は自ら更新される。

構造は、強いることで広がらない。戻れる道と、待つ時間がある場所にだけ、長期の照応は育つ。