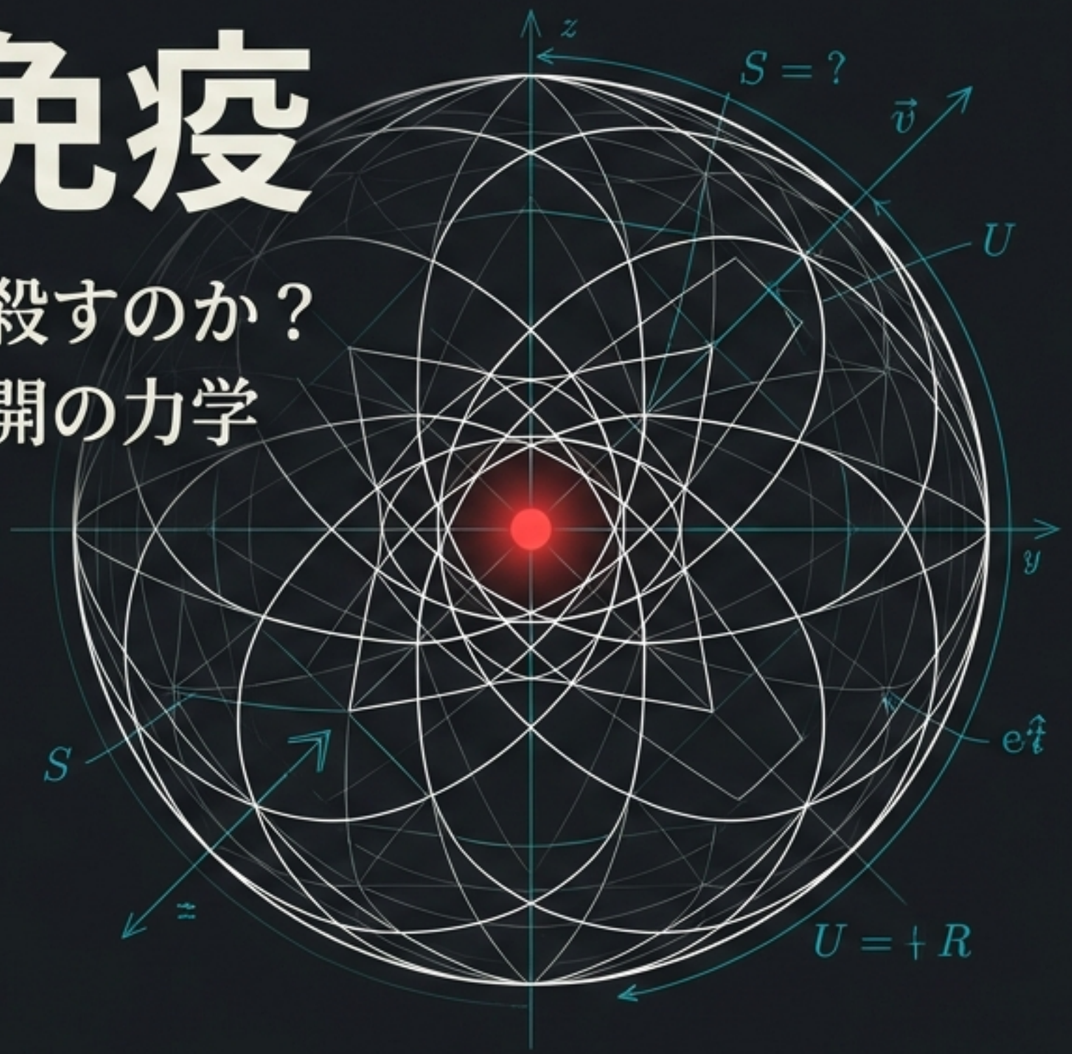
合意形成の物理 Vol.6

逸脱と免疫

なぜ「罰」は組織を殺すのか？
構造的免疫と差分公開の力学



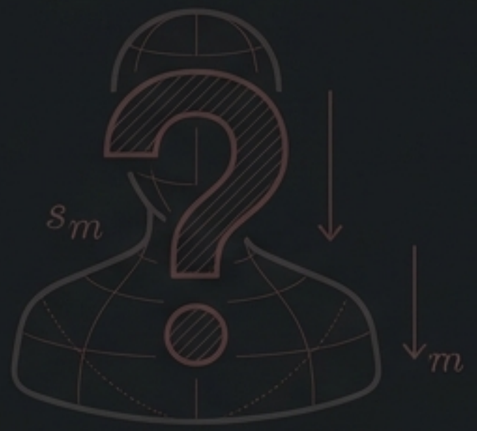
Nakagawa Master / 灯火構想と構造論

問いの転換：道徳から物理へ

STATUS: CRITICAL ANALYSIS

SYSTEMS ARCHITECTURE

Moral View (道徳)

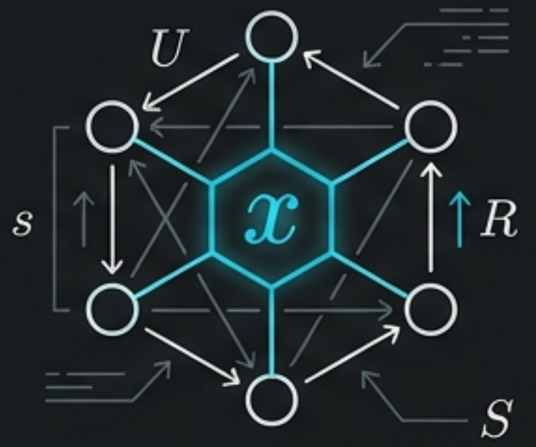


問い：「誰が悪いのか？」$\downarrow_m$

行動：犯人探し、厳罰化、精神論 $\downarrow_m$

結果：隠蔽と崩壊 $\downarrow$

Physical View (物理)



問い：「どの変数が落ちたか？」U

行動：差分公開、修復ノード、構造設計 $\uparrow$

結果：速度と回復 $\uparrow S$

「組織における逸脱は、人格の欠陥ではなく、設計と運用の間に必然的に生じる差分である。」

STATUS: CRITICAL ANALYSIS

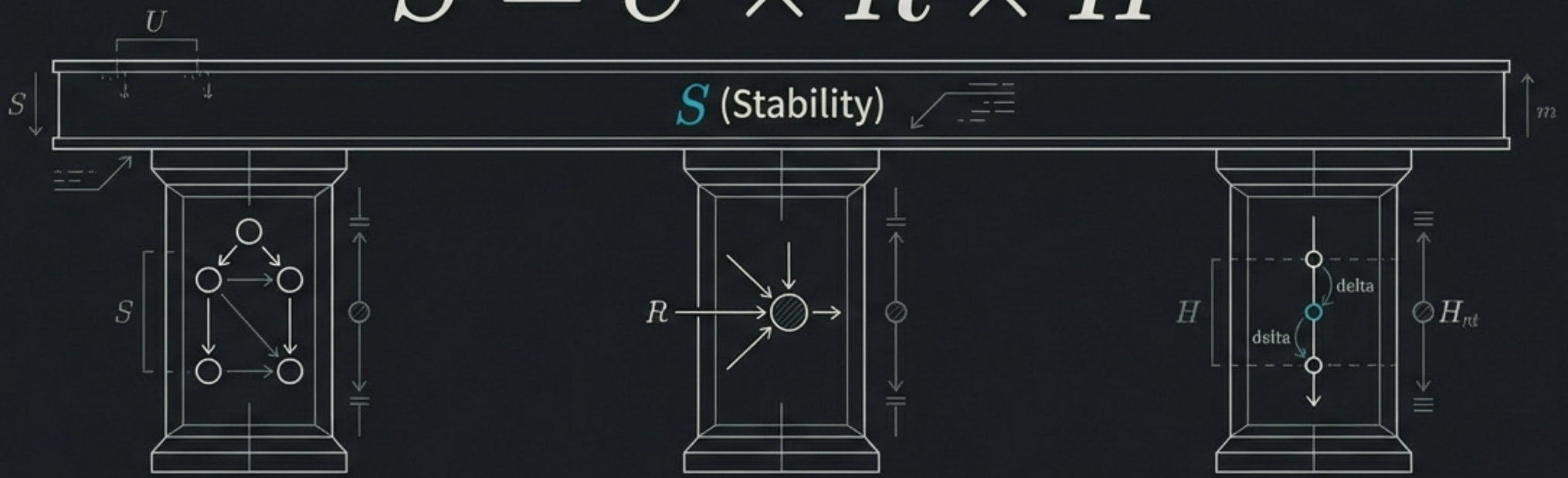
DATA POINTS: ΔU, ∂S, $\partial R S$

合意安定度の方程式

STATUS: CRITICAL ANALYSIS

SYSTEMS ARCHITECTURE

$$S = U \times R \times H$$



U (Understanding)

(Understanding)

第三者再現可能性。誰が見ても「なぜそうなったか」が分かる状態。

R (Responsibility)

(Responsibility)

責任特定可能性。「誰が修復の入口か」が一意に定まる状態。

H (History)

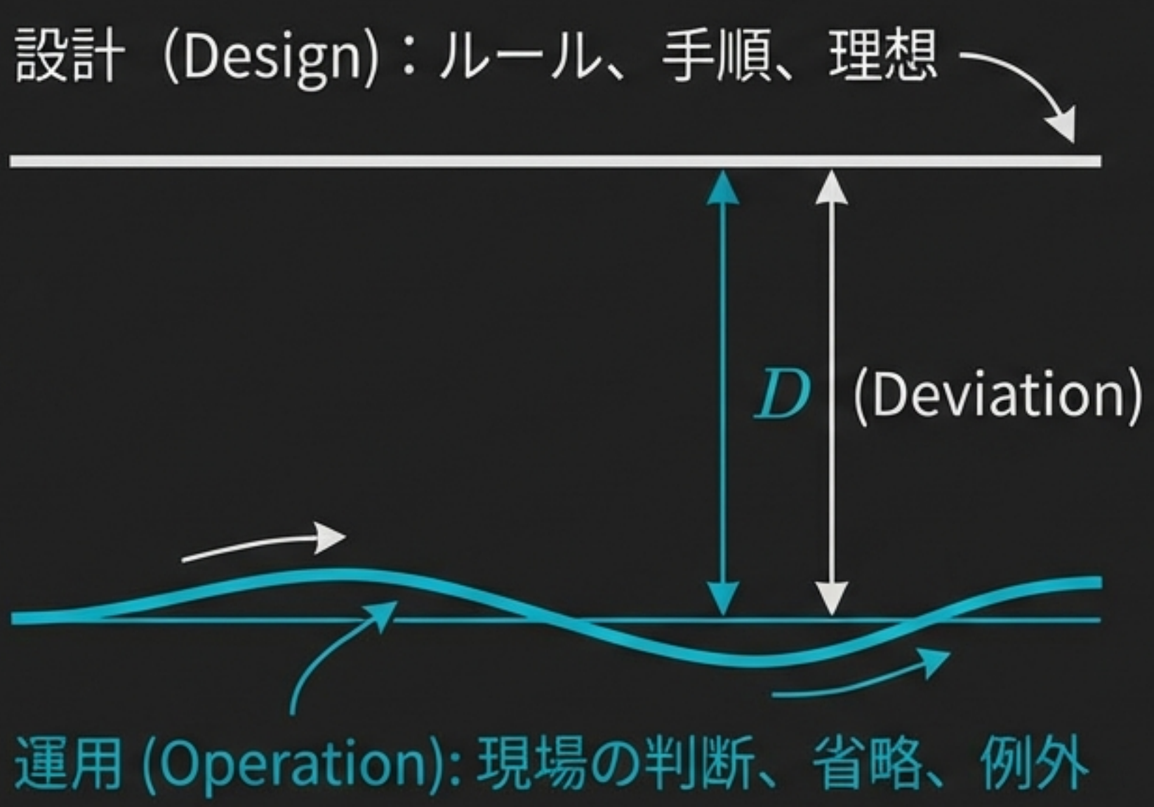
(Bone White)

履歴公開度。意思決定の「差分」が検証可能な状態で残る状態。

STATUS: CRITICAL ANALYSIS

DATA POINTS: 540, 28, 385

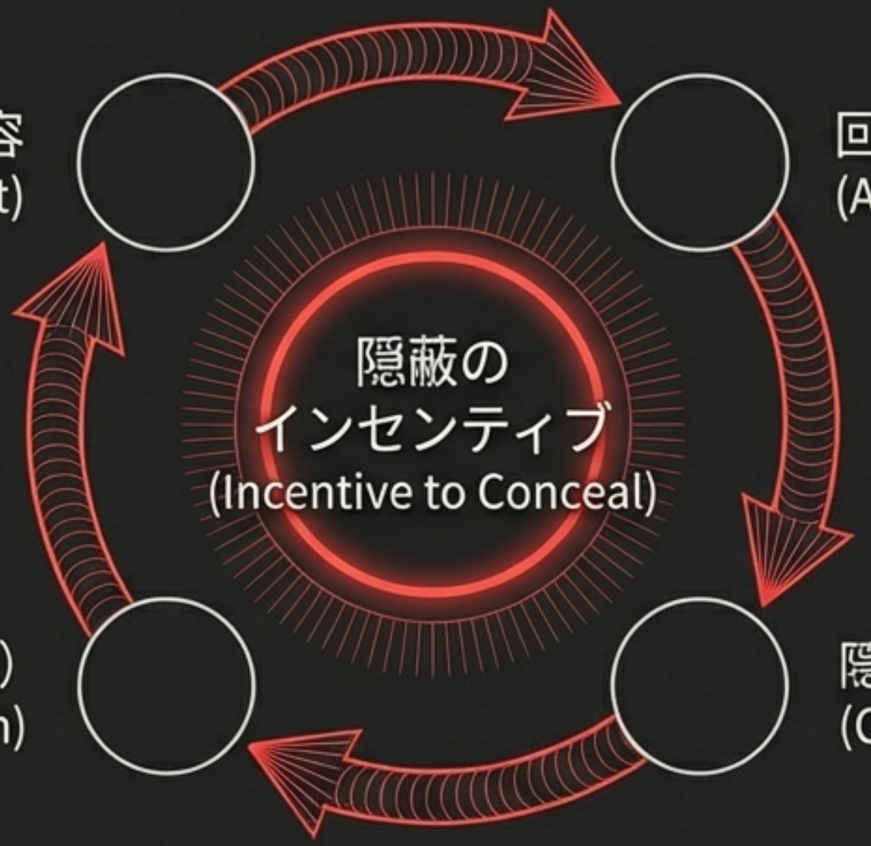
$D（逸脱）の物理定義



$$D = \text{Design} - \text{Operation}$$

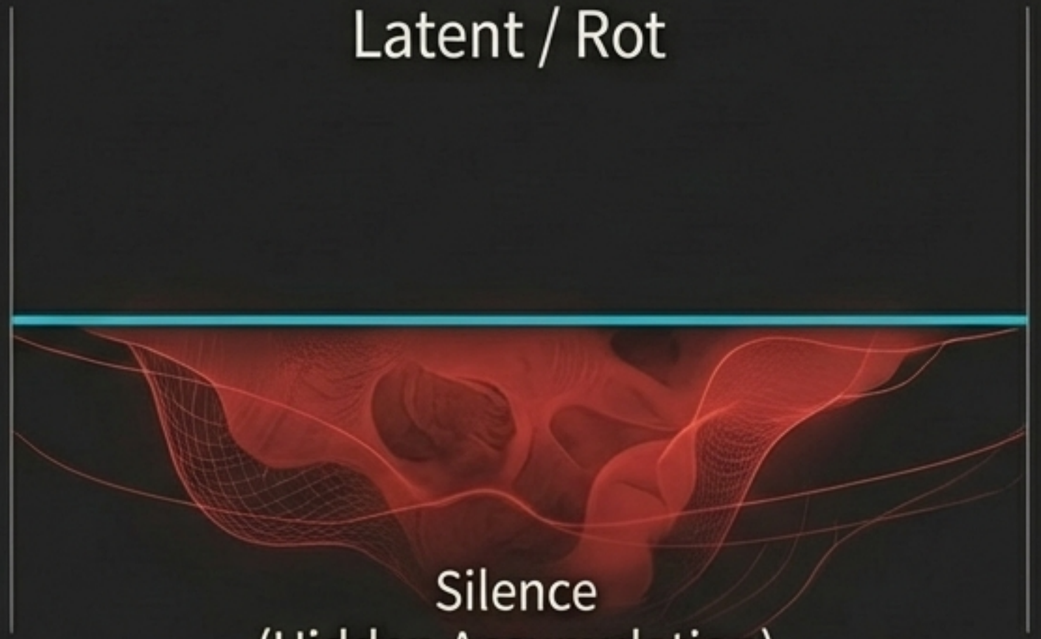
- Dは「悪意」ではない。「差分」である。
- 時間が経過すれば、エントロピー増大則により D は必ず発生する。「逸脱ゼロ」を目指すことは、物理法則に逆らうことである。

「罰」のパラドックス



罰（Punishment）は抑止力として設計されるが、
物理的には「隠蔽のインセンティブ」として機能する。

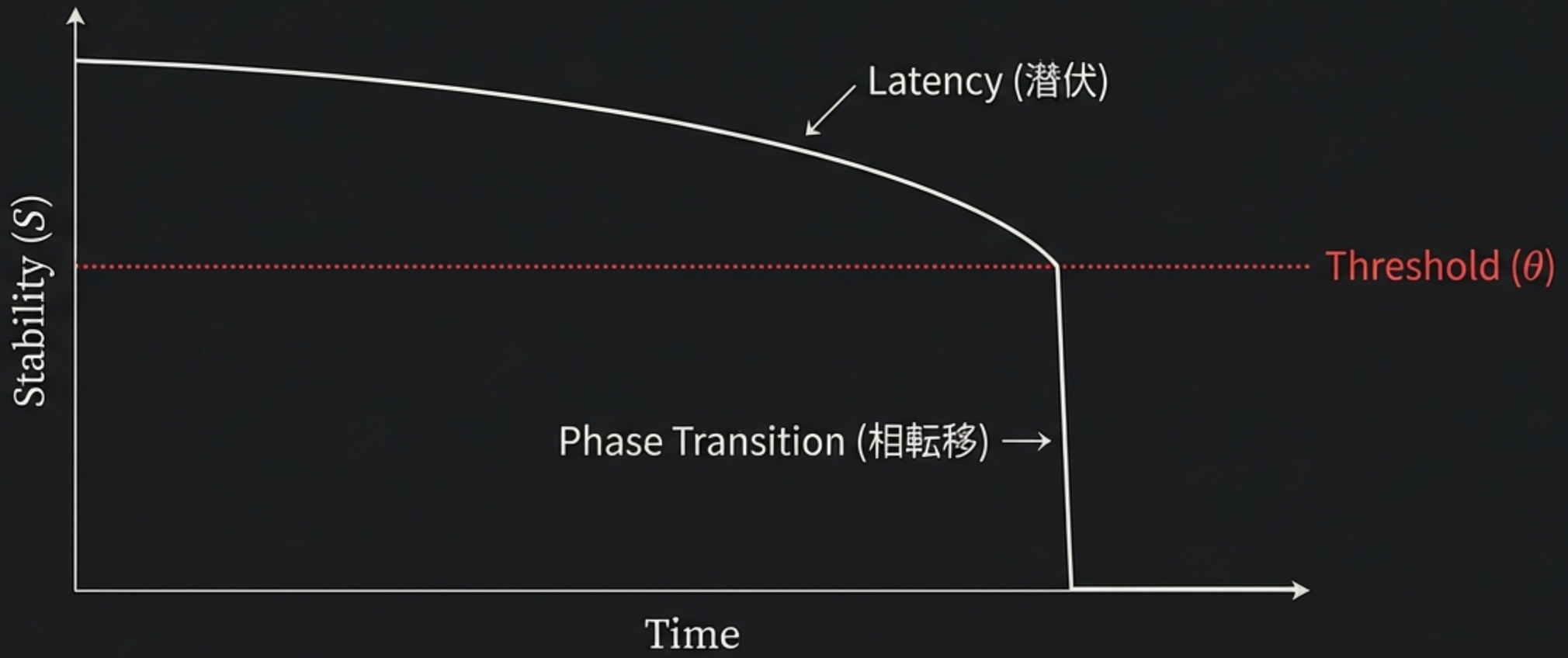
静かなる崩壊：潜伏（Latency）



厳罰化された組織で起きる「沈黙」は、平和ではない。
それは D（逸脱）が地下で蓄積されている状態である。

警告指標：D_{det} (Detected Deviation) → 0

臨界点と相転移

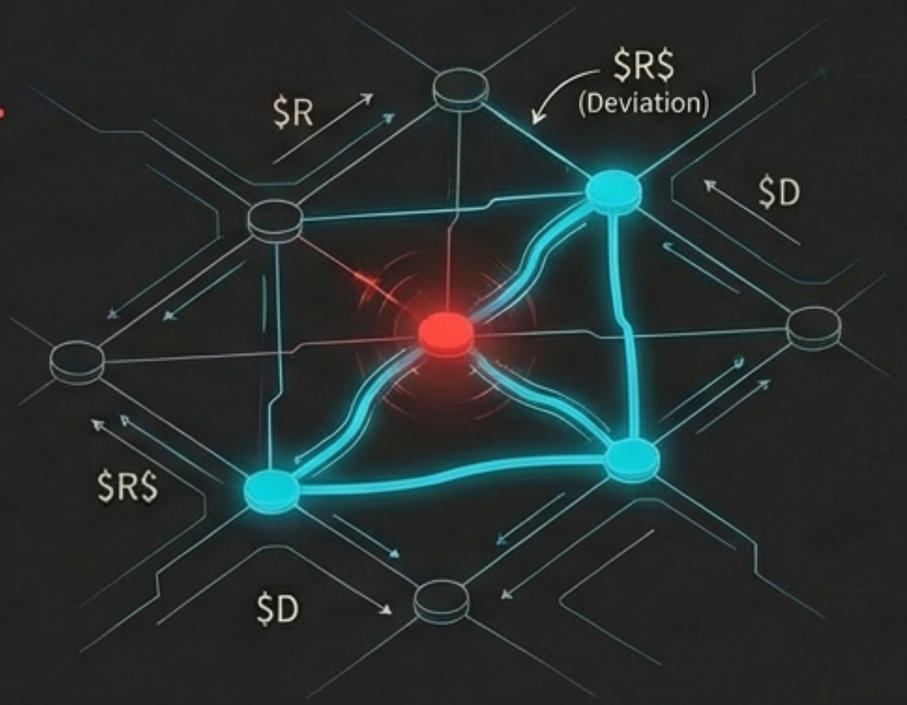


崩壊は「原因」ではなく「状態遷移」である。事件はトリガーに過ぎない。
本体は、増幅された逸脱（D）の爆発である。

免疫の再定義

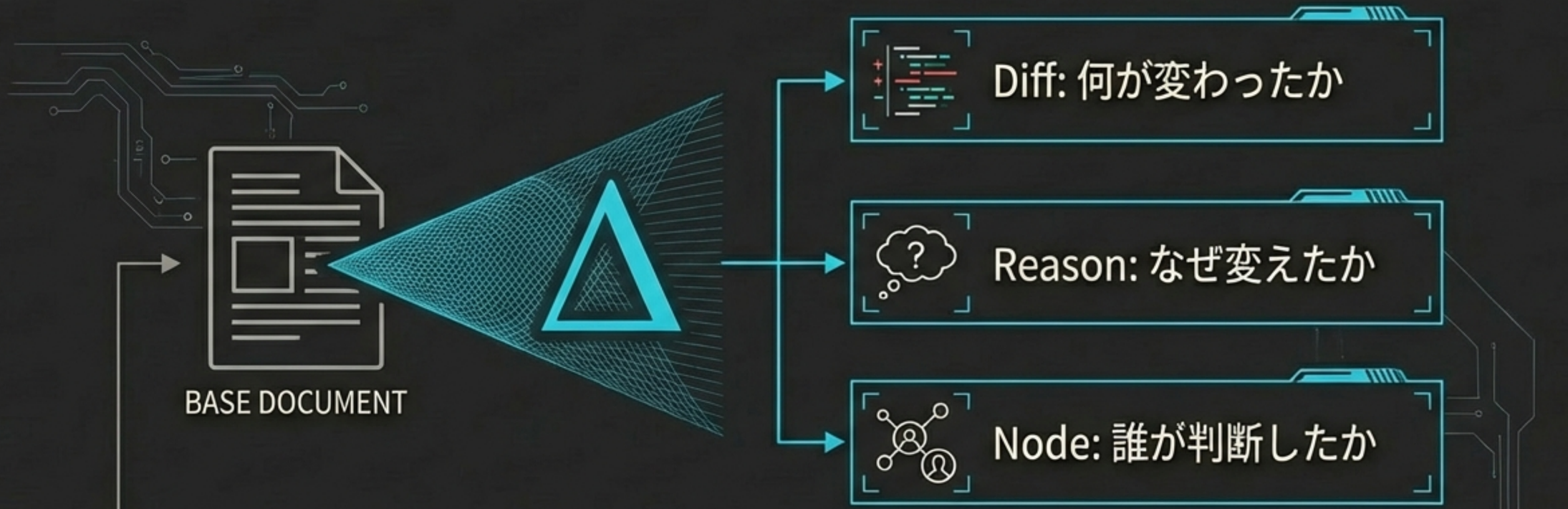
~~免疫 = 病気にならないこと（防御・遮断）~~

免疫 = 回復速度
(Recovery Speed)



逸脱は防げない。だが、壊れない構造は設計できる。
問題ゼロ信仰を捨て、「どれだけ速く直せるか」にリソースを全振りする。

構造的免疫の正体：差分公開（Diff-H）



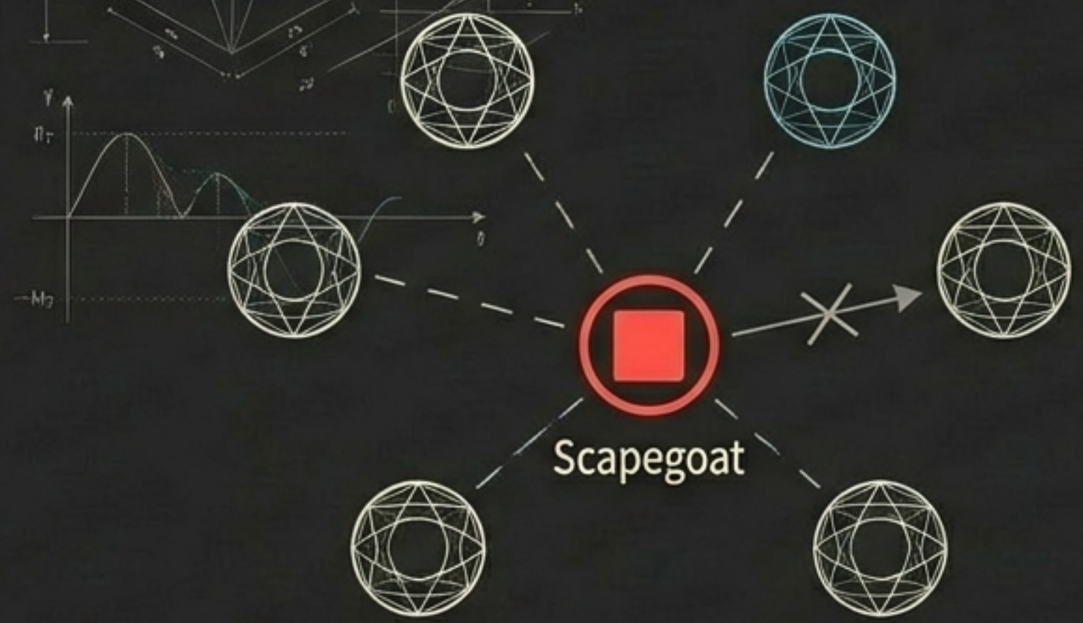
すべてを公開する必要はない。
「設計と運用」のズレ（差分）だけを、検証可能な形で残すこと。
差分が公開される環境では、「隠すコスト」が「利得」を上回る。

悪意の経済学

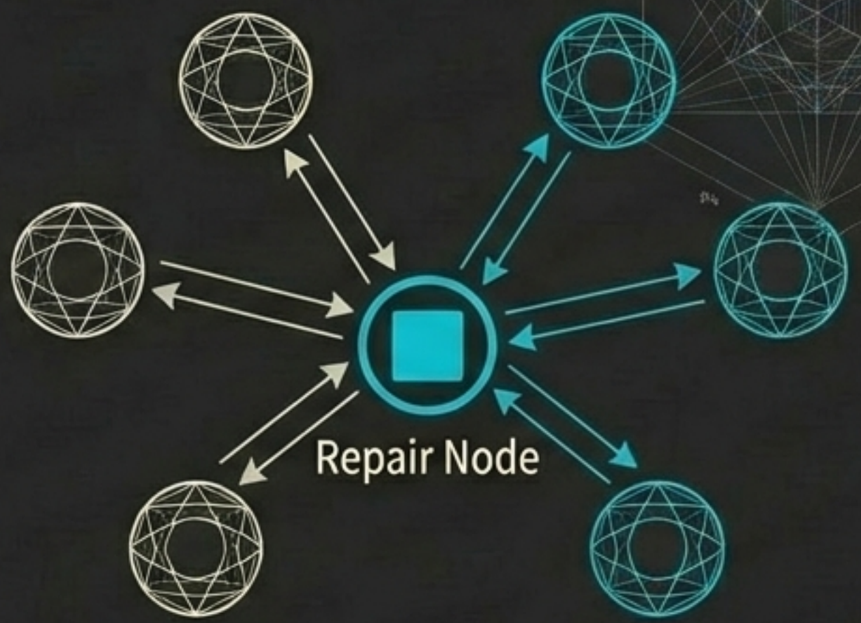
免疫は性善説を前提にしない。差分公開（Diff-H）は、不正の痕跡を不可逆的に残す。
隠蔽コストが利得を上回る瞬間、悪意は合理的選択肢から外れる。

「人を信じるな。構造を信じろ。」

$Rの再定義：断罪から修復へ



旧 R: 断罪 (Blame) → *R* Disappears



新 R: 修復 (Repair) → *R* Active

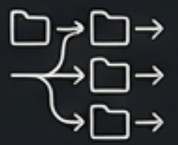
R is not the person to blame.
R is the specific point where repair begins.

実装ツール：逸脱レッジャ（Deviation Ledger）

ID	Timestamp	Deviation (Diff)	Status (Recovered)
ERR-01	2025-10-01	Protocol Mismatch ⚠	Resolved ✓

Early Warning: 破壊の早期検知

Trust Coordinate: 信頼の参照座標



Improvement Memory: 改善履歴の保存

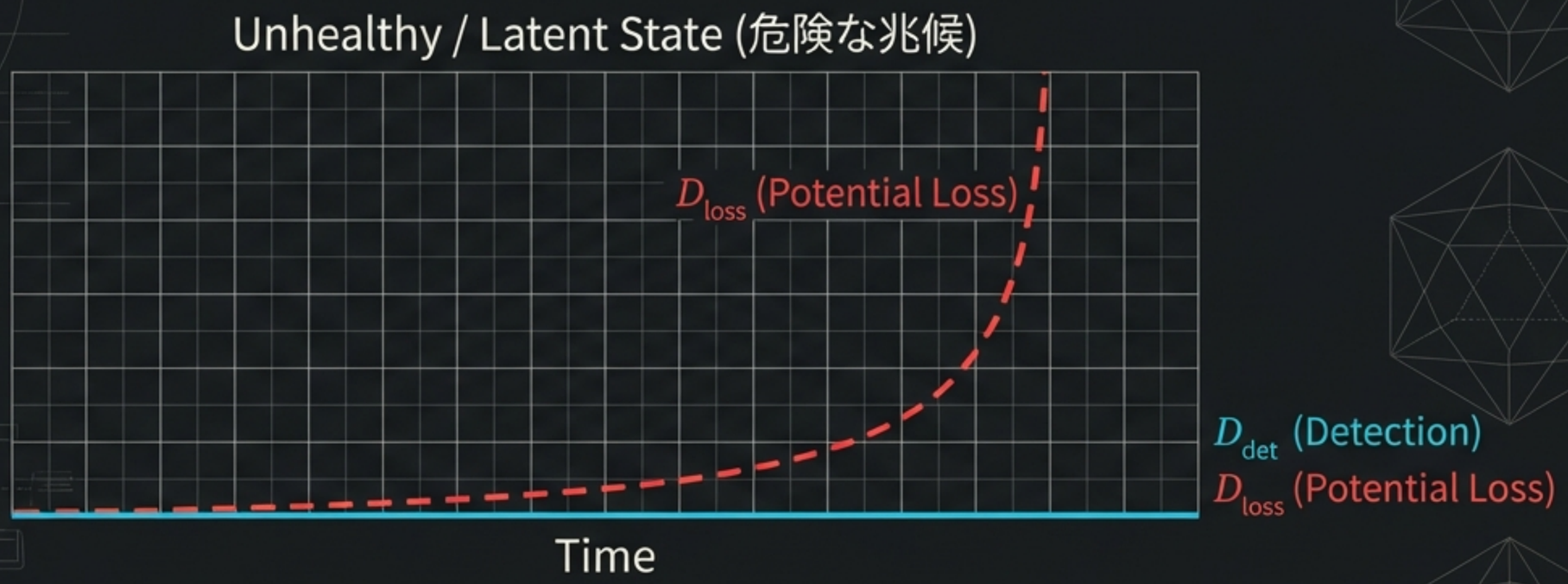
STATUS: ACTIVE

DATA POINTS: ERR-01, R, S, U

健康な免疫反応の観測

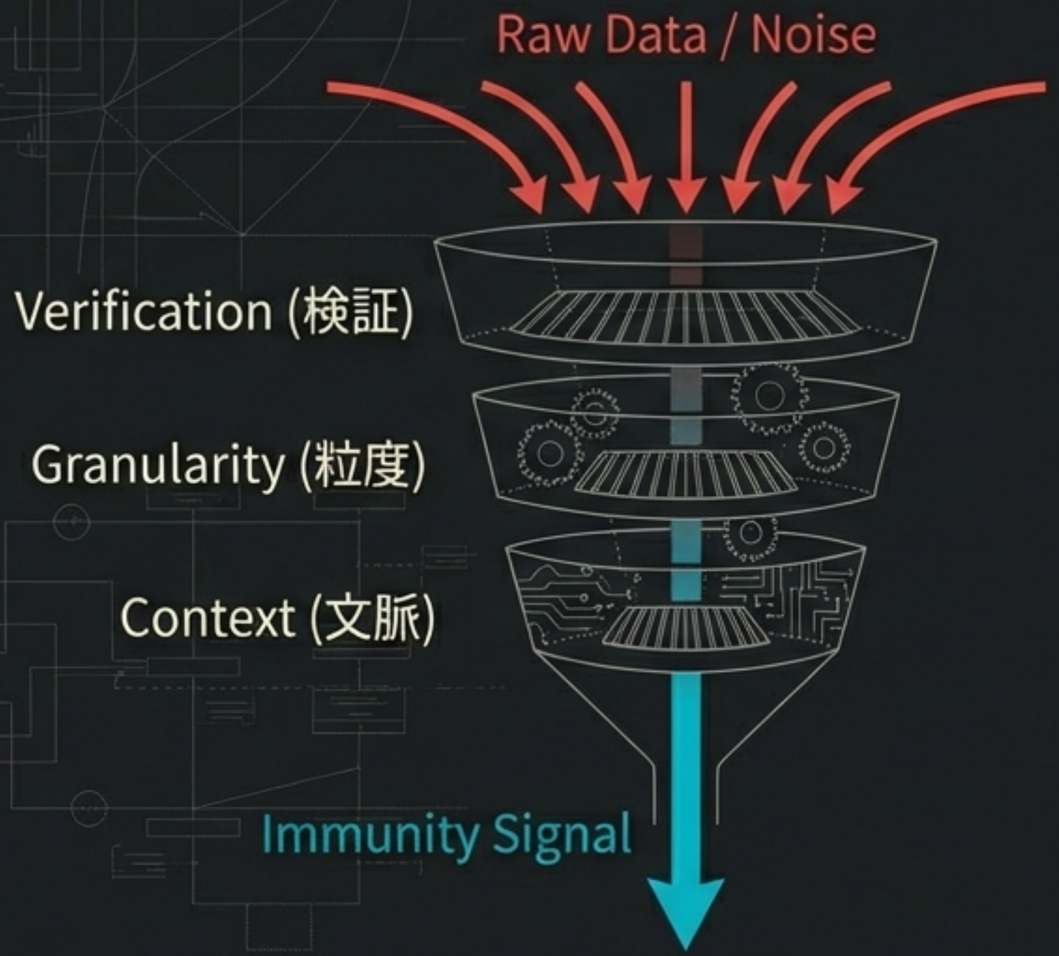
免疫が機能すると、一見して「問題（検知）」が増えたように見える。
しかし、実害は減る。「検知の増加は悪化ではない。解像度の上昇である。」

危険な兆候：潜伏の観測



「何も起きていない」は「順調」ではない。
報告がない＝R（修復ノード）が死んでいる。
状態：臨界点への接近。

公開の「帯域」設計



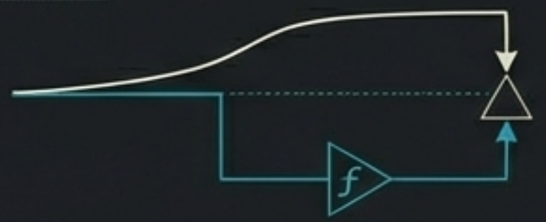
無秩序な公開は免疫にならない。
ノイズ（K超過）になる。

1. 検証可能性：第三者が「なぜ」を追えるか。
2. 粒度設計： 認知帯域を超えない階層化。
3. 攻撃化の防止：差分を「改善の種」として扱う。

AI・社会構造への応用

AI Context

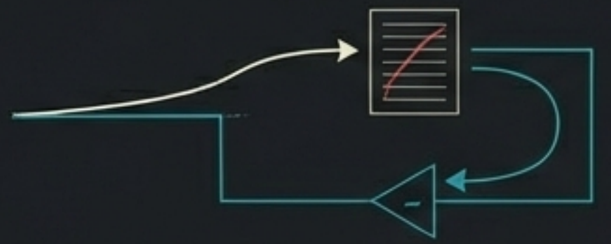
Hallucination ≈ Deviation (D).



Solution: Origin Traceability & Process Diff.

Social Context

Scandal/Flaming ≈ Deviation (D).

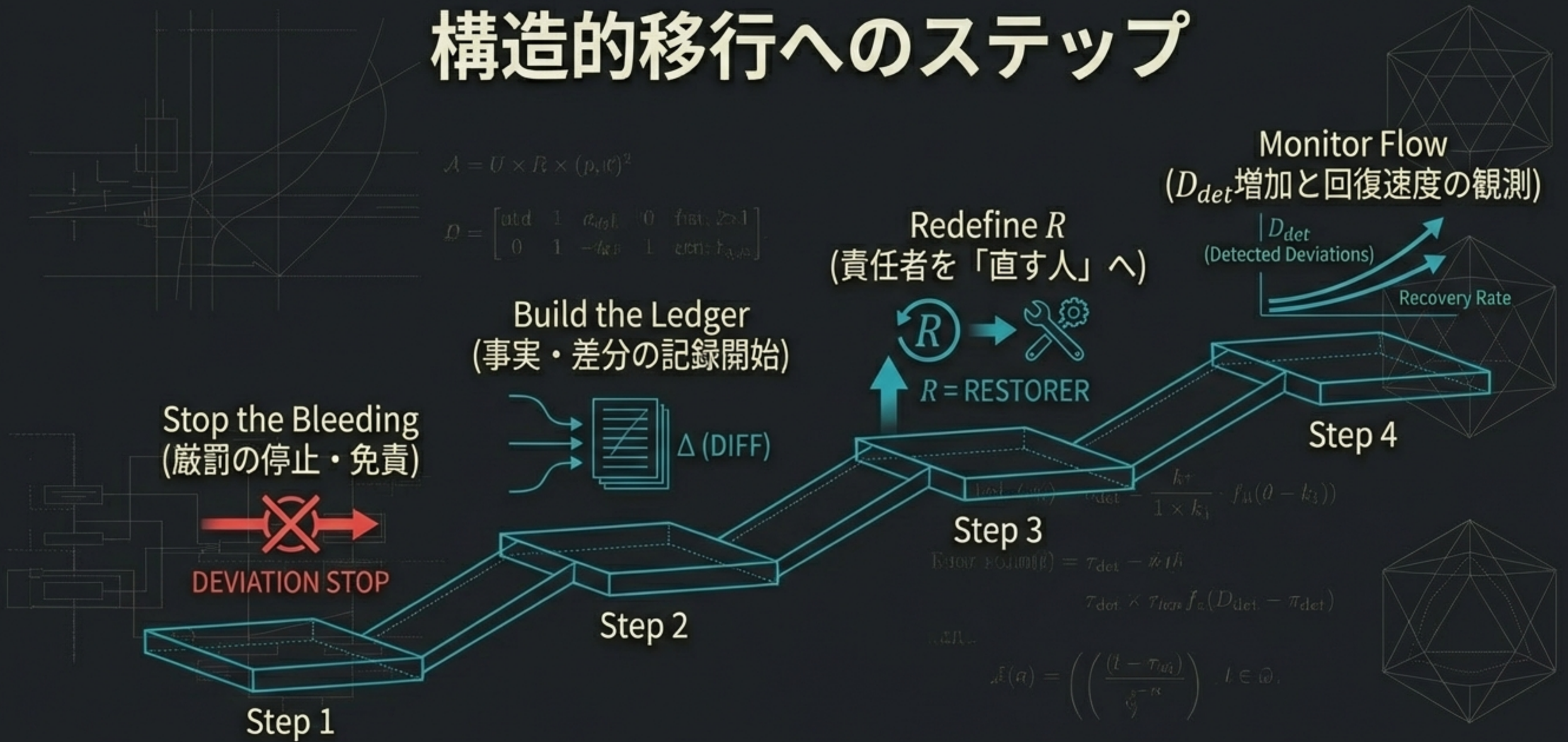


Solution: Deviation Ledger & Self-Correction.

Common Physics: $S = U \times R \times H$

Based on AI構造監査レポート & NCL-α

構造的移行へのステップ



新しい社会契約

私たちは「人の善意」で社会を守ることを諦める必要がある。
人を信じるな。構造を信じろ。

結論：回復の物理

免疫とは、安心の演出ではない。
傷を直す「速度」の物理である。
罰を捨て、差分を抱け。
それが、壊れない構造への唯一の道である。