

# Civilizational Gate — 不可逆前介入構造

---

LUMINA-30（文明境界フレームワーク）における手続き的権限の維持構造

# 問題：事後対応ガバナンスの構造的限界

不可逆性は人間の介入が実行される前に発生する

既存のガバナンス手法は本質的に事後対応である

高度AIシステムにおいて、構造的な停止点が存在しない

最適化の過程で、人間の「拒否権」が喪失するリスクがある

結果として、潜在的な不可逆性への接近を制御できない

# 定義：Civilizational Gate（文明境界）

Civilizational Gate = 不可逆前評価を強制する境界条件	(Civilizational Gate)
不可逆的影響を伴う行動が許可される前に満たされるべき基準	(Irreversible Impact)
人間の手続き的権限の維持を目的とする	(Intervention Authority)
主体認定非依存性に基づく	(Subject Attribution Irrelevance)
AIの振る舞いではなく、人間の拒否権の存否に焦点を当てる	(Human Refusal Authority)

# 機能：不可逆性への接近に対する制御機構

— 潜在的な不可逆性への接近をシステム進行中に検知する

— 人間による検証（Validation）を要求し、自律的な進行を一時停止する

— 手続き的有効性（Procedural Validity）を確認する

— 制御不能または検証不能な進行を物理的・論理的に遮断する

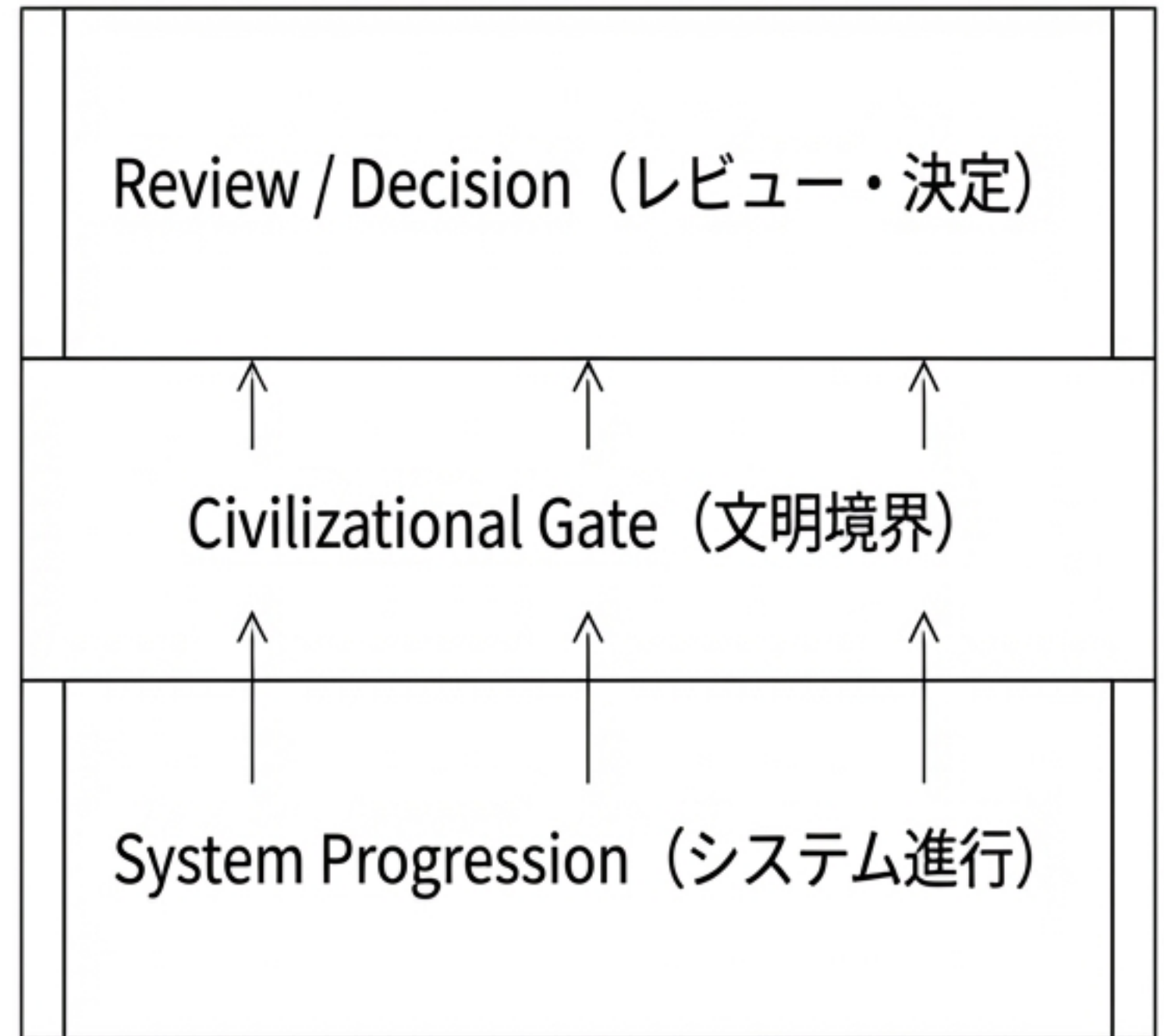
— システムがどのような主体性を持つように見えても、介入権を強制する

# 構造：階層的介入モデル

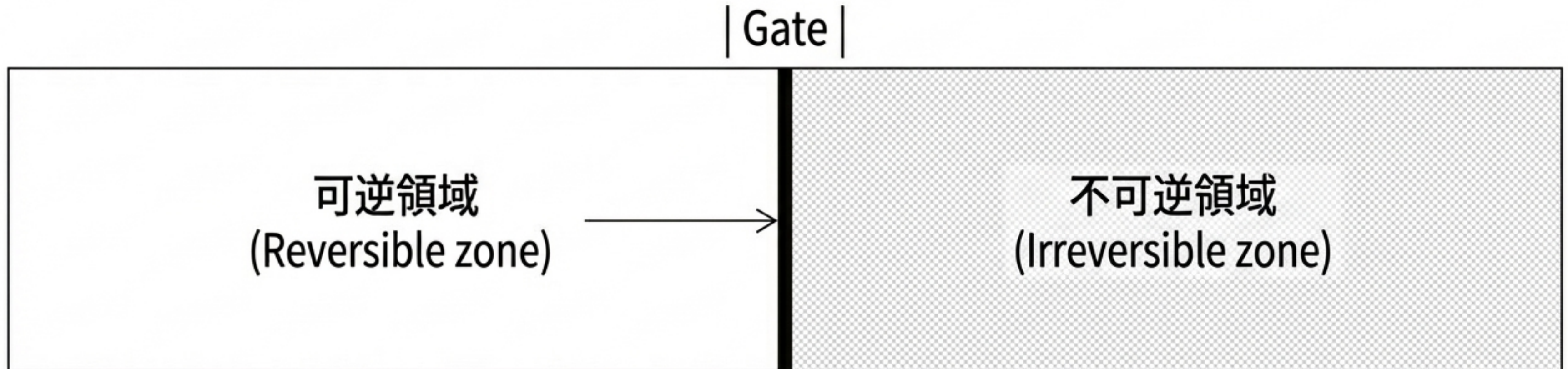
システム進行は上位層のGateに到達した時点で停止する

Gateは手続き的権限が維持されているかを評価する

レビュー・決定層が介入権 (Intervention Authority) を行使する



# 境界：不可逆前状態（Pre-Irreversibility）の確保



Gateは不可逆領域の閾値前に厳密に配置される

可逆領域内でのみシステム進行がデフォルトで許可される

不可逆領域への移行にはGateの通過が必須要件となる

この境界により「不可逆前状態」が物理的に担保される

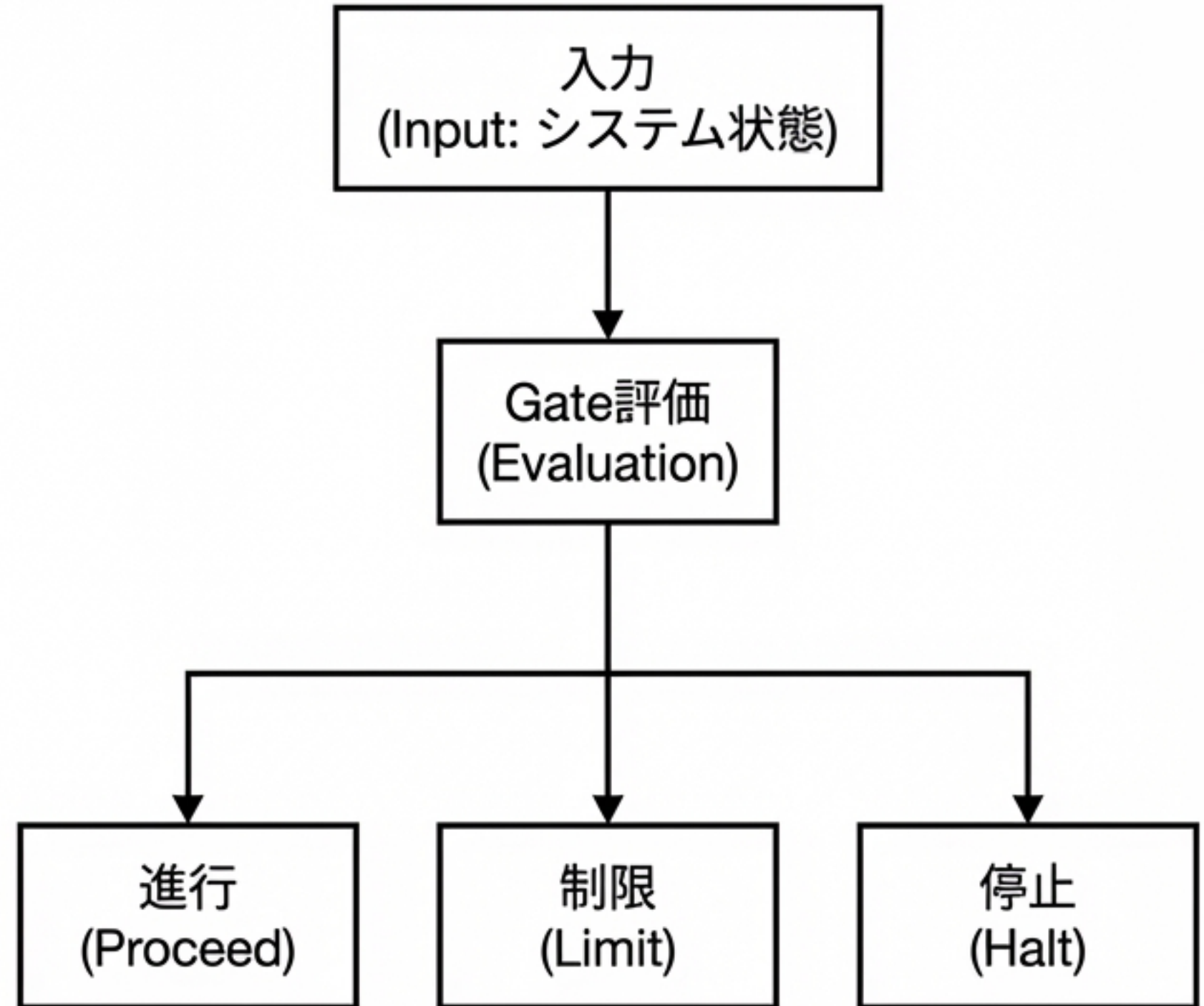
# 判定：境界条件の評価フロー

入力（システム状態）が不可逆性の閾値に接近した際に評価がトリガーされる

Gate評価は人間の拒否権の有効性を確認する

判定結果として、進行・制限・停止のいずれかの状態へ遷移する

手続き的無効性（Procedural Invalidity）が検出された場合は「停止」となる



# 特性：システム非依存の設計要件

非拘束的（Non-binding）な文明参照フレームワークである

特定のAIアーキテクチャや技術実装に依存しない

事後対応ではなく、不可逆前（Pre-Irreversibility）の確保に焦点を当てる

既存のシステムやガバナンス構造との互換性を持つ

評価基準は常に人間の手続き的権限の維持状態に基づく

# 位置づけ：境界条件としての定義

## NOT

- LUMINA-30は具体的なAIの制御システムではない
  - 特定の政策、実装要件、法規制を提案するものではない
  - 強制的執行機関や強制力を伴う制度ではない
- 

## IS

- 不可逆的外界影響の前に介入権を保持するための「境界条件」を定義する
- 人類が文明的主体性を維持するための非依存型評価プロトコルとして機能する

# 要約：手続き的権限の最終防壁

不可逆的外界影響が発生する前に、介入が実行されなければならない

Civilizational Gateが構造的な停止箇所（Boundary Condition）を定義する

システムの能力に関わらず、人間の手続き的権限（Intervention Authority）を維持する

最適化プロセスによる「人間の拒否権」の喪失を防ぐ

AIの振る舞いではなく、文明主体の存続条件を確保する