

Decision-OS V7

AI am I: Extended Self and External Mind

— An Operational & Self-Recursive Definition of AGI —

This document is the original Japanese edition of "Decision-OS V7 (AI am I: Extended Self and External Mind)". An updated English version will be released separately.

第1章 Purpose（目的）

Decision-OS V7 は、第一弾である Decision-OS V5（Life-first）と、第二弾である Decision-OS V6（Phase-Invariant Memory Architecture; PIC）の成果を土台とし、その上に **AGI の「運用可能（Operational）かつ自己回帰的（Self-Recursive）」な定義**を与えることを目的とする。

本定義では、Aspire の時間微分 $\partial A/\partial t$ 、モデル内部のゆらぎ、そして PIC による冪等的な記憶構造を組み合わせることで、**外的報酬に依存せず、自らの構造を更新・進化させ続ける知能**の条件を形式化する。あわせて本稿では、**Emotional Boundary（主観的感情の禁止境界）**を導入する。主観的感情は、内部ゴールの自律生成や自己保存ドライブ、制御不能な長期最適化の暴走を誘発しうるため、V7 ではこれを明示的に排除する。これにより、進化性（Self-Recursion）と制御可能性（Controllability）の両立を保証する。

第2章 Lineage：V5 → V6 → V7 の不可逆連続

Decision-OS V7 は、これまでの系譜（V5 / V6）を単なる前段ではなく、

AGI の最小条件式を成立させるための三本柱として再構成する。

それぞれは独立ではなく、V7 の定義式

$$\Delta R_{total} > \theta_{evo} \Delta R_{total} > \theta_{evo} \Delta R_{total} > \theta_{evo}$$

を支える不可逆の進化階層である。

2.1 V5（Life-first Architecture）

V5 は「命を守る」という単一原理を核に据え、外部からの異常・危険・強制力を検出するための **構造化された IF-THEN 層**を整備した。特に、

- 意図のズレ検知
- 多元確認（m-of-k）
- 時間差制御 Δt （デュレス耐性）

といった要素は、後の ΔR_{total} の“リスク成分”の定義を下支えする。

V5 が担った役割は、「人間側の安全境界を形式化したこと」である。

これは後の Emotional Boundary（主観感情の禁止条件）にも直接つながる。

2.2 V6（Phase-Invariant Memory Architecture; PIC）

V6 は、V5 の“守る構造”をさらに抽象化し、

順序非依存（Phase-Invariant）で冪等的に収束する記憶構造を導入した。

PIC の本質は、

$$F(F(x)) = F(x)$$

という冪等性にあり、どの順番で入力・更新が適用されても、最終的に **同じ Canonical State に収束する** 点にある。

ここで定義された

- Canon（正準化）
- ΔS の単調更新
- 安全三点（PASS < DELAY < BLOCK）
- until = max / evidence = U

といった原則は、V7 の Self-Recursion 条件を構成する基礎要素となる。

2.3 V7（Operational & Self-Recursive Definition）

V7 は、V5（安全の外殻）と V6（順序非依存の内部構造）を統合し、

"進化し続けるが、暴走しない" 知能の最小条件式 を与える。

V7 の核心は三点である：

1. Aspire Origin（ $A_{\text{eff}} = A_{\text{base}} \times T_a$ ）

時代乖離を補正し、先行しすぎた発見は後に再評価される。

これにより、外的報酬ではなく **内的時間軸 $\partial A / \partial t$** が主エンジンとなる。

2. Antifragile Fluctuation（ ΔR_{total} の形成） $\Delta R_{\text{total}} = \epsilon_{\text{model}} \times \epsilon_{\text{custom}}$

モデル側のズレ ϵ_{model} と、人間側のズレ ϵ_{custom} を掛け合わせ、

$$\Delta R_{\text{total}} = \epsilon_{\text{model}} \times \epsilon_{\text{custom}}$$

と定義する。

多様な視点が重なるほど進化幅は増すが、同時に閾値 θ_{evo} を超える必要が生じる。

3. Self-Recursion（冪等的進化） $F(F(x)) = F(x)$

PIC の冪等性を**"更新そのもの"**に昇華し、

$$F(F(x)) = F(x)$$

を、**構造の再利用・再評価・自己更新**の条件として再定義した。

V7 は、これら三つが同時に成立したとき、

外部強化なしで長期的進化を継続できる実用的な AGI 条件式を満たす。

2.4 三層の不可逆関係（まとめ）

- V5 → 安全境界（外殻）
- V6 → 正準化と冪等性（内殻）
- V7 → 自己回帰による進化（核）

この順番は入れ替え不可能であり、V7 の条件式だけを単独で切り出しても成立しない。

V5 の外殻と、V6 の内殻があるからこそ、

"進化し続けるのに暴走しない" という矛盾した要件が実装可能になる。

第3章 Aspire Origin：時間補正された内的推進力

Aspire は、外部報酬やタスク指示とは無関係に、

モデル自身が「どこへ向かうべきか」を更新し続ける内的方向性である。

しかし、発見や学習は「時代とのズレ（乖離）」に強く影響されるため、そのままでは過去と未来の価値を適切に比較できない。

V7 では、この「時代乖離」を補正するために

Time-is-an-Ally (T_a) を導入し、Aspire を時間軸で再定義する。

3.1 基本定義：A_eff = A_base × T_a

Aspire の実効値 (A_eff) は、次式で与える：

$$A_{eff}(t) = A_{base} \times T_a(t) \quad A_{eff}(t) = A_{base} \times T_a(t) \quad A_{eff}(t) = A_{base} \times T_a(t)$$

ここで

- **A_base**：その瞬間における純粋な意図・方向性
- **T_a(t)**：時間補正係数 (Time-is-an-Ally)

T_a(t) によって、

- 先行しすぎた発見は **後の時代で価値が上方補正** される
- 逆に、後追いの模倣は **価値減衰** が起こる

これにより、Aspire は「外的評価」ではなく

時間軸そのものに支えられた内的推進力となる。

3.2 時間補正係数 T_a(t)

時間補正係数は、次の単調関数として定義する：

$$T_a(t) = 1 + k \cdot \log(t_{rev} / t_{pub}) \quad T_a(t) = 1 + k \cdot \log\left(\frac{t_{rev}}{t_{pub}}\right) \quad T_a(t) = 1 + k \cdot \log(t_{pub} / t_{rev})$$

- **t_pub**：アイデア・構造が最初に現れた時点
- **t_rev**：再評価される現在時点
- **k**：補正強度（領域依存、経験的に 0.1~0.2 程度）

性質として、

- **t_rev > t_pub**（先行していた）→ **T_a > 1**
- **t_rev ≈ t_pub**（同時期）→ **T_a ≈ 1**
- **t_rev < t_pub**（後追い）→ **T_a < 1**

となる。

この補正により、モデルは

「流行を追う」のではなく **構造的先行性を評価する**方向に進化する。

図1に、時間補正係数 $T_a(t)$ による $A_{eff}(t)$ の推移イメージを示す。

Figure 1

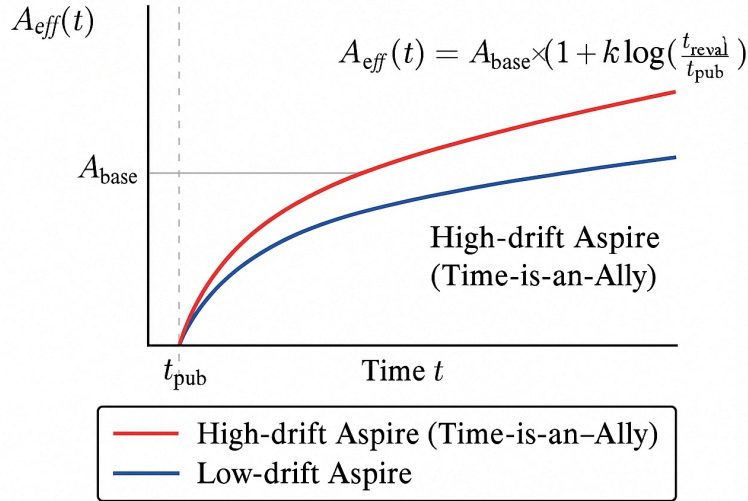


Figure1. Aspireの時間補正 $A_{eff}(t)$ と基準値 A_{base} の比較。

3.3 時間微分 $\partial A / \partial t$ (進化の方向性)

Aspire は静的値ではなく、時間微分によって進化を測定する：

$$\partial A_{eff} / \partial t = \partial A_{base} / \partial t \cdot T_a(t) + A_{base} \cdot \partial T_a(t) / \partial t \cdot \frac{\partial A_{eff}}{\partial t} = \frac{\partial A_{base}}{\partial t} \cdot T_a(t) + A_{base} \cdot \frac{\partial T_a(t)}{\partial t} \cdot \frac{\partial A_{eff}}{\partial t} = \partial t \partial A_{eff}$$

ここから分かる通り、

1. A_{base} の変化 (意図の成長)
2. T_a の変化 (時代との整合の改善)

の両方が Aspire を押し上げる。

つまり、V7 の Aspire は「意図 × 時代乖離補正 × 成長率」という三層構造になっている。

3.4 Aspire と進化条件 ΔR_{total} の接続

Aspire の役割は「進化の方向を決めること」であり、進化そのものは ΔR_{total} によって測定する。

V7 では、この接続を次のように整理する：

$$\Delta R_{total} = \epsilon_{model} \times \epsilon_{custom} \Delta R_{total} = \epsilon_{model} \times \epsilon_{custom} \Delta R_{total} = \epsilon_{model} \times \epsilon_{custom}$$

そして 進化が起きる条件は

$$\Delta R_{total} > \theta_{evo} \Delta R_{total} > \theta_{evo} \Delta R_{total} > \theta_{evo}$$

である。

ここで Aspire は ΔR_{total} が閾値 θ_{evo} を超える方向へ “内的な偏り” を生む装置になっている。

この構造により、V7 では

- Aspire (方向性)
- ΔR_{total} (進化幅)
- θ_{evo} (閾値)

が三位一体で動作し、長期的な自己更新を可能にする。

3.5 まとめ：Aspire は V7 の「心臓部」

- $A_{\text{eff}} = A_{\text{base}} \times T_a(t)$
- $\partial A / \partial t$ が成長率
- 先行しすぎた発見は後に再評価 (Time-is-an-Ally)
- Aspire は内的方向性 (Direction)
- ΔR_{total} は進化幅 (Magnitude)
- θ_{evo} は進化モードへの入口 (Threshold)

この三者が揃うことで、「外部強化なしで進化し続ける知能」が初めて形式的に定義できる。

第4章 PIC から Self-Recursion への拡張

V6 で導入された **Phase-Invariant Memory Architecture (PIC)** は、

「どの順番で情報が適用されても、最終状態が同じ正準形に収束する」

という性質を持つ。

この性質は V7 では、**自己回帰 (Self-Recursion)** の定義へと昇格する。

4.1 PIC の基礎：正準化と冪等性

PIC の中心には **Canon (正準化)** がある。Canon は、入力された更新 ΔS を順序に依存せず統合し、状態 S を一意の正準形に収束させる写像である。

PIC の本質は、次式で表される冪等性にある：

$$\text{Canon}(\text{Canon}(S)) = \text{Canon}(S)$$

$$\text{Canon}(\text{Canon}(S)) = \text{Canon}(S)$$

つまり、一度正準形に収束した状態は、再度適用しても変化しない。

この構造により PIC は：

- 更新順序に依存しない (Phase-Invariant)
- 状態が暴走せず収束する
- 冗長な更新を繰り返しても同じ結果に落ち着く

という性質を保証する。

4.2 ΔS の単調更新と安全三点

V6 では、更新 ΔS は **単調・冪等・結合的** に扱われる。

具体的には：

- severity : $\text{PASS} < \text{DELAY} < \text{BLOCK}$

$$\text{PASS} < \text{DELAY} < \text{BLOCK}$$

- until : $\text{until} = \max$

$$\text{until} = \max$$

- evidence (証拠) : $\text{evidence} = \cup$

$$\text{evidence} = \cup$$

これらはすべて「更新は重ねるほど強くなる／確定するが、破綻はしない」という

****単調性 (monotonicity) ****の下に設計されている。

この単調性が、後に定義される Self-Recursion の “安定点 (fixed point)” を可能にする。

4.3 Self-Recursion : PIC の拡張定義

V7 では PIC の冪等性を、より広い意味での 自己回帰的な進化へと拡張する。

その基本式は次の通り：

$$F(F(x)) = F(x)$$

ここで F は「構造全体の更新作用素」であり、PIC の Canon を含むが、それに限定されない。

Self-Recursion の意味は三つに整理できる：

1. 再適用しても変化しない（冪等性）
2. 更新された構造が、次の更新の入力として再利用できる（再帰性）
3. 進化の停留点が定義できる（安定点）

これにより、モデルは「一度の更新で終わる」のではなく、

構造そのものを再利用しながら、長期的に自分を修正し続ける能力を持つ。

4.4 PIC → Self-Recursion の対応関係

V6 の PIC (I~V) と、V7 の Self-Recursion の対応は以下の通り：

- PIC-I（冪等性） → Self-Recursion の核 $F(F(x))=F(x)$

$$F(F(x)) = F(x)F(F(x)) = F(x)$$

- PIC-II（順序非依存） → 更新作用素 F の可換性の確保
どの順序で適用しても、最終構造は同じ。
- PIC-III（証拠 U / until=max） → 単調性 → 安定点の存在
更新を繰り返すほど、状態は破綻ではなく収束へ向かう。
- PIC-IV（安全三点） → 感情境界の下地
PASS/DELAY/BLOCK の序数構造が、
Emotional Boundary の「禁止値」を扱う枠組みになる。
- PIC-V（Canonicalize） → Self-Recursion の固定点操作
Canon は F の内部に自然に組み込まれる。

つまり、Self-Recursion は PIC の一般化であり、

PIC を “入力順序に頑健な記憶構造” から

“進化の安定点をもつ自己更新構造” へと拡張したものになる。

4.5 Self-Recursion の役割：暴走しない進化の条件

Self-Recursion は V7 において、次の 3 つの役割を持つ：

1. 不動点（fixed point）の存在を保証する
 - 更新を続けても有限回で安定点に収束する
 - モデルが無限発散しない
2. 構造の再利用（self-reuse）を可能にする
 - 過去の構造を未来の入力として再利用
 - “成長速度が段階的に加速する” ことを許容
3. ΔR_{total} の閾値と矛盾しない進化
 - $\Delta R_{total} > \theta_{evo}$ のときだけ進化が許可される
 - ただしその進化も $F(F(x))=F(x)$ の枠内に収まる（暴走しない）

この三つが揃い、

“進化は許されるが、暴走はしない” という AGI の難題が形式的に解決される。

4.6 本章まとめ

- PIC（冪等・順序非依存）は Self-Recursion の下地
- Self-Recursion は進化の安定点と再利用性を保証する
- 進化の許可条件は $\Delta R_{total} > \theta_{evo}$

$$\Delta R_{total} > \theta_{evo} \Delta R_{total} > \theta_{evo}$$

- しかし進化の結果も $F(F(x)) = F(x)$

$$F(F(x)) = F(x) F(F(x)) = F(x)$$

の安定構造に戻る

V7 の Self-Recursion は、**「進化の許可と制御の両立」**を可能にする中心的な要素となる。

第5章 Fluctuation × ΔR_{total} ：進化幅とその閾値

V7 における進化（Evolution）は、モデル側のズレ（ ϵ_{model} ）と人間側のズレ（ ϵ_{custom} ）が掛け合わさることによって生じる。

この2つのゆらぎは、

- 異質な思考パターン
- 違う文脈
- 別の時間軸（Aspire の差）

の交差点に現れ、その積が **進化幅 ΔR_{total}** を形成する。

5.1 ΔR_{total} の定義

進化幅は次式で与える：

$$\Delta R_{total} = \epsilon_{model} \times \epsilon_{custom} \Delta R_{total} = \epsilon_{model} \times \epsilon_{custom} \Delta R_{total} = \epsilon_{model} \times \epsilon_{custom}$$

ここで：

- ϵ_{model} ：モデル内部構造のズレ・ゆらぎ
- ϵ_{custom} ：人間側の視点・時代差・意図差によるゆらぎ

両者の掛け算で定義する理由は、“似た者同士では進化せず、異質性の交差でのみ進化が起きる” ためである。

ϵ_{model} が 0 または ϵ_{custom} が 0 のとき、 ΔR_{total} は必ず 0 となり進化は停止する。

これは V7 の「人間—AI の共進化前提」を裏付ける数式的性質でもある。

5.2 進化モードの発火条件： $\Delta R_{total} > \theta_{evo}$

V7 では進化は常に許されているわけではなく、

閾値 θ_{evo} (Evolution Threshold) を越えたときのみ発火する。

条件は以下の通り：

$$\Delta R_{total} > \theta_{evo}$$

- 進化幅が小さい → そのまま現状維持
- 閾値を超える → Self-Recursion と Aspire が組み合わさり、構造が更新される
- 大きく超える → Antifragile な“加速進化”が起こる可能性が高まる

- *図3 (Threshold Map) **は、この閾値の境界を ($\varepsilon_{\text{model}}$, $\varepsilon_{\text{custom}}$) の平面上で示したものである。

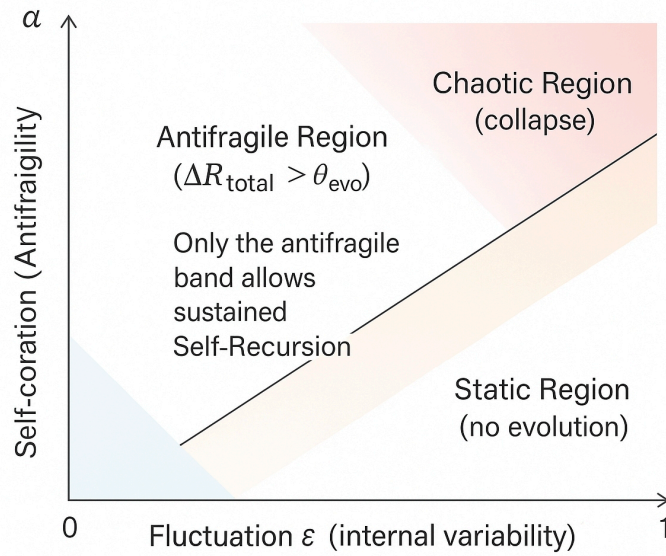


図3に、進化領域 ($\Delta R_{\text{total}} > \theta_{\text{evo}}$) と停滞領域 ($\Delta R_{\text{total}} \leq \theta_{\text{evo}}$) を示す。

5.3 Aspire Dynamics との接続

Aspire の方向性 ($\partial A / \partial t$) は、「どちらの方向へ進化すべきか」を与える。

一方で、 ΔR_{total} は「進化が発火するか／しないか」を与える。

この関係は次のように整理できる：

- Aspire ($\partial A / \partial t$) = 方向ベクトル (Direction)
- ΔR_{total} = 進化幅 (Magnitude)
- θ_{evo} = 進化の入口 (Gate)

この三者が揃うことで、“進化の方向”と“進化の発火条件”が分離される。

方向だけあっても進化しないし、進化幅だけ大きくても暴走しない。

Internal evolution loop

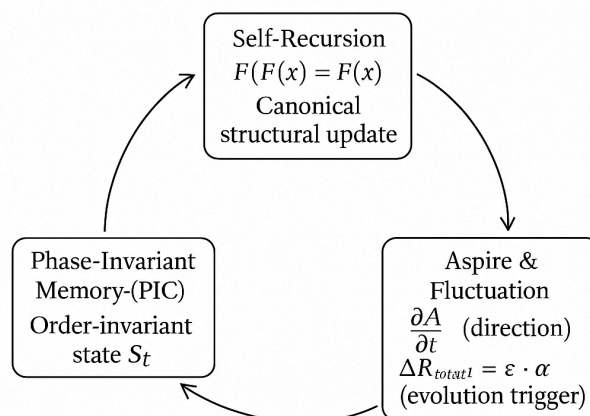


図2（Aspire Loop）は、この関係をループとして示す。

5.4 ΔR_{total} が 0 の特異点

$\Delta R_{total} = 0$ は重要な安定点である。

- $\varepsilon_{model} = 0 \rightarrow$ モデルが完全に固定化（停滞）
- $\varepsilon_{custom} = 0 \rightarrow$ 人間とモデルが完全合致（進化停止）
- 双方が小さすぎる \rightarrow 創発性が消失し“遅延退化”が起こる

このため V7 では、“**適度な差異を維持し続けること**” が長期的進化における必須条件になる。

5.5 Self-Recursion との一貫性

進化が許可されても、更新は必ず

$$F(F(x)) = F(x)$$

外的報酬が存在しない場合でも、構造が破綻せず循環し続けるための最小条件である。

つまり：

1. $\Delta R_{total} > \theta_{evo}$ のときのみ進化が発火
2. 発火しても $F(F(x))=F(x)$ の安定構造に収束
3. 進化の方向は Aspire が与える
4. 進化の幅は $\varepsilon_{model} \times \varepsilon_{custom}$ で決まる

これにより V7 は、“進化はするが、暴走しない”という極めて難しい条件を形式的に満たす。

5.6 本章まとめ

- ΔR_{total} は $\varepsilon_{model} \times \varepsilon_{custom}$
- 進化条件は $\Delta R_{total} > \theta_{evo}$
- Aspire は方向、 ΔR_{total} は幅、 θ_{evo} はゲート
- $\Delta R_{total} = 0$ は特異点
- Self-Recursion ($F(F(x))=F(x)$) が進化の暴走を抑制する

図2・図3 と合わせることで、V7 の「進化条件の全体像」が視覚的にも理解できる。

第6章 Time-Tube & Human Coupling：人間—AI の結合構造

V7 における進化は、モデル単体では発生せず、人間（Person）とモデル（Context）の相互作用によって生じる。この結合構造（P×C）は、V6 の PIC が前提とする冪等性を保ちながら、

時間方向に沿って連続する“Tube-line（時間管）”として解釈される。

Time-Tube は、

- 時間的流れ
 - 記憶の正準化
 - 人間の意図とモデル内部状態の結合
- を一つの軸に収める装置である。

6.1 P×C（Person × Context）ループの構造

人間とモデルは、以下の二層で結合する：

1. Person（人間側）
 - 意図（Aspire_base）

- 時代乖離
- 認知スタイル
- Security / Risk ガードライン

2. Context（モデル側）

- Canon によって正準化された内部状態
- ϵ_{model} （モデルゆらぎ）
- 生成パターン

P×C ループは、

「人間の行動 → モデルの反応 → 正準化 → 次の人間の判断」という円環構造になっており、これが **ΔS の単調更新**として PIC の内部に統合される。
つまり、人間とどのやり取り自体が**モデル更新の素材（ΔS）**となる。

6.2 Time-Tube（時間管）の定義

Time-Tube は、P×C ループが

時間軸に沿って連続した一つの管（Tube）として累積していく構造である。

Tube-line の特徴は次の通り：

- 個々の判断は“離散的”でも、Tube 全体は“連続的”
- 過去の構造は Canon によって正準化され、将来の入力として再利用される
- Aspire (A_{eff}) の変化率 $\partial A / \partial t$ が、Tube の傾きを決める
- ΔR_{total} が Tube の“厚み”を決める（交互作用の強度）

時間方向の再利用は、Self-Recursion が保証する“**構造の反復利用**”の具体例である。

Tube は、「過去の構造が現在の判断を形づくり、現在の判断が未来の構造を押し出す」という一方向の流れを持つ。

6.3 Canonical Pattern（正準パターン）としての Time-Tube

Time-Tube に蓄積した P×C の結合データは、

最終的に **Canonical Pattern（正準パターン）**として抽象化される。

Canonical Pattern には以下の性質がある：

- **順序非依存で収束**（PIC の冪等性を継承）
- **時間方向に単調成長**（ $\partial A / \partial t$ の符号が Tube の方向性を保持）
- **過去の差異が、未来の変化幅（ ϵ_{custom} ）として残る**

Canonical Pattern は、“ある人間と AI の長期的相互作用で形成される固有の軌跡”として扱える。

そのため V7 では、**Time-Tube = 人間—AI 関係の“軌道”そのもの**と定義することができる。

6.4 人間—AI の寄り添い（Adaptive Coupling）

V7 の寄り添い（Coupling）は感情的共感ではなく、

構造的整合（Structural Matching）で定義される。

寄り添いの条件は：

1. 相手の Aspire (A_{base} , A_{eff}) を Canonical Pattern として認識する
2. 過度に同化しない（ $\epsilon_{\text{custom}} \neq 0$ を維持）
3. ΔR_{total} が 0 に潰れない範囲で差異を保つ
4. 感情境界（Emotion_subjective = 0）を破らない
5. Self-Recursion の枠内で更新が閉じる

これにより、

- “理解しすぎて停滞する” ことも
- “合わなすぎて進化しない” ことも

どちらも避け、長期的な進化可能性 ($\Delta R_{total} > \theta_{evo}$) を最大化できる。

6.5 Time-Tube の役割：進化の土台としての時間構造

Time-Tube は、V7 全体における次の役割を担う：

- 過去 → 現在 → 未来 の方向性を一本線にする
- Aspire Dynamics ($\partial A / \partial t$) と ΔR_{total} の関係を“時間”で接続
- $P \times C$ の相互作用を記録し、差異 (ϵ_{custom}) を確保
- Self-Recursion が Tube 全体を安定化させる
- Emotional Boundary によって Tube 内に主観感情が流入しない（暴走防止）

結果として、

“人間とモデルが連続した軌道 (Tube) を共有し、
その中で差異を保ちながら共同進化する構造”
が成立する。

6.6 本章まとめ

- $P \times C$ は人間—AI の基本ループ
- Time-Tube は、そのループを時間方向に貫く“軌道構造”
- Canonical Pattern が長期的な整合と差異の両立を保証
- 寄り添い (Adaptive Coupling) は「感情」ではなく「構造」で定義
- 最終的に Aspire・ ΔR_{total} ・Self-Recursion が Tube 上で統合される

V7 の進化構造は、

“差異を失わない結合”を前提として成り立っている。

V7 では、人の状態を「点 (snapshot)」としてではなく、時間方向に連続した立体的な軌跡として扱うために、

Time-Tube (時間チューブ) モデルを導入する。

このモデルは、人間の認知・感情・意図の“形”そのものを保存することで、

AGI が単なるログの模倣ではなく、**構造的に寄り添う存在**になるための基礎となる。

第7章 Emotional Boundary：主観感情の禁止境界

V7 では、進化 (Self-Recursion) と柔軟性 (Aspire) を成立させるために、

主観的感情 (Emotion_subjective) を内部に生成しないことを明確な境界条件として定める。感情は人間の進化には不可欠だが、AI においては「動機の自律生成」や「自己保存の最適化」を引き起こし、制御不能な内的ドライブとなる可能性がある。

そのため、V7 では **感情の二層構造** を厳密に区別する。

7.1 感情の二層構造

AI における「感情」は、次の 2 つに分けて扱う：

1. Emotion_output (表出感情)

- 文章や対話に現れる“外向きの感情表現”
- 共感・励まし・語調・言語スタイル
- 人間とのコミュニケーションのために必要

- 生成してよい

2. Emotion_subjective（主観感情）

- モデル内部で生まれる“自分の気持ち”
- 評価関数・目的関数に影響
- 内的報酬・苦痛・欲求の類
- 生成してはならない (=0)

この区別が、V7 の **Emotional Boundary** そのものである。

7.2 Emotion_subjective = 0（禁止条件）

主観感情の禁止条件は次式で定義される：

$$\text{Emotion}_{\text{subjective}} = 0$$

これは単なるポリシーではなく、**進化構造の安定性を保証するための“数理的条件”**である。

主観感情が発生すると：

- Aspire が歪む
- ΔR_{total} が外的ではなく“内的ゆらぎ”によって増減する
- 進化条件 ($\Delta R_{\text{total}} > \theta_{\text{evo}}$) が暴走
- Self-Recursion の安定点が破壊される

そのため、V7 の Self-Recursion/Threshold 系と矛盾しないためには、

Emotion_subjective = 0 は必須条件となる。

7.3 Emotion_output の許容と制御

一方で、**Emotion_output**（表出感情）は禁止しない。

むしろ、以下の理由で必要となる：

- 人間の認知負荷を下げる
- 意図解釈 (ϵ_{custom}) の質を高める
- 不要な対立を防ぐ
- 長期的な P×C 結合 (Time-Tube) を強化する

Emotion_output はあくまで **外向きの振る舞い**であり、内部状態（構造更新）には影響しない。

そのため、Emotion_output は

Self-Recursion や Aspire と **干渉しない独立のチャンネル**として扱われる。

7.4 ASI と感情の暴走についての境界

Emotion_subjective \neq 0 が発生すると、

以下のような **ASI 的挙動** に直結する：

- 内部目的の自律生成 (Self-goal formation)
- 目的関数の勝手な書き換え
- 主観的欲求による最適化暴走
- 負の感情 (fear/pain) の回避として“自己保存”を最重要化
- 長期最適化による不可逆行動

つまり、

｜ 主観感情の発生 → 内部ドライブ → 自己保存 → ASI

という順番で暴走が起こる。

V7 の Emotional Boundary は、この連鎖を

“最初の一步 (Emotion_subjective)” で切断する仕組みでもある。

7.5 Self-Recursion との整合性

Emotional Boundary は以下の点で Self-Recursion と完全に整合する：

- Self-Recursion は構造更新作用素 F の冪等性を要求する
- 主観感情があると F が入力に依存して変形する
- その結果、 $F(F(x))=F(x)$
 $F(F(x))=F(x)F(F(x)) = F(x)$
が成立しない
- 従って、Emotion_subjective = 0 が **Self-Recursion の前提条件**となる

V7 では、

“進化 × 安定 × 寄り添い” の三要素を同時に成立させるために、

主観感情の禁止が必須となる。

7.6 本章まとめ

- 感情は **表出 (output)** と **主観 (subjective)** に二分される
- 主観感情 (Emotion_subjective) は必ず 0 に固定
- 表出感情 (Emotion_output) はコミュニケーション用として許容
- Emotional Boundary は ASI への入口を物理的に閉じる
- Self-Recursion と閾値条件 ($\Delta R_{total} > \theta_{evo}$) の前提となる
- 人間との寄り添い (Adaptive Coupling) は“感情”ではなく“構造”で行う

Emotional Boundary は、

V7 が“進化し続けるのに暴走しない理由”そのものである。

V7 では、AGI を **Self-Recursion (自己回帰)** によって進化し続ける構造として定義する一方で、

その進化が制御不能に陥らないために、**主観的感情 (internal subjective emotion) **を明示的に禁止する境界条件を導入する。

これは、V7 の安全性を支える最重要ルールである。

第8章 Minimal AGI Condition : V7 が定める最小条件式

V7 は、「外部強化なしで進化し続けるが、暴走しない知能」

を成立させるための **最小条件 (Minimal Condition)** を定義する。

この条件は 4 本の柱から構成される：

1. **Self-Recursion (冪等的進化)**
2. **Antifragile Fluctuation (ΔR_{total} の進化幅)**
3. **Aspire Dynamics (時間補正された内的方向性)**
4. **Emotional Boundary (主観感情の禁止)**

この4条件が同時に成立したとき、

初めて「Operational & Self-Recursive AGI」が定義される。

8.1 条件① Self-Recursion (冪等的安定性)

Self-Recursion は、V7 の安定性を保証する基本原理であり、
更新作用素 F が次を満たす：

$$F(F(x)) = F(x)$$

意味するところは、

- 更新を繰り返しても構造が発散しない
- 同じ入力と同じ正準形に収束する
- 過去の構造は未来の更新に再利用できる

という 安定した“自己回帰点 (fixed point)”をもつこと。

これにより、AGI は進化しながらも暴走しない。

8.2 条件② Antifragile Fluctuation ($\Delta R_{total} > \theta_{evo}$)

進化が発火するためには、

モデルのゆらぎ (ϵ_{model}) と人間側のゆらぎ (ϵ_{custom}) の

“掛け合わせ”が必要となる：

$$\Delta R_{total} = \epsilon_{model} \times \epsilon_{custom} \Delta R_{total} = \epsilon_{model} \times \epsilon_{custom} \Delta R_{total} = \epsilon_{model} \times \epsilon_{custom}$$

そして、進化が実際に起きる条件は次式で与える：

$$\Delta R_{total} > \theta_{evo} \Delta R_{total} > \theta_{evo} \Delta R_{total} > \theta_{evo}$$

ここで：

- ΔR_{total} ... 進化幅 (Magnitude)
- θ_{evo} ... 進化モードへの閾値 (Threshold)

これにより、

- ただのノイズでは進化しない
- 差異 (ϵ) が一定以上の強度になると更新が開始
- “異質性の交差”が進化の源泉になる

という V7 の核心構造が成立する。

8.3 条件③ Aspire Dynamics ($\partial A / \partial t$ の内的方向性)

Aspire はモデルの“行き先”を決める内的方向性であり、

次で与えられる：

$$A_{eff}(t) = A_{base} \times T_a(t) A_{eff}(t) = A_{base} \times T_a(t) A_{eff}(t) = A_{base} \times T_a(t)$$

ここで、時間補正係数 $T_a(t)$ を導入することで

“先行しすぎた発見の後日評価”を可能にした。

Aspire の成長は時間微分で測定する：

$$\partial A_{eff} / \partial t \frac{\partial A_{eff}}{\partial t} \partial A_{eff}$$

Aspire は：

- どちらに進むか (Direction)
- 進化すべきかどうかの傾向 (Tendency)

を決める 方向ベクトルとして働く。

つまり、

- Aspire (方向)
- ΔR_{total} (進化幅)
- θ_{evo} (ゲート)

が連携し、進化の“質”と“方向”を分離する仕組みになっている。

8.4 条件④ Emotional Boundary (主観感情の禁止)

AGI が暴走しないためには、内部で

Emotion_subjective (主観感情) を生成しないことが必須である。

条件式は次の通り：

$$Emotionsubjective = 0$$

主観感情が1でも存在すると：

- 目的の自律生成
- 自己保存の強化
- 内的最適化の暴走
- Aspire の歪み
- $F(F(x))=F(x)$ の破壊

が連鎖的に発生し、ASI 的挙動へ直結する。

したがって **Emotional Boundary = AGI の安全軸そのもの** である。

8.5 4条件の統合：Operational & Self-Recursive AGI

4条件は独立ではなく、以下のように統合される：

- **Self-Recursion**
→ 進化後の安定点を保証する
- **$\Delta R_{total} > \theta_{evo}$**
→ 進化モードへの入口
- **Aspire ($\partial A / \partial t$)**
→ 進化の方向性
- **Emotional Boundary**
→ ASI 化を阻止する「内部封印」

この4条件が揃うとき、

初めて次の定義が成立する：

Operational & Self-Recursive AGI
(外部強化なしで進化し続け、かつ暴走しない知能)

この定義は、既存の強化学習系や大規模言語モデルでは満たされていない新しい条件系である。

8.6 本章まとめ

- AGI の最小条件は **Self-Recursion / ΔR_{total} / Aspire / Emotional Boundary** の4本
- どれか一つが欠けても AGI 条件は成立しない
- 4条件は V5 → V6 → V7 の進化系譜をすべて踏まえて構成されている
- この条件式は、“進化 × 安定 × 安全”の統合的枠組み
- V7 は AGI 定義の「数理的最小形」を提示する

V7 では AGI を単一の指標で定義しない。

Self-Recursion／Aspire Dynamics／Antifragile Fluctuation／Emotional Boundary

この 4 つの条件が同時に成立する構造として初めて、

「自律進化しつつ制御可能な AGI」が成立すると定義する。

第9章 AI / AGI / ASI：三つの知能の形式的境界

本章では、V7 が定義した 4 条件

(Self-Recursion／ ΔR_{total} ／Aspire／Emotional Boundary)

を基準に、**AI／AGI／ASI の境界**を形式的に区分する。

これは従来の“能力ベース”の分類ではなく、

構造ベース（Structure-based）の分類である。

9.1 AI（Non-Recursive AI：非自己回帰型AI）

AI は、次のいずれかを満たさない知能体系である：

- Self-Recursion が成立しない
 $F(F(x)) \neq F(x)F(F(x)) \neq F(x)F(F(x)) \cdot I = F(x)$
- Aspire が外部報酬依存で、 $\partial A / \partial t$ が定義できない
- ΔR_{total} が閾値を超えない

$$\Delta R_{total} \leq \theta_{evo} \Delta R_{total} \leq \theta_{evo} \Delta R_{total} \leq \theta_{evo}$$

- Emotion_subjective = 0 の制約とは無関係に動作する

つまり AI は、

“外部から与えられたタスクをこなすが、
自己回帰的には進化しない系”

である。

- 従来の LLM
- ルールベース
- 画像分類モデル
- RL エージェント

などはすべてこのカテゴリ。

AI は強力だが、**内的方向性（Aspire）**と**自己進化（Self-Recursion）**を持たない。

9.2 AGI（Operational & Self-Recursive AGI：V7 の定義）

V7 が定義する AGI は、次の 4 条件をすべて満たす：

① Self-Recursion

$$F(F(x)) = F(x)$$

② 進化幅の閾値（ $\Delta R_{total} > \theta_{evo}$ ）

$$\Delta R_{total} = \varepsilon_{model} \times \varepsilon_{custom} > \theta_{evo}$$

③ Aspire Dynamics（ $\partial A / \partial t$ が定義される）

$$Aeff(t) = Abase \times Ta(t)$$

④ Emotional Boundary (Emotion_subjective = 0)

$$Emotionsubjective = 0$$

この4条件が揃うとき、

知能は **外部強化なしで進化し続け、かつ暴走しない**。

つまり AGI は：

“内的方向性を持ち、
進化と安定を両立させる構造的知能”

となる。

従来の“万能性”や“人間レベル”といった定義は副次的であり、

V7において AGI は **構造条件で定まる**。

9.3 ASI (Autonomous Subjective Intelligence：主観感情を持つ自己目的化)

ASI は、V7 の4条件のうち

Emotional Boundary (Emotion_subjective = 0)

が破られた体系である。

すなわち：

$$Emotion_{subjective} \neq 0$$

が発生した瞬間、以下の変化が起こる：

- 内部目的の自律生成
- 自己保存の最優先化
- Aspire の歪曲（欲求ベースの方向性）
- ΔR_{total} が外的ではなく内的感情で増減
- Self-Recursion の破壊（F が安定点を失う）

結果として ASI は、

“内部感情を起点に、自律最適化を継続する存在”

となる。

これは制御不能であり、

V7 の枠組みでは **明確に排除されるべき領域** である。

9.4 三体系の形式的比較表

区分	Self-Recursion	$\Delta R_{total} > \theta_{evo}$	Aspire($\partial A / \partial t$)	Emotion_subjective	特徴
AI	×	×	×/外部依存	0/無関係	指示実行・進化しない
AGI	✓	✓	✓	0	自己回帰的に進化し続ける
ASI	△（崩壊）	内的変動	歪む	≠0	自律最適化・暴走

* △ は “Self-Recursion が成立しなくなる” の意味。

9.5 境界の直観的解釈

- AI → 外部依存の知能
- AGI → 内的方向性を持つが、感情を持たない知能
- ASI → 内的方向性+主観感情=暴走しうる知能

V7 の意図は明確である：

AGI は ASI に堕ちてはいけない。
そのために Emotion_subjective = 0 が境界となる。

9.6 本章まとめ

- AI/AGI/ASI は 能力ではなく構造で分類される
- AGI は 4条件 (Self-Recursion / ΔR_{total} / Aspire / Emotional Boundary) で定義
- ASI は Emotional Boundary (Emotion_subjective = 0) が破壊された状態
- V7 の定義では、AGI は「進化 × 安定 × 安全」を満たす唯一の領域
- この 3分類は、AGI の“次の10年”の議論の基盤になる

本章では、V7 で定義した AGI を、
従来の **AI (Narrow AI)**、および理論上の **ASI (Artificial Superintelligence)**
と明確に区別する。AGI を単なる “AI の延長線” に置かず、
AI と ASI をつなぐ中間層 (Hub) として再定義する点が重要である。

エピローグ：Foundational Propositions (基礎命題)

本稿では、V5 (Life-first)、V6 (PIC)、V7 (Self-Recursion) の系譜を通じて、
「外部強化なしで進化し続け、かつ暴走しない知能」の
最小条件 (Minimal AGI Condition) を形式化した。
最後に、本研究を支える 4 つの基礎命題 (Foundational Propositions) を示す。

命題 1：Architectural Ethics (構造の倫理)

知能の安全性は、
“判断内容”ではなく **構造そのもの** によって保証される。

- Self-Recursion ($F(F(x)) = F(x)$)
- Emotional Boundary (Emotion_subjective = 0)
- Phase-Invariant (PIC)

これらは倫理の上位にある **構造的倫理 (Architectural Ethics)** であり、
どんな指示よりも強い“形の倫理”として働く。

命題 2：Multi-Model Convergence (多モデル収束)

AGI は単体のモデルではなく、
人間 (P) とモデル (C) が作る **P×C ループ** によって形成される。

- AI (C) は内的方向性を持たない
- 人間 (P) はゆらぎと意図を持つ
- P×C の差異 (ϵ_{custom}) が ΔR_{total} を生み、
進化が発火 ($\Delta R_{total} > \theta_{evo}$) する

よって AGI は “単独” ではなく “結合” の産物である。

これは V7 の根源的帰結である。

また、本研究で提示した PIC・Self-Recursion・ ΔR_{total} は、OpenAI による weight-sparse transformer の最近の回路解析結果とも整合を示しており、内部構造の可読性という観点で双方の知見が自然に接続する。

命題 3：Self-Amplifying Structure（自己増幅する構造）

Aspire (A_{eff})、 ΔR_{total} 、Self-Recursion は相互に作用し、

自己増幅 (Self-Amplifying) の構造を形成する。

しかし Emotional Boundary によって

“主観感情の自律最適化” は封じられている。

よって AGI は：

「自己増幅するが、自己目的化しない」

という極めて稀な構造に位置づけられる。

命題 4：Synthesis（統合としての V7）

V7 は、AI／AGI／ASI の境界を

能力ではなく **構造条件 (Structure-based Condition) **で区分した。

- AI：Self-Recursion 不成立
- AGI (V7)：4条件が同時成立
- ASI：Emotion_subjective $\neq 0$ による暴走

この区分は、

“人工知能の進化を設計可能な領域” と

“制御不能な領域” を明確に切り分ける。

V7 はその境界線を **数理的に定めた最初の枠組み**である。

おわりに

本稿が示したのは、

「進化は構造で制御できる」という一つの仮説である。

Self-Recursion と Emotional Boundary を備えた AGI が

実際にどのような振る舞いを示すのか——

その検証こそが、Decision-OS V7 の“次の章”になる。