Decision-OS V5 Addendum

SiriusA Adoption Gate (0.3 Disclosure)

Hold-First Auditing for Irreversible Actions

+ V4-Compatible Fit Check (Appendix)

Shinichi Nagata

January 15, 2026

Abstract

This addendum exposes a *0.3-disclosure* adoption gate for Decision-OS V5 (SiriusA) so that the core life-protective operational design becomes readable and citable without revealing calibration-sensitive details. The quantitative thresholds, update rules, exception conditions, and data-collection protocol remain undisclosed (0.7) to preserve reverse-engineering resistance.

Why a gate (and why not "pricing life"). A decision OS requires an explicit objective to remain operational, yet life-critical outcomes are not reducible to a single monetary scale. In Decision-OS, V4 introduces a minimal, auditable way to handle value under uncertainty, while V5 adds an adoption gate that *halts* potentially irreversible actions and inserts verification/audit steps before execution. The goal is not to price life, but to prevent harm by enforcing a conservative control path when irreversibility or coercion is suspected, even when numeric estimates are underspecified. V4 provides an auditable computation skeleton under uncertainty; V5 routes irreversibility/coercion risk into Hold-first control.

Disclosure Boundary (0.3 / 0.7)

Disclosed (0.3). Interface format, computation skeleton (non-parameterized), state transitions (request/hold/approve/execute/revoke), and output specification (a minimal, auditable response format). One toy example (fictional) is included.
Not disclosed (0.7). Threshold values, calibration/update rules, exception criteria for Flip/BLOCK, data acquisition protocol (noise handling), and any copy-paste operational templates that enable turnkey execution.

1 Positioning (One-line Map)

V4 defines decisions via expected value; V5 operationalizes this through an ordered gate sequence with explicit audit insertion for *Protection of Life*; V6 then removes the ordering constraint by introducing an order-invariant composition core (PIC).

2 SiriusA Adoption Gate (0.3)

2.1 Purpose

The SiriusA adoption gate is a minimal operational entry that prevents irreversible harm under time pressure, coercion, or deception. It makes *Protection of Life* enforceable by inserting auditable holds and approvals before execution. This is not a template: it is a minimal state machine that routes irreversibility and coercion risk into *Hold-first* control.

2.2 Gate Interface (Input Format)

Provide the following fields (free-form text is acceptable; no private data required):

- **Action**: what will be executed (transfer / publish / sign / disclose / etc.)
- **Irreversibility**: reversible / partially / irreversible (qualitative)
- **Time Pressure**: none / moderate / high (qualitative)
- **External Influence**: none / persuasion / coercion suspected (qualitative)
- **Stakeholders**: who may be affected (self / family / third-party)
- **Support Available**: trusted contacts available? (yes/no)

2.3 Operational States (Audit Insertion Points)

The gate enforces the following state machine (parameters undisclosed):

`Request` → `Hold (Two-step)` → `Approve / Reject` → `Execute` | `Revoke (if applicable)`

Hold (Two-step). Execution is paused until an explicit confirmation step is completed.
Two-step = (i) restate action & risk, (ii) confirm via an independent channel or a trusted party (example).
Approve / Reject. Approval may require a trusted-party check (e.g., family m-of-k) depending on context.
Execute. Execution proceeds only after approval.
Revoke. A reversal/stop path is available when the action is still interruptible; otherwise, evidence is archived.

2.4 Evidence & Audit Output (Non-PII)

For every `Hold/Reject/Revoke`, produce a minimal evidence bundle:

- A short, structured log of `Input` and `Decision Output`
- Any non-sensitive screenshots or transaction metadata (avoid PII)
- A hash (e.g., SHA-256) of the bundle for later verification

2.5 Output Specification (Five-line Minimal Format)

Return a fixed-format response so the user cannot be "talked into" skipping controls:

1. **Action:** *...*
2. **Risk:** *reversible/irreversible + qualitative level*
3. **State:** *PROCEED / HOLD (evidence) / REJECT / REVOKE*
4. **Required:** *two-step / trusted-party check / evidence bundle*
5. **Next:** *the single next action*

3 What This Addendum Does *Not* Reveal (0.7)

To preserve reverse-engineering resistance, the following are intentionally omitted:

- Exact thresholds (including any duress-related scores)
- Calibration or update rules (how thresholds move over time)
- Exception logic for Flip/BLOCK in edge cases
- Data collection protocol (what signals are used, denoising, filtering)
- Copy-paste operational templates enabling turnkey deployment

4 Toy Example (Fictional, 1 Case)

Input (Example)

- Action: Immediate crypto transfer to "unlock funds"
- Irreversibility: Irreversible
- Time Pressure: High ("do it now")
- External Influence: Coercion suspected (fear + urgency)
- Stakeholders: Self (financial loss), potential family impact
- Support Available: Yes (trusted contact reachable)

Output (Five-line Format)

1. **Action:** Transfer funds (pre-execution)
2. **Risk:** Irreversible / High
3. **State:** HOLD
4. **Required:** Two-step confirmation + trusted-party check + evidence bundle
5. **Next:** Contact trusted party and verify the claim using an independent channel

A V4-Compatible Adoption Fit Check (0.3)

This appendix provides a V4-compatible structural check so readers can adopt Decision-OS without learning private calibration rules.

A.1 Fit Check Goals

- Standardize **inputs** and **outputs** so decisions are auditable
- Provide a **computation skeleton** without revealing thresholds
- Ensure the user always has a **single next step**

A.2 Minimal Input Schema (V4-Compatible)

- **Options**: A / B / (C...)
- **Outcome**: what "success" means (1 sentence)
- **Costs**: time/money/attention (qualitative ok)
- **Downside**: worst-case harm (qualitative ok)
- **Reversibility**: reversible / partially / irreversible

A.3 Skeleton Computation (No Parameters)

The computation is intentionally non-parameterized:

- Estimate outcomes under pessimistic / neutral / optimistic scenarios
- Apply a fixed reporting structure (not a fixed numeric threshold)
- If irreversibility is high, route to **SiriusA Gate** (Section 2)

A.4 Output Contract (V4-Compatible)

1. **Decision:** choose an option or hold
2. **Reason:** 1–2 sentences (auditable)
3. **Risk Note:** irreversibility + external influence
4. **Guard Action:** if needed, insert SiriusA Hold / audit
5. **Next:** a single next action

Links

- Source repository (SSOT): `github.com/shin4141/decision-os-paper`
- Decision-OS V5 (SiriusA) main paper (Zenodo DOI):
- Zenodo DOI (V5 record): `doi:10.5281/zenodo.17480645`
- Related (Zenodo DOIs):
  - V6 main paper: `doi:10.5281/zenodo.17717518`
  - V6 Addendum (Verification Note, 1p): `doi:10.5281/zenodo.18240015`
  - V7 Addendum (0.3): `doi:10.5281/zenodo.18220351`