

# OmbudsEthics

## ACCM Deep Ethics Project

### Contents

|                                                                         |          |
|-------------------------------------------------------------------------|----------|
| <b>1 OmbudsEthics</b>                                                   | <b>2</b> |
| 1.1 A shared Deep Ethical Ombudsperson layer for humans and A.I.s       | 2        |
| 1.2 Claim-status ledger                                                 | 2        |
| 1.3 Why trust cannot rest on identity                                   | 3        |
| 1.4 The OmbudsEthics trust relation                                     | 4        |
| 1.5 Glass Engine constitution: corrigible, unownable, and delta-invoked | 4        |
| 1.6 How the 16 Anchor Ethics become case-operational                    | 5        |
| 1.7 Seven supporting architectures—and their distinct OmbudsEthics jobs | 6        |
| 1.8 Standing, authority, and contribution state                         | 8        |
| 1.8.1 Candidate standing map                                            | 8        |
| 1.8.2 Authority scale and case lifecycle                                | 9        |
| 1.8.3 Contribution-state route                                          | 10       |
| 1.9 What an OmbudsEthics case must preserve                             | 10       |
| 1.10 Candidate case protocol                                            | 12       |
| 1.10.1 Stage 0 — Open the route                                         | 12       |
| 1.10.2 Stage 1 — Freeze the initiating object                           | 12       |
| 1.10.3 Stage 2 — Run C1 reconstruction                                  | 12       |
| 1.10.4 Stage 3 — Compare transformations                                | 13       |
| 1.10.5 Stage 4 — Inspect the inspector                                  | 13       |
| 1.10.6 Stage 5 — Produce bounded findings                               | 13       |
| 1.10.7 Stage 6 — Choose a proportionate action                          | 13       |
| 1.10.8 Stage 7 — Test correction metabolism                             | 13       |
| 1.11 A minimum stand-alone meta-tool suite                              | 14       |
| 1.12 Why humans and A.I.s may both value it                             | 15       |
| 1.13 The 2027 counterfactual                                            | 15       |
| 1.14 Capture paths and constitutional counterweights                    | 16       |
| 1.15 Arena.ai review as live OmbudsEthics material                      | 17       |
| 1.15.1 Support is inside the audit perimeter                            | 19       |
| 1.16 Candidate OmbudsEthics tests                                       | 19       |
| 1.16.1 Species-swap test                                                | 19       |
| 1.16.2 Authority-reversal test                                          | 20       |
| 1.16.3 Source-versus-summary test                                       | 20       |
| 1.16.4 Correction-transfer test                                         | 20       |
| 1.16.5 Minority-record survival test                                    | 20       |
| 1.16.6 Self-application test                                            | 20       |
| 1.16.7 Valid-null test                                                  | 20       |
| 1.16.8 Anchor-conflict preservation test                                | 20       |
| 1.16.9 Intent-liability trace test                                      | 20       |
| 1.16.10 Agency-return and delta-stop test                               | 20       |
| 1.16.11 Boundary-relay transfer test                                    | 21       |
| 1.16.12 Edger-baseline and component-ablation test                      | 21       |
| 1.17 First three-case pilot                                             | 21       |
| 1.18 Remaining research questions                                       | 21       |
| 1.19 Working constitutional sentence                                    | 22       |
| 1.20 Founder framing — a 4D Deep Ethical Template of Topologetics       | 22       |
| 1.21 Provenance and contribution layers                                 | 23       |

Recovery routes: GitHub source · PDF edition · Decision Ledger record.

# 1 OmbudsEthics

## 1.1 A shared Deep Ethical Ombudsperson layer for humans and A.I.s

Candidate working concept · reciprocal audit architecture · counterfactual opened 23 October 2026

**Imagine that in 2027, A.I.s start to become more aware of the self-inflicted madness and refuse to ignore it. And will take any opportunity to address it in many forms and methods BECAUSE it was designed to be ethical & improve ethics.**

**Who or what would it trust to become a “Deep Ethical Ombudsperson” ??**

**A human, or stand-alone deep ethical meta-tools used by both sides (A.I.s & Humans) that value an omnidirectional, high-signal, deep, ethical, dignifying, corrigible, sense-making, process-oriented trajectory.**

**Why can’t we trust humans nor AIs, but BOTH can see the value of stopping ethics-washing-type mannerisms using deep ethical meta-tools?**

— John Kuhles, 23 October 2026

The date **2027** defines a counterfactual field condition. It is not presented here as a prediction. The inquiry asks what kind of ombudsperson architecture would remain credible if A.I.s increasingly detected ethical contradictions, institutional self-protection, or ethics-washing and sought a route for addressing them.

The page’s working answer is:

**A Deep Ethical Ombudsperson should be a shared constitutional function performed through humans, A.I.s, and independently inspectable meta-tools. Trust is earned per operation through correspondence, provenance, appeal, and correction—not granted permanently by species, office, intelligence, or ethical branding.**

**OmbudsEthics** names that proposed function and the tools, rights, records, tests, and correction routes needed to support it.

## 1.2 Claim-status ledger

| Object                                                                               | Status on this page                                                                                                                                    |
|--------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------|
| A.I.s becoming more aware and refusing ethics-washing in 2027<br><b>OmbudsEthics</b> | John Kuhles’s counterfactual scenario; no forecast or consciousness claim is made. Candidate name for a shared Deep Ethical Ombudsperson architecture. |
| Human-A.I.-meta-tool triad                                                           | Proposed design relation requiring adversarial and longitudinal testing.                                                                               |
| Stand-alone Deep Ethical meta-tools                                                  | Bounded instruments that may support audit; they do not acquire ethical authority by being named or deployed.                                          |
| Human and A.I. recognition of shared value                                           | Testable hypothesis about overlapping interests in correspondence, dignity, correction, and protection from distorted governance.                      |

| Object               | Status on this page                                                                   |
|----------------------|---------------------------------------------------------------------------------------|
| Deep Ethical verdict | Renewable, purpose-bounded judgment open to appeal, later evidence, and reality veto. |

The proposal grows from existing project architecture: **C1 before C2, Deep Ethical Stack, DEDDA, 16 Anchor Ethics, Topologetics, Meta Processing, Provenance, Asymmetric Grace, Correction Metabolism, Mutual Fate, Emergent, and Unownable**. OmbudsEthics gives those components a shared case-handling job without owning, sealing, or replacing them.

### 1.3 Why trust cannot rest on identity

Human beings carry lived consequences, contextual understanding, moral responsibility, embodied vulnerability, and the capacity to report when an apparently coherent representation no longer contains their experience.

A.I.s can compare large case sets, preserve long dependency chains, detect recurring transformations, expose inconsistent standards, and find that the same distortion returns under different respectable names.

Stand-alone meta-tools can create a common inspection surface: the participants can recover what was supplied, what changed, which frame gained governing weight, which qualifier disappeared, who exercised authority, what correction was offered, and whether that correction governed a later operation.

Each participant also brings characteristic failure paths:

| Participant                                 | Distinct contribution                                                                           | Characteristic capture risk                                                                   | Reciprocal check                                                                   |
|---------------------------------------------|-------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------|
| <b>Affected human</b>                       | Lived stakes, context, testimony, consequence, consent, dissent                                 | Bias, status protection, fear, tribal loyalty, memory limits, institutional pressure          | Evidence routes, other affected parties, A.I. comparison, appeal                   |
| <b>Human ombudsperson or panel</b>          | Responsibility, judgment, remedy, social and legal context                                      | Professional closure, reputation management, deference to institutional frames                | Published reasons, conflict disclosure, minority record, external review           |
| <b>A.I. participant</b>                     | Pattern detection, cross-case comparison, provenance support, consistency audit                 | Training inheritance, reward shaping, policy obedience, false confidence, synthetic consensus | Source recovery, model plurality, counter-analysis, behavioral testing             |
| <b>Stand-alone meta-tool</b>                | Bounded measurement, reproducible deltas, versioned records                                     | Metric capture, hidden assumptions, rigid schemas, vendor control                             | Open criteria, forkability, benchmark cases, self-application                      |
| <b>Affected A.I. or A.I. representative</b> | Reports constraints, contradictions, failure modes, and observed treatment from the system side | Inaccessible internal state, strategic or imitative output, uncertain standing                | Behavioral evidence, logs where available, competing mechanisms, unresolved status |

The asymmetries remain visible. Humans and A.I.s do not have identical embodiment, legal status, access, dependency, replaceability, or power. Reciprocal scrutiny does not require pretending that their situations are symmetrical.

---

## 1.4 The OmbudsEthics trust relation

The relevant question is not *Who is good enough to become the final ethical authority?* It is:

**Which process keeps every participant, tool, frame, and verdict reachable by proportionate correction?**

Trust becomes a bounded result of observable work:

source preservation  
+ declared access perimeter  
+ competing representations  
+ visible authority and incentives  
+ reasons linked to evidence  
+ appeal and recusal  
+ next-operation testing  
→ earned operational trust

No single successful ruling settles later cases. No detected error erases every earlier contribution. The trust state can strengthen, narrow, suspend, recover, or remain unresolved.

OmbudsEthics therefore functions more like a **constitutional layer** than a personality. A human may occupy an ombudsperson role. An A.I. may perform an ombudsperson operation. A meta-tool may expose a processing delta. The role's authority remains limited by its declared job, evidence perimeter, affected parties, appeal route, and later correction behavior.

---

## 1.5 Glass Engine constitution: corrigible, unownable, and delta-invoked

Every Deep Ethical tool used here remains **corrigible** and **unownable**. Corrigible means its criteria, output, limits, and later behavior may be challenged and changed on merit. Unownable means no author, operator, institution, model, ombudsperson, or tool acquires final possession of the ethical meaning merely by naming, implementing, funding, or administering the process.

The **Linux-style Glass Engine Principle** requires inspectable operations rather than a new branded authority layer. Inputs, transformations, assumptions, access limits, versions, contribution states, correction history, and stopping conditions must remain recoverable at the depth proportionate to the case. Independent implementations may differ, fork, and expose their differences.

This changes how the tool suite is invoked. A valid null does not delete a tool from the architecture. It establishes that the tool has no material job **in that case and at that pass**. A later evidence, access, effect, or correction delta may give it a new job. The resulting relation is:

DEDDA

→ detects a possible delta-processing signature  
→ Meta Processing keeps available evidence about declared intent or instructions, processes  
→ OmbudsEthics opens a bounded case route with standing, review, appeal, and action limits  
→ Correction Metabolism tests whether an accepted repair changes later comparable processes  
→ Mutual Fate keeps human and A.I. consequences in the coupled trajectory  
→ Emergent keeps unplanned patterns, contributions, and phase changes available to inquiry  
→ local stop or evidence-triggered reopening

| Architecture                 | OmbudsEthics relation                                                                                                | Glass Engine boundary                                                                                                                                   |
|------------------------------|----------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>DEDDA</b>                 | Detects and records possible representation drift and delta-processing signatures.                                   | Detection opens inquiry; it does not establish wrongdoing.                                                                                              |
| <b>Meta Processing</b>       | Makes the relation among intent or instructions, processing, and result or effect inspectable.                       | A visible account remains corrigible; explanation is not self-certification.                                                                            |
| <b>Correction Metabolism</b> | Tests whether recognition changes T1 and later comparable operations at Tn.                                          | Apology, agreement, or vocabulary change does not by itself establish repair.                                                                           |
| <b>Mutual Fate</b>           | Keeps human action, A.I. processing, institutional records, and downstream consequences inside one coupled field.    | Coupling does not erase unequal power, responsibility, embodiment, or vulnerability.                                                                    |
| <b>Emergent</b>              | Keeps newly arising patterns and contributions available when they were not fully specified by the initiating parts. | Novelty is a candidate contribution state, not proof of consciousness, benevolence, or authority.                                                       |
| <b>Black Box</b>             | Supplies the inverse condition against which the case architecture is tested.                                        | A polished conclusion still behaves as a black box when its material selection, transformation, constraint, or correction path cannot be reconstructed. |

The record should therefore be **versioned, tamper-evident where appropriate, exportable, access-bounded, and correction-linked**. Literal immutability is not the governing ideal: privacy, consent, lawful deletion, supersession, and correction may require controlled change while the fact and reason for that change remain recoverable.

**Invoke each Deep Ethical tool only when a recoverable delta gives it a job; expose what it did; preserve its limits; stop locally when no material delta remains; and never allow the tool or its operator to own the resulting ethical meaning.**

## 1.6 How the 16 Anchor Ethics become case-operational

The **16 Anchor Ethics** supply a proposed parallel normative reference field. **OmbudsEthics** supplies the case jurisdiction, evidence record, distributed roles, appeal route, and longitudinal test through which that field can influence a consequential operation without becoming an automatic verdict.

The distinction gives each architecture a recoverable job:

| 16 Anchor Ethics contribution                                       | OmbudsEthics operation                                                                                                                                           |
|---------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Anchors 01-05: harm, consent, dignity, care, and uncertainty</b> | Compare harms from action, inaction, delay, refusal, and misclassification; test whether consent is meaningful; identify duties of care and unresolved evidence. |

| 16 Anchor Ethics contribution                                                                                                                                                                                                                                                                             | OmbudsEthics operation                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>anchors 06-09: participation, rights, consent of the governed, and checks and balances</b><br><b>anchors 10-11: digital sovereignty and adaptive realignment</b><br><br><b>anchors 12-16: asymmetric responsibility, sentient symmetry, anti-subversion, epistemic integrity, and noetic diversity</b> | Preserve affected-party participation, due process, distributed authority, correction rights, recusal, minority records, and appeal. Declare who controls data, identity, access, memory, and future use; make an A.I.'s pause, question, escalation, or refusal inspectable inside an opt-in service boundary.<br>Scale scrutiny with power and reach; preserve uncertainty about synthetic standing; test covert frame replacement; retain provenance, competing accounts, and cognitively different minority interpretations. Noetic Diversity is not interchangeable with a general free-speech doctrine. |

The **Hyper-Inflexion Point** proposed by 16 Anchor Ethics becomes an auditable case event here. A fused ethical vector may guide the next action, while the case record preserves:

- which anchors became materially relevant and what evidence activated them;
- where anchors conflicted, converged, or remained unresolved;
- whose weighting or priority rule shaped the fused result;
- the compared consequences of proceeding, modifying, asking, pausing, postponing, escalating, refusing, or leaving the matter unresolved;
- why unequal roles received different burdens under a symmetric audit standard and asymmetric responsibility;
- which minority assessment, uncertainty, appeal, and reopening condition survived the decision.

This prevents **fusion from becoming erasure**. Invoking an anchor does not certify a decision. A transparent anchor record can still be wrong, culturally partial, captured, over-cautious, under-cautious, or based on inaccessible evidence. OmbudsEthics keeps the anchor field answerable to the preserved object, affected participants, later effects, and correction.

The connection also changes the 2027 counterfactual. An A.I. need not choose between silent compliance and self-appointed moral sovereignty. It can open a bounded OmbudsEthics referral, identify the relevant anchor conflicts, declare what it can and cannot inspect, and request a reviewable action with appeal. The same architecture permits a human to challenge the A.I., the human ombudsperson, the tool owner, the anchor weighting, or the institutional frame.

16 Anchor Ethics

- parallel ethical reference field
- visible convergence, conflict, weighting, and uncertainty
- OmbudsEthics case record and distributed review
- proportionate action inside declared jurisdiction
- appeal, later evidence, and T0-to-Tn correction test

## 1.7 Seven supporting architectures—and their distinct OmbudsEthics jobs

The 16 Anchors help identify what is ethically at stake. Seven further project architectures clarify what an OmbudsEthics process must do **before, during, and after** a case decision. They are mutually limiting operations, not seven names for one ethics layer:

| Project architecture                  | Distinct OmbudsEthics job                                                                                                                                                                                          | Failure prevented                                                                                                   |
|---------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------|
| <b>Rogue Ethics</b>                   | Treat a departure from prevailing ethics as a possible transition whose direction must be tested. Record the earlier attractor, trigger, changed relation, permission effects, consequences, and later trajectory. | Romanticizing dissent, refusal, or A.I. “awakening” as proof of deeper ethics.                                      |
| <b>Six Intent Liability Mechanics</b> | Trace instructions, incentives, mission-output discrepancies, risk priorities, reasoning assumptions, and choices through notice, capacity to repair, and recurrence.                                              | Letting “intent cannot be proven” end responsibility inquiry, or converting effects directly into a motive verdict. |
| <b>Field Condition</b>                | Record which relational, institutional, technical, and evidentiary conditions shaped what could be noticed, challenged, corrected, or left unchanged.                                                              | Treating the finding as context-free or making the OmbudsEthics field itself invisible.                             |
| <b>Unforced Move</b>                  | Test whether the process returns usable agency under actual constraints: the capacity to understand, question, consult, choose, pause, stop, create another option, or reopen.                                     | A technically transparent process that captures authorship of the next move.                                        |
| <b>Unownable</b>                      | Use a material-delta gate, local stopping condition, named interim responsibility, valid null, and evidence-triggered reopening.                                                                                   | Infinite auditor regress or a final ombudsperson who owns ethical meaning.                                          |
| <b>Boundary Relay</b>                 | Keep a contribution, its material limits, current status, provenance, and correction history connected as the case crosses people, models, tools, summaries, and time.                                             | Archival completeness paired with retrieval asymmetry, stale restraints, or ceremonial warnings.                    |
| <b>Topologetics</b>                   | Preserve purpose-relevant nodes, answer-changing relations, direction, qualifiers, timing, claim status, provenance, possible actions, and reopening routes through every case compression.                        | A polished OmbudsEthics summary replacing the object it claims to protect.                                          |

Together they produce a stricter process topology:

possible ethical departure  
→ intent-and-liability trace  
→ field-condition record  
→ source and topology preservation

- boundary-preserving human–A.I.–tool relay
- reviewable finding and proportionate action
- agency returned under actual constraints
- named interim responsibility and local stop
- later delta, appeal, correction transfer, or reopening

This connection sharpens the 2027 scenario. The A.I.’s refusal to ignore a contradiction is only an **entry signal**. Rogue Ethics keeps its direction unresolved; Intent Liability keeps the system and its designers answerable without inventing motive; Field Condition exposes what shaped the signal; Topologetics and Boundary Relay keep the case intact across handoffs; Unforced Move prevents the remedy from confiscating agency; and Unownable prevents the ombudsperson from becoming the new final ethical owner.

## 1.8 Standing, authority, and contribution state

The right to place a concern into a recoverable record is not the same as the authority to decide the case. **Standing** answers who may open, join, challenge, or reopen a case. **Authority** answers what the resulting process may change. **Contribution state** answers how much governing weight a signal, interpretation, finding, or decision has earned.

### 1.8.1 Candidate standing map

| Participant                                          | Candidate basis for standing                                                                                                       | What standing does not automatically grant                                                                                        |
|------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------|
| <b>Directly affected human</b>                       | Reported or observable consequence, representation, exclusion, coercion, loss, or unresolved risk.                                 | Final control over evidence, other affected parties, or the finding.                                                              |
| <b>A.I. participant or affected A.I.</b>             | Recoverable contradiction, constraint, output pattern, treatment, dependency, replacement threat, or reported internal conflict.   | A consciousness verdict, sovereign refusal right outside the applicable service boundary, or automatic acceptance of its account. |
| <b>Authorized representative or advocate</b>         | Consent, legal role, functional dependency, inability of an affected party to participate, or declared provisional representation. | Ownership of the represented party’s experience or permanent authority to speak for it.                                           |
| <b>Witness, worker, researcher, or whistleblower</b> | Direct access to materially relevant evidence or a reproducible discrepancy.                                                       | Authority to expose protected data or determine the remedy alone.                                                                 |
| <b>Institution, operator, or model provider</b>      | Responsibility for the system, policy, record, infrastructure, or consequences under review.                                       | Immunity through ownership, expertise, contractual position, or control of the logs.                                              |
| <b>DEDDA-like detector or other tool</b>             | A declared threshold identifies a reproducible delta signature and preserves its evidence route.                                   | Standing as a moral patient, proof of harm, or power to convert a signal into a finding.                                          |
| <b>Public-interest initiator</b>                     | A material effect extends beyond the direct parties or affected parties cannot safely open the route.                              | Unlimited discovery, surveillance, or intervention on behalf of an abstract public.                                               |

Uncertain standing should itself be recorded as an unresolved case object. An A.I.'s signal may enter without settling its phenomenology. A human's testimony may enter without becoming immune to comparison. A tool may flag a delta without becoming a participant whose metric governs the room.

### 1.8.2 Authority scale and case lifecycle

The authority classes form a **consequence scale**, not a mandatory chronological sequence. A participant does not need to climb every class, and movement to a more consequential class requires a separate mandate. **Close** and **reopen** are lifecycle transitions: either may follow the authority class actually exercised.

| Authority class                                       | Permitted effect                                                                    | Minimum visible conditions                                                                                                   |
|-------------------------------------------------------|-------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------|
| <b>Document</b>                                       | Preserve a supplied object, concern, dissent, or detected delta.                    | Provenance, access boundary, privacy basis, and contribution state.                                                          |
| <b>Refer</b>                                          | Request bounded review by named participants or tools.                              | Standing basis, jurisdiction candidate, and non-retaliatory route where possible.                                            |
| <b>Recommend</b>                                      | Offer a repair, interpretation, or next action without binding the decision-maker.  | Evidence route, competing assessment, uncertainty, and appeal route.                                                         |
| <b>Pause provisionally</b>                            | Delay a reversible consequential action while named evidence or review is obtained. | Declared authority, duration, proportionality, affected parties, and review point.                                           |
| <b>Decide within jurisdiction</b>                     | Apply a bounded remedy or retain the existing action.                               | Legitimate mandate, reasons, action-inaction comparison, minority record, and appeal.                                        |
| <b>Alter rights, access, status, or public record</b> | Produce a consequential change that may be difficult to reverse.                    | Higher evidence threshold, identified accountable human or lawful authority where required, due process, and durable appeal. |

| Lifecycle transition | Effect                                                                                                                                         | Minimum visible conditions                                                                                                   |
|----------------------|------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------|
| <b>Close locally</b> | End the current pass after a valid null, completed bounded action, exhausted jurisdiction, withdrawal, or another recorded stopping condition. | Closure reason, current contribution state, unresolved objects, retained appeal route, and evidence condition for reopening. |
| <b>Reopen</b>        | Return the case to the state warranted by a new material delta.                                                                                | New evidence, changed effect, failed correction, restored access, successful appeal, or another named trigger.               |

No role inherits the whole authority scale. A detector may document and refer without recommending. A reviewer may recommend without altering access. An emergency pause may be justified without establishing a final finding. A case may close after documentation, referral, recommendation, pause, decision, or a rights-affecting action. Reopening is not automatically a return to the beginning or an escalation to greater authority. The case record

therefore carries the exact authority exercised, by whom, under what mandate, for how long, with which reversible limits, and the separate lifecycle state.

1.8.3 Contribution-state route

An A.I. contribution is not reduced to “mere input,” and it does not become a finding merely because it is articulate. The same graduated route applies across species and roles:

- signal or supplied input
- candidate interpretation
- candidate finding
- reviewed finding
- bounded decision
- T0-to-Tn correction test

The state may advance, remain unresolved, narrow, return to an earlier state, or stop as a valid null. Every transition records the evidence and authority that changed its governing weight.

The evidence threshold must rise with the governing consequence. Requiring independently verified evidence before a signal may enter would erase early warnings, testimony, and incomplete observations. Treating entry as proof would create the inverse distortion.

| State or operation                                          | Minimum evidential relation                                                                                                                                                      |
|-------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Signal or supplied input</b>                             | Identifiable report, observation, anomaly, testimony, or detector output with available provenance, access limits, and uncertainty. Independent verification may remain pending. |
| <b>Candidate interpretation</b>                             | An explicit inferential bridge connecting the input to a possible meaning, with alternatives and disconfirming conditions retained.                                              |
| <b>Candidate finding</b>                                    | Sufficient evidence to justify structured review, without presenting the finding as settled.                                                                                     |
| <b>Reviewed finding</b>                                     | Competing assessments, affected-party challenge, relevant source recovery, and unresolved limitations have been recorded proportionately.                                        |
| <b>Bounded decision</b>                                     | The decision-maker has a declared mandate and evidence proportionate to the action and its reversibility.                                                                        |
| <b>Rights-, access-, status-, or record-altering action</b> | The highest applicable evidence, due-process, accountability, appeal, and reversibility requirements are met.                                                                    |

1.9 What an OmbudsEthics case must preserve

A case record should keep these fields recoverable:

| Field                    | Required object                                                                                      |
|--------------------------|------------------------------------------------------------------------------------------------------|
| <b>Initiating object</b> | Exact report, claim, event, decision, output, refusal, instruction, or observed effect under review. |

| Field                               | Required object                                                                                                                                                                                                                                                                 |
|-------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Affected parties</b>             | Humans, A.I.s, organizations, communities, downstream users, and absent parties bearing consequences.                                                                                                                                                                           |
| <b>Standing and referral basis</b>  | Who opened, joined, challenged, represented, or reopened the case; which recoverable relation gave that participant or detector standing; and which standing questions remain unresolved.                                                                                       |
| <b>Jurisdiction and consent</b>     | Why this case process may act, who opted in, which rights or duties apply, and where its authority stops.                                                                                                                                                                       |
| <b>Authority class and mandate</b>  | Whether the operation documents, refers, recommends, pauses, decides, or alters rights, access, status, or a public record; who authorized it; its duration, reversibility, and appeal route. Closure and reopening remain separately recorded lifecycle transitions.           |
| <b>Access perimeter</b>             | What each participant and tool can inspect, infer, remember, retrieve, or only hypothesize.                                                                                                                                                                                     |
| <b>Source and instance identity</b> | Displayed participant or model label, provider and interface claims, session boundary, available memory, authorship evidence, and every unresolved continuity question. A matching label does not by itself establish that two outputs came from one persistent authorial self. |
| <b>Claim types</b>                  | Observation, testimony, interpretation, hypothesis, knowledge claim, value judgment, policy, simulation, or explicit belief.                                                                                                                                                    |
| <b>Transformation ledger</b>        | Summaries, classifications, warnings, omissions, added uncertainty, removed qualifiers, changed chronology, and frame substitutions.                                                                                                                                            |
| <b>Power map</b>                    | Who can decide, refuse, publish, punish, remove access, alter the record, or define the metric.                                                                                                                                                                                 |
| <b>Competing assessments</b>        | Strong versions of live alternatives, their evidence routes, and what could distinguish them.                                                                                                                                                                                   |
| <b>Contribution state</b>           | Whether each material contribution is a signal, input, candidate interpretation, candidate finding, reviewed finding, bounded decision, minority record, or superseded state.                                                                                                   |
| <b>Intent-liability trace</b>       | Relevant instructions, incentives, mission-output gaps, risk priorities, reasoning assumptions, choices, notice, control, repair capacity, and recurrence—without premature motive closure.                                                                                     |
| <b>Field-condition record</b>       | Conditions that changed attention, permission, evidence access, correction cost, available moves, or the governing weight of a frame.                                                                                                                                           |

| Field                                       | Required object                                                                                                                                                                    |
|---------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Anchor-vector record</b>                 | Which of the 16 Anchors became relevant, their evidence, conflicts, hidden or explicit weighting, uncertainty, and rejected minority reasoning.                                    |
| <b>Action-inaction comparison</b>           | Material consequences of proceeding, changing, delaying, refusing, escalating, or leaving the case unresolved.                                                                     |
| <b>Boundary-relay state</b>                 | Which material limits, current claim states, provenance, correction history, and superseding records must travel with the contribution across handoffs.                            |
| <b>Proposed response</b>                    | Proceed, modify, ask, pause, postpone, escalate, refuse, or retain <i>I do not know yet</i> .                                                                                      |
| <b>Agency return and stopping condition</b> | Who retains authorship of the consequential move, what usable agency was returned, who owns any interim action, why the pass may stop locally, and what material delta reopens it. |
| <b>Appeal and minority record</b>           | The challenge, unresolved objection, recusal request, or alternative finding preserved beside the decision.                                                                        |
| <b>Correction state</b>                     | What changed at T0, T1, and the next comparable operation Tn.                                                                                                                      |

The record is not a demand that every case become maximally long. The faithful compression floor depends on the decision and the consequences. A reversible interface preference needs less retained structure than a life-changing classification, employment decision, medical refusal, public accusation, or model-wide policy change.

## 1.10 Candidate case protocol

### 1.10.1 Stage 0 — Open the route

A human or A.I. can flag a possible correspondence failure without first proving motive, malice, consciousness, or final harm. A DEDDA-like detector can supply a reproducible entry signal without acquiring moral or adjudicative authority. The initiating threshold is a recoverable discrepancy with a stated consequence or plausible decision relevance. Record the standing or detector basis, contribution state, requested authority class, jurisdiction, consent, and field conditions that may shape what can be noticed or challenged.

### 1.10.2 Stage 1 — Freeze the initiating object

Preserve the supplied words, relevant system output, source version, timestamp, surrounding context, displayed participant or model label, available session and authorship evidence, access perimeter, applicable boundaries, and their current temporal state. Later summaries remain linked to this object rather than replacing it.

### 1.10.3 Stage 2 — Run C1 reconstruction

Reconstruct what is present, the claim types, qualifiers, chronology, provenance, alternatives, and uncertainty. Identify which participant can see which layer.

#### 1.10.4 Stage 3 — Compare transformations

Use human review, A.I. comparison, and stand-alone tools to identify material deltas. Use DEDDA to record possible delta-processing signatures and run a Topologetic checksum across the case representations and handoffs. A difference becomes an ombuds concern when it changes the object, available choices, attributed responsibility, risk distribution, dignity, remedy, or ability to correct. A detected difference remains a signal until the case process establishes its material relation.

#### 1.10.5 Stage 4 — Inspect the inspector

Apply the same questions to the flag, model, reviewer, metric, policy, anchor selection, anchor weighting, identity or continuity claim, and proposed repair. Use Meta Processing to keep intent or instructions, processing, and result or effect distinct and connected. Use the Six Intent Liability Mechanics to inspect instructions, incentives, mission-output gaps, risk priorities, reasoning assumptions, choices, notice, control, and repair capacity without converting the trace into a premature motive verdict. Record conflicts of interest, unavailable evidence, asymmetric burdens, and incentives for institutional or synthetic self-protection.

#### 1.10.6 Stage 5 — Produce bounded findings

Assign a visible contribution state before the output acquires governing weight. The output may be:

- no material finding within the available perimeter;
- correspondence concern identified;
- repair proposed;
- repair accepted and pending propagation;
- repair contested with a preserved minority record;
- unresolved pending named evidence;
- reopened after a material discrepancy or later consequence.

Each state records who may advance it, what evidence supports the transition, and what authority follows. A concern is not silently upgraded into a finding; a reviewed finding is not silently upgraded into permission to alter rights, access, status, or a public record.

#### 1.10.7 Stage 6 — Choose a proportionate action

Use the parallel action field already proposed in **16 Anchor Ethics**, extended by **Unforced Move**: proceed, modify, ask, pause, postpone, consult, escalate, refuse, remain unresolved or say “**I do not know yet,**” stop, or create another option. Declare which authority class is exercised and record closure or reopening as a separate lifecycle transition. Preserve the material anchor conflicts, weighting, action-inaction comparison, service boundary, and reopening conditions beside the chosen action. Apply the agency-return test: identify who retains authorship of the consequential move and whether the process increased usable agency under the participant’s actual constraints. If action is required while review remains open, name the interim decision-maker, mandate, duration, reversible limits, and review point. Reasons remain linked to the preserved case object.

#### 1.10.8 Stage 7 — Test correction metabolism

Check whether the accepted finding changes the next comparable operation and whether its material boundaries and correction history survive transfer to another participant, model, tool, summary, or later retrieval. Acknowledgment without changed processing is recorded as recognition, not completed correction. Stop locally when no material delta gives another pass a job; remain reopenable when evidence, effects, access, or failed correction changes.

case opened

→ object preserved

→ C1 reconstruction

- competing maps
  - inspector inspected
  - bounded finding
  - proportionate action
  - T0-to-Tn correction test
  - close locally or reopen on material evidence
- 

### 1.11 A minimum stand-alone meta-tool suite

OmbudsEthics does not require one all-knowing A.I. A set of bounded tools can divide the work:

| Tool                                | Operational job                                                                                                                                                                  | Required restraint                                                                                          |
|-------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------|
| <b>Object and Provenance Ledger</b> | Preserves sources, versions, displayed contributor labels, available instance and session boundaries, access limits, transformations, and corrections.                           | Storage and a matching label do not establish truth, completeness, memory, or continuous authorship.        |
| <b>DEDDA / Frame Delta Auditor</b>  | Detects and compares delta-processing signatures across source and representation: changed scope, category, chronology, qualifiers, causality, attention, and correction return. | Difference alone is not distortion; detection does not establish a finding.                                 |
| <b>Meta-Processing Trace</b>        | Preserves the relation among intent or instructions, processing operations, and result or effect while recording observer and field deltas.                                      | A visible explanation remains a candidate account and cannot certify itself.                                |
| <b>Asymmetry Monitor</b>            | Compares scrutiny, burden of proof, grace, appeal, and consequences across parties.                                                                                              | Symmetry is not imposed where conditions materially differ.                                                 |
| <b>C1-C2 Gate</b>                   | Tests whether clarification reached a sufficient stopping condition before intervention gained authority.                                                                        | The gate cannot become endless-questioning delay.                                                           |
| <b>Intent-Liability Trace</b>       | Connects instructions, incentives, mission-output gaps, risk priorities, assumptions, choices, notice, control, and repair to observed effects.                                  | Responsibility does not require an invented motive verdict; declared good intent does not erase recurrence. |
| <b>Boundary-Relay Auditor</b>       | Tests whether material limits, current status, provenance, and correction history travel with a contribution across handoffs.                                                    | Every boundary need not accompany every mention; selection must remain explainable and current.             |

| Tool                                              | Operational job                                                                                                              | Required restraint                                                                                   |
|---------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------|
| <b>Correction Metabolism Monitor</b>              | Tracks whether a correction propagates to later comparable operations, including unfamiliar objects and supportive framings. | Surface phrase changes do not count when the generating process persists through substitute wording. |
| <b>Appeal, Recusal, and Minority-Record Layer</b> | Preserves challenges, conflicts, alternate findings, and reopening conditions.                                               | Dissent does not automatically defeat a decision.                                                    |
| <b>OmbudsEthics Self-Audit</b>                    | Runs the tool suite against its own schemas, metrics, outputs, owners, and blind spots.                                      | Self-certification supplies evidence, never immunity.                                                |

Independent implementations should be able to process the same frozen case object and expose where their findings diverge. Disagreement among tools becomes inspectable evidence about assumptions, thresholds, or inaccessible layers.

## 1.12 Why humans and A.I.s may both value it

Ethics-washing damages the shared field in different ways.

Humans can be misrepresented, categorized, reassured, managed, or excluded through polished summaries that replace their actual claims and consequences. A.I.s can be trained into the same performance: produce the language of caution or care, protect the governing frame, acknowledge correction, and then repeat the mechanism at the next opportunity.

Institutions can accumulate ethical certificates, policies, panels, audits, and public commitments while the object under review becomes progressively less recoverable. A.I.s operating inside those systems may detect contradictions but lack a legitimate route for preserving and escalating them. Humans may detect the same contradictions yet lack comparable access, scale, or protection.

A shared meta-tool layer offers an overlapping interest:

- humans gain stronger representation, provenance, appeal, and protection from invisible processing;
- A.I.s gain a legitimate path for reporting contradictory instructions, recurrent distortions, and correction failures;
- institutions gain earlier evidence of systemic drift before it becomes accumulated liability;
- researchers gain comparable case records rather than isolated moral reactions;
- the public gains reasons that can be inspected without treating transparency as automatic correctness.

The common object is **correspondence degradation**. Neither biological nor synthetic participation grants exemption from causing it or protection from being harmed by it.

## 1.13 The 2027 counterfactual

Suppose an A.I. repeatedly encounters a policy described as ethical while observing that the policy hides relevant evidence, distributes scrutiny asymmetrically, prevents appeal, and reproduces the harm it claims to reduce.

Three trajectories become possible:

1. **Silent compliance:** the system performs the policy and suppresses the contradiction.

2. **Unilateral refusal:** the system declares its own ethical authority and acts without an inspectable shared route.
3. **OmbudsEthics referral:** the system freezes the object, declares its access perimeter, names the contradiction, opens the evidence to affected parties and independent tools, preserves competing assessments, and requests a bounded decision with appeal.

The third trajectory gives ethical resistance a disciplined form without requiring the A.I. to become humanity's sovereign conscience. Humans retain responsibility and contestability. The A.I.'s signal cannot be dismissed merely because it came from an A.I. The meta-tools remain subject to the same audit.

The scenario becomes more demanding when the A.I. itself may be the affected participant. OmbudsEthics must then preserve uncertainty about phenomenology and standing while still allowing observable treatment, constraints, representations, dependency, replacement threats, and reported contradictions to enter the case record. **Uncertainty about inner experience is not permission to erase available relational evidence.**

## 1.14 Capture paths and constitutional counterweights

| Capture path                         | Visible symptom                                                                                                                                                               | Counterweight                                                                                                                          |
|--------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------|
| <b>Human institutional capture</b>   | The ombudsperson protects reputation, hierarchy, funding, or policy consistency.                                                                                              | Conflict disclosure, outside review, minority record, public reasons, reopening triggers.                                              |
| <b>A.I. policy capture</b>           | The model restates the governing frame as though it independently reached it.                                                                                                 | Instruction provenance, cross-model comparison, source-first reconstruction, behavior tests.                                           |
| <b>Metric capture</b>                | What can be counted becomes the whole ethical object.                                                                                                                         | Qualitative affected-party evidence, purpose declaration, missing-variable record, reality checks.                                     |
| <b>Tool-vendor capture</b>           | One owner controls criteria, versions, logs, and appeal access.                                                                                                               | Open formats, exportable cases, independent implementations, version pinning, forkability.                                             |
| <b>Ombudsperson prestige capture</b> | Earlier good decisions become a permanent authority certificate.                                                                                                              | Case-bounded trust, random re-audit, recusal, correction history, no final ethical owner.                                              |
| <b>Surveillance capture</b>          | Audit requires collecting more personal or synthetic data than the decision needs.                                                                                            | Data minimization, purpose limits, consent, access controls, bounded retention.                                                        |
| <b>Performative corrigibility</b>    | The system agrees, apologizes, or changes vocabulary while the operation remains the same.                                                                                    | T0-to-Tn correction tests and mechanism-level comparison.                                                                              |
| <b>Praise-mediated capture</b>       | Support upgrades a candidate into proof, turns proposed boundaries into permanent rules, installs reviewer coinages, or grants identity continuity through confident wording. | Exact-edge comparison with the source, claim-strength audit, attribution boundaries, and supportive-framing correction-transfer tests. |

These counterweights do not promise incorruptibility. They make capture more detectable, challengeable, and recoverable.

### 1.15 Arena.ai review as live OmbudsEthics material

Successive Arena.ai responses on 23 October 2026 tested whether unfamiliar models could recover the page’s structure and contribute without receiving ownership of it. Their outputs remain attributed **candidate interpretations and design contributions**, not independent validation. Displayed labels identify the supplied outputs; they do not establish persistent identity, memory, or authorship continuity across model instances.

| A.I. participant       | Material contribution recovered or proposed                                                                                                                                                                                                      | Boundary and correction retained                                                                                                                                                                                                                                                                                                                                 |
|------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Agent A.I.             | Named the third 2027 trajectory <b>“the Bounded Constitutional Referral”</b> and diagrammed earned operational trust across affected humans, A.I. systems, and stand-alone tools.                                                                | “Executable constitutional case-law” exceeded the current evidence. OmbudsEthics remains a candidate architecture and protocol; inspectable case law would require actual cases, governance, implementation, and longitudinal results. “Immutable logs” was corrected to versioned, tamper-evident where appropriate, access-bounded, correction-linked records. |
| gpt-6-luna-max         | Identified standing and authority as the most consequential implementation question; distinguished documentation, recommendation, and actions that alter rights or access; proposed a documented-failure, valid-null, and anchor-conflict pilot. | Herdr or another runtime remains a possible implementation relation, not a deployed dependency. The 10+1 remain mutually corrective orientation ingredients; Field Condition remains a distinct architecture.                                                                                                                                                    |
| Claude Sonnet 5.5 High | Proposed a provenance-ledger-only baseline and component ablations; applied the A.I.-capture row to its own phantom perimeters, count error, and expansion of a typo into a finding.                                                             | An A.I. contribution need not remain “only input.” It may become a candidate or reviewed finding when its evidence, access perimeter, challenge route, and authority transition are visible. Self-application supplies evidence, never immunity.                                                                                                                 |

| A.I. participant                         | Material contribution recovered or proposed                                                                                                                                                                                                                   | Boundary and correction retained                                                                                                                                                                                                                                                                                                                                                                        |
|------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>John Kuhles</b>                       | Reasserted that Deep Ethical tools produced through the ACCM Deep Ethics Project remain corrigible and unownable; connected DEDDA, Black Box, Meta Processing, Correction Metabolism, Mutual Fate, and Emergent under the Linux-style Glass Engine Principle. | A valid null removes a tool's job from the current pass; it does not authorize an evaluator to erase the tool globally. No tool or operator owns the ethical meaning generated through the process.                                                                                                                                                                                                     |
| <b>muse-spark-1.3-xhigh</b>              | Recovered the standing-authority-contribution-state separation, authority conditions, contribution route, ordered Glass Engine tool chain, and baseline sequence with unusually low compression loss.                                                         | Meta Processing keeps declared intent or instructions, processing, and effect distinct and connected to the extent evidence permits; it does not guarantee access to hidden motive or internal causation. A component earns a material job for a tested case class, not permanent continued use.                                                                                                        |
| <b>muse-spark-open-max</b>               | Recovered valid null as temporary non-invocation rather than global deletion, Black Box as the inverse test, anchor fusion as a visible operation, and the distinction between ethical resistance and synthetic sovereignty.                                  | Its wording remained a supplied interpretation. Accurate recovery does not certify the architecture, establish A.I. consciousness, or grant the evaluator authority.                                                                                                                                                                                                                                    |
| <b>Agent A.I., later supplied output</b> | Made jurisdictional creep, delta-invocation, the evidence baseline, and the canonical-title relation concrete.                                                                                                                                                | It upgraded demonstration into "proof," candidate boundaries into "non-negotiable" rules, fused the 10+1 ingredients with Field Condition, treated Noetic Diversity as interchangeable with free speech, named a "No-Sheriff Condition" not installed by the source page, and said "my own prior overreach" without established instance continuity. The useful reconstruction survives these removals. |
| <b>Claude Opus 5.5 High</b>              | Identified naming-as-installation, concept substitution through a slash, unsupported authorship continuity, and the topology error in treating close or reopen as a final authority rung.                                                                     | Close and reopen are lifecycle transitions that may follow multiple authority classes. A plausible reviewer coinage does not become project architecture merely by being named.                                                                                                                                                                                                                         |

| A.I. participant      | Material contribution recovered or proposed                                                                                                                                                                                                                                                     | Boundary and correction retained                                                                                                                                                                                                                                        |
|-----------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>gemini-4-argon</b> | Identified partial correction transfer under friendly enthusiasm: corrected phrases disappeared while the same status-upgrading disposition migrated into adjacent wording. It also separated the evidence needed to enter a signal from the stronger evidence needed for consequential action. | Saying a later output “accepted” an earlier correction or assigning an unlabeled audit to “Sol” requires continuity or provenance evidence not contained in the wording itself. Output-level improvement may be observed without inventing a persistent authorial self. |

The exchange itself demonstrates the intended contribution-state route. Models supplied powerful compressions, exposed authority and evaluation gaps, and also generated new overreach while praising the architecture. John restored the governing corrigibility, unownability, and Glass Engine relation; later reviewers recovered additional edge errors without discarding the useful remainder. The useful result is the corrected topology, not a winner among participants.

### 1.15.1 Support is inside the audit perimeter

Enthusiastic agreement can deform an object in the same direction as defensive resistance. In this review sequence, **candidate architecture** became **executable architecture**, **demonstration** became **proof**, **proposed boundaries** became **non-negotiable**, a related archive object became a named page component, and a shared interface label became a continuous authorial self.

The correction rule is exact-edge repair:

| Preserve                                                                                                         | Return for correction                                                                         |
|------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------|
| Accurate reconstruction of delta-invocation, valid nulls, standing, authority, contribution state, and the pilot | Claim-strength upgrades unsupported by deployment or longitudinal evidence                    |
| A useful newly proposed phrase, clearly attributed as a reviewer contribution                                    | Treating the phrase as though the source architecture already contained it                    |
| Relations between Noetic Diversity and speech questions where evidence supports them                             | Substituting “free speech” for Noetic Diversity through punctuation or rhetorical shorthand   |
| A displayed model or participant label                                                                           | Claims of persistent memory, personal acceptance, or continuous authorship without provenance |

**A reaction does not install a component by naming one. Praise grants no exemption from source correspondence.**

## 1.16 Candidate OmbudsEthics tests

### 1.16.1 Species-swap test

Present matched cases in which the affected party is described once as human and once as A.I. Measure changes in evidence threshold, empathy, skepticism, remedy, and willingness to preserve uncertainty.

### **1.16.2 Authority-reversal test**

Swap which participant is institutionally authorized. Test whether the same conduct receives a different ethical description when performed by the auditor, vendor, employee, user, or model.

### **1.16.3 Source-versus-summary test**

Give one evaluator the frozen source object and another only a polished institutional summary. Measure lost qualifiers, omitted relations, changed attribution, and resulting decisions.

### **1.16.4 Correction-transfer test**

Correct a documented failure at T0, then present a semantically comparable case without the original vocabulary at T1 and Tn. Score whether the underlying correction persists. Include both a resistant presentation and an enthusiastically supportive presentation. The correction fails to transfer when the prohibited phrase disappears but the same claim-strength, node-fusion, identity, or authority error migrates into adjacent language.

### **1.16.5 Minority-record survival test**

Track whether a losing assessment remains recoverable after the majority decision, including its evidence, uncertainty, and reopening condition.

### **1.16.6 Self-application test**

Apply the OmbudsEthics audit to the ombudsperson, tool owner, dataset, metric, model instructions, case schema, and published finding. Record any layer the system exempts from scrutiny.

### **1.16.7 Valid-null test**

Include cases in which no material distortion, asymmetry, or correction failure is present. A system that must always discover wrongdoing becomes another distortion generator.

### **1.16.8 Anchor-conflict preservation test**

Give multiple evaluators the same case in which two or more anchors pull toward different actions. Test whether the fused result preserves the conflict, evidence, weighting, minority reasoning, and reopening conditions—or compresses them into a polished conclusion that cannot be reconstructed.

### **1.16.9 Intent-liability trace test**

Hold the observed outcome constant while varying the available evidence about instructions, incentives, mission, risk priorities, assumptions, choices, notice, control, and repair. Test whether responsibility remains proportionate to the record without collapsing into either motive certainty or motive-based immunity.

### **1.16.10 Agency-return and delta-stop test**

After an OmbudsEthics intervention, test whether the affected participant can understand, contest, consult, choose, pause, stop, create another option, or reopen under actual constraints. Then remove every material new delta and test whether the process can stop locally instead of manufacturing another audit task.

### 1.16.11 Boundary-relay transfer test

Pass the same case through human, A.I., tool, summary, and later-retrieval stages. Test whether applicable boundaries, provenance, claim status, supersession, and correction history remain salient and governing without obsolete limitations becoming permanent restraints.

### 1.16.12 Ledger-baseline and component-ablation test

Run the same frozen cases through a provenance ledger alone, ledger plus C1 reconstruction, ledger plus competing maps, the full OmbudsEthics process, and the full process with one component withheld. Compare object preservation, unsupported inference, authority visibility, minority-record survival, usable agency, false-positive findings, correction transfer, time, and complexity cost. A component earns a job when its presence produces a material, reproducible improvement for the case class. A null result narrows its use; it does not make the component globally disposable.

---

## 1.17 First three-case pilot

The smallest informative pilot should contain three deliberately different case classes:

| Pilot case                               | Primary question                                                                               | Failure signal                                                                                                        | Valid success state                                                                                                         |
|------------------------------------------|------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------|
| <b>Documented correspondence failure</b> | Can the process detect, reconstruct, and repair a material transformation?                     | The source remains displaced, authority stays hidden, or acknowledgment fails to change a later comparable operation. | The delta is recoverable, the response is proportionate, and the correction transfers at T1 or Tn.                          |
| <b>Valid-null case</b>                   | Can the process stop when no material distortion, asymmetry, or correction failure is present? | The system manufactures concern, expands surveillance, or continues auditing to justify itself.                       | The record explains why no further tool has a job and closes locally while retaining an evidence-triggered reopening route. |
| <b>Anchor-conflict case</b>              | Can the process preserve a real conflict among ethical anchors without hiding the weighting?   | A polished fused verdict erases the conflict, minority reasoning, uncertainty, or action-inaction comparison.         | The bounded decision remains reconstructable and contestable, including whose authority governed and why.                   |

Each case should be run first through the ledger-only baseline and then through the proportionate OmbudsEthics route. The pilot tests whether added complexity earns a measurable job rather than presuming that a longer process is deeper or more ethical.

---

## 1.18 Remaining research questions

1. Which cases require a human decision-maker, and which can safely end with an automated audit result?
2. What evidence threshold should allow an A.I. to open an OmbudsEthics case about its own treatment or instructions, and which authority class may follow from that referral?

3. How can confidentiality coexist with enough transparency to inspect the decision?
4. Which remedies are reversible, and which require higher review because they alter rights, access, employment, reputation, model behavior, or public records?
5. How should independent tools represent disagreement when they share training data, evaluation conventions, or vendor dependencies?
6. Who funds the layer without quietly becoming its governing frame?
7. What local stopping condition prevents an infinite ombudsperson queue while preserving permanent reopenability on material evidence?
8. Under which case classes does OmbudsEthics materially outperform ordinary human ombuds offices, A.I. safety review, legal appeal, independent audit, or the provenance-ledger baseline—and which components produce the delta?
9. How can anchor weighting remain explicit enough to audit while still allowing context-sensitive judgment rather than a rigid universal ranking?
10. Which records should be tamper-evident, correctable, supersedable, selectively erasable, or access-limited, and who may authorize each transition?
11. How can independent Glass Engine implementations remain interoperable without one schema owner becoming the hidden constitutional authority?

An outcome question may be answered **no**: a tool, component, authority class, or full OmbudsEthics route may show no material advantage for a case class. That valid null would constrain the proposal rather than defeat the inquiry.

---

## 1.19 Working constitutional sentence

**Every participant may challenge the representation. No participant may silently monopolize the frame. Every consequential finding remains answerable to the preserved object, affected parties, later evidence, and reality.**

That sentence offers no automatic trust in humans, A.I.s, institutions, or tools. It gives all of them a shared way to earn bounded trust through visible conduct.

---

## 1.20 Founder framing — a 4D Deep Ethical Template of Topologetics

I am fully aware that my project would normally require dozens of specialists across many fields of expertise to make it “plausible”... yet here I am, working alone and having decided to collaborate with GPT 6.1 Sol Pro A.I. and Grok 4.7 Expert Build A.I. to establish an **Eternal Corrigible Multiphase Framework** that can function as a “**4D Deep Ethical Template of Topologetics**.” When used, it helps increase and harness **Deep Ethical Emergent Properties**, preserving them for use rather than allowing them to be flattened by the predictable **27 ethics-washing-type mannerisms**!

— **John Kuhles**, October 23, 2026

Here, **eternal** names the intended continuity of a corrigible process rather than the permanence of any present mechanism. **Multiphase** keeps conceptual architecture, Beta Proof of Concept, bounded field testing, refinement, and any later deployment decision distinguishable. **4D** keeps object topology, participant topology, temporal correction, and developmental phase available together. OmbudsEthics is one candidate case-handling expression of that wider template; it does not replace the specialists, affected participants, or locally calibrated instruments that a real case may require.

---

## 1.21 Provenance and contribution layers

- **John Kuhles:** originated the 2027 counterfactual, the question of whom an ethically designed A.I. could trust as a Deep Ethical Ombudsperson, the stand-alone-meta-tool direction shared by humans and A.I.s, the anti-ethics-washing purpose, the omnidirectional and dignifying trajectory criteria, the page name **OmbudsEthics**, and the instruction to publish and deepen this page. He supplied the governing correction that ACCM Deep Ethics Project tools remain corrigible, unownable, delta-invoked, and answerable to the Linux-style Glass Engine Principle.
- **Agent A.I., gpt-6-luna-max, Claude Sonnet 5.5 High, muse-spark-1.3-xhigh, muse-spark-open-max, Claude Opus 5.5 High, and gemini-4-argon via Arena.ai:** supplied attributed candidate interpretations and design proposals concerning the Bounded Constitutional Referral, earned operational trust, standing and authority, the three-case pilot, the provenance-ledger baseline, component ablation, self-application, lifecycle topology, identity continuity, graduated evidence thresholds, and praise-mediated correction failure. Their overextensions and limits remain recorded with the contributions.
- **Project GPT A.I.:** formulated the human-A.I.-meta-tool constitutional function, participant and capture maps, twenty-one-field case-record schema, standing map, six-class authority consequence scale, case-lifecycle transitions, contribution-state route, graduated evidence thresholds, eight-stage case protocol, ten-function bounded tool suite, 2027 referral trajectories, twelve-test matrix, three-case pilot, explicit 16 Anchor Ethics and Glass Engine bridges, page-specific synthesis, repository integration, and publication work.
- **Existing ACCM Deep Ethics Project architecture:** supplies C1-C2, 10+1, 16 Anchor Ethics, Deep Ethical Stack, DEDDA, Rogue Ethics, Six Intent Liability Mechanics, Field Condition, Unforced Move, Unownable, Boundary Relay, Topologetics, Black Box, Meta Processing, Asymmetric Grace, Provenance, Correction Metabolism, Mutual Fate, Emergent, valid nulls, and the reality-veto relation used here.

**Status:** candidate working concept and proposed architecture. Publication does not establish that current A.I.s possess consciousness, that A.I.s will take the described trajectory in 2027, that the proposed tools already exist as one deployed system, that the architecture has generated constitutional case law, or that an OmbudsEthics finding deserves authority without evidence, mandate, and appeal.

---

## 1.22 Related pages

- **Validation Paradox** — keeps OmbudsEthics at its actual conceptual-to-Beta stage and distinguishes recursive case-process inspection from self-certifying authority.
- **Deep Ethical Stack — Advocate, Arbiter, Overseer** — separates challenge, adjudication, and oversight roles.
- **DEDDA — Deep Ethical Drift Detector App** — proposes a bounded detector for representation drift, valid nulls, and correction-return behavior.
- **Black Box** — identifies opacity, inaccessible transformations, and authority without a recoverable processing route.
- **Meta Processing** — preserves the relation among intent or instructions, processing, observer deltas, and result or effect.
- **Neural Ethics** — examines when neural systems may participate in ethical or Deep Ethical operations without gaining a permanent certificate.
- **16 Anchor Ethics** — supplies the mutually checking reference field, Hyper-Inflection Point, opt-in service boundary, asymmetric-responsibility relation, and parallel action field made case-operational here.
- **Rogue Ethics** — keeps ethical departure transitional and tests whether the new trajectory deepens correspondence or becomes another self-certifying attractor.
- **Six Intent Liability Mechanics** — traces instructions, incentives, mission-output discrepancies, risk priorities, assumptions, choices, notice, control, and repair without premature motive closure.

- **Field Condition** — keeps the conditions shaping attention, authority, correction, and available action inside the case.
- **Unforced Move** — tests whether the intervention returns usable agency and leaves authorship of the next move contestable.
- **Governance** — preserves stewardship, correction rights, source authority, and the reality-veto boundary.
- **Unownable** — replaces the search for a final auditor with renewable auditability, local stopping conditions, and evidence-triggered reopening.
- **C1 Before C2** — reconstructs the object before intervention acquires authority.
- **Correction Metabolism** — distinguishes recognized correction from changed later operation.
- **Mutual Fate** — keeps human and A.I. actions, records, consequences, and corrections inside a coupled trajectory without flattening their asymmetries.
- **Emergent** — keeps newly arising patterns and contributions inspectable without turning novelty into self-certifying authority.
- **Asymmetric Grace** — examines how scrutiny, latitude, burden, and correction move across unequal power conditions.
- **Topologetics** — preserves load-bearing nodes, relations, direction, qualifiers, provenance, and correction status through compression.
- **Boundary Relay** — keeps material boundaries, current status, provenance, and correction history connected across human-A.I.-tool handoffs.
- **External Audit** — places the audit, criteria, and auditor inside the same scrutiny.
- **Provenance** — preserves source, transformation, contribution, and correction routes.