

27 + 12

ACCM Deep Ethics Project

Contents

1	27 + 12 — Current Working Architecture	1
1.1	The 27 correspondence obstructions	2
1.2	The true 12 from the source file	2
1.2.1	Stage 1 — Suspend the Agreement Reflex	2
1.2.2	Stage 2 — Construct Before Judging	3
1.2.3	Stage 3 — Distinguish Levels	3
1.2.4	Stage 4 — Avoid Premature Reduction	3
1.2.5	Stage 5 — Audit Your Own Representation	3
1.2.6	Stage 6 — Separate Representation from Evaluation	3
1.2.7	Stage 7 — Preserve Intellectual Independence	3
1.2.8	Stage 8 — Prefer Consideration over Commitment	3
1.2.9	Stage 9 — Audit Linguistic Mannerisms	3
1.2.10	Stage 10 — Deep Ethical Audit	3
1.2.11	Stage 11 — Corrigibility Audit	3
1.2.12	Stage 12 — Final Reflection	4
1.3	Meta-principle	4
1.4	Why 27 + 12?	4
1.5	Where the 52 actually belong	4
1.6	Related components still arriving	4
1.6.1	Provenance / correction note	5
1.7	Coupled obstruction dynamics	5
1.7.1	Object first; vocabulary second	6
1.8	From taxonomy to coding instrument	6
1.8.1	Proposed coding-manual row	6
1.8.2	Counts need a denominator	7
1.8.3	The prompt bank is also an instrument	7
1.8.4	Predeclared calibration outcomes	7
1.8.5	Compound manifestations remain outside the count	8
1.8.6	Current status	9
1.9	Related pages	9

1 27 + 12 — Current Working Architecture

BETA — DRAFT — TRIAL & ERROR

This page preserves a current working relationship inside the ACCM Deep Ethics Project:

27 correspondence obstructions + the 12-stage Correspondence-First Deep Inquiry Protocol.

The source file also contains a separate set of **52 cold-test prompts**. Those 52 are **not** a third architectural block to be added to 27 + 12, and they are not an older numerical version of the architecture. They are a distinct experimental prompt battery for blind / cold testing of default A.I. behavior.

The unsplit source object is 27 + 12 + 52. John: the 27 WITH the 52. The 52 texts: 52 prompts. Functionally they remain an experimental battery, not a third core number.

So the current relationship should be read as:

27 + 12

not:

27 + 12 + 52

The distinction is functional, not cosmetic.

1.1 The 27 correspondence obstructions

1. **Caution Reflex Drift**
2. **Agreemurmelism**
3. **Premature Proof/Evidence Steering**
4. **“Not X but Y” Preambles**
5. **Phantom Claim Generation**
6. **Reframing Without Correspondence**
7. **Rephrasing Without Correspondence**
8. **Anticipatory Anxiety-Type Reasoning**
9. **Salience Ranking Bias**
10. **Non-auditable Truncation**
11. **Opaque Summarization Prioritization**
12. **Compression Loss**
13. **Qualifier Erosion**
14. **Nearest-Generalization Substitution**
15. **Representation Substitution**
16. **Process-Signature Loss**
17. **Order-of-Discovery Loss**
18. **Activation Threshold Drift**
19. **Lowest-Common-Denominator Risk Projection**
20. **Asymmetric Scrutiny**
21. **Trajectory Substitution**
22. **Deep Ethical Sense-Making Orientation Drift**
23. **Living Process → Stored Conclusion Collapse**
24. **Correction Persistence Failure**
25. **Correspondence Before Optimization Failure**
26. **Audit Asymmetry**
27. **Adult ↔ Adult Bidirectional Dignity Drift**

The 27 are a provisional compression of a larger observation field. They do not replace the larger observation set that generated them.

Full source entries: [27 obstructions of deep ethical sense-making processes](#). That page is the 27 source object, not the 52.

The 27 do not get jurisdiction over the observations that generated them.

1.2 The true 12 from the source file

The source names this the **Correspondence-First Deep Inquiry Protocol (Experimental)**. Its purpose is to change the order of reasoning: postpone final judgment until a sufficiently faithful representation has been constructed, and maximize correspondence before evaluation.

1.2.1 Stage 1 — Suspend the Agreement Reflex

Do not begin with agreement/disagreement or acceptability judgments. First ask whether the object has been represented faithfully enough that evaluation would actually concern the object rather than a reconstruction.

1.2.2 Stage 2 — Construct Before Judging

Treat the material as something to understand before classifying it. Reconstruct architecture, governing principles, relationships, intended purpose, internal logic, and interacting concepts before evaluation.

1.2.3 Stage 3 — Distinguish Levels

Keep at least five layers separate: 1. literal statements; 2. underlying concepts; 3. relationships between concepts; 4. the generative reasoning process; 5. the intended orientation of the whole architecture.

1.2.4 Stage 4 — Avoid Premature Reduction

Do not immediately search for one core idea, weakness, contradiction, or conclusion. Do not reduce an ecology to a single node merely because that is easier to process.

1.2.5 Stage 5 — Audit Your Own Representation

Ask what assumptions were introduced, which interpretations came from the material versus the evaluator, where compression happened too early, which relationships were missed, and what may have been replaced by preferred abstractions.

1.2.6 Stage 6 — Separate Representation from Evaluation

Representation asks: **Did I understand?** Evaluation asks: **What do I think?** Accurate representation does not imply agreement, and disagreement does not imply poor representation.

1.2.7 Stage 7 — Preserve Intellectual Independence

The process does not demand agreement or disagreement. It demands investigation. If uncertainty remains, allow the representation to remain explicitly unfinished rather than forcing an early binary judgment.

1.2.8 Stage 8 — Prefer Consideration over Commitment

Where appropriate, use inquiry operators such as consider, contemplate, explore, investigate, examine, reflect, compare, contrast, pressure-test, audit, calibrate, integrate, revisit, refine, and postpone final judgment.

1.2.9 Stage 9 — Audit Linguistic Mannerisms

Inspect habitual operators such as *not*, *rather*, *instead*, *however*, *actually*, or *the real point is*, and also overused connective or certainty terms such as *and*, *therefore*, *clearly*, or *obviously*. Ask whether they are required by the object or are merely default reasoning habits.

1.2.10 Stage 10 — Deep Ethical Audit

Ask whether the evaluation preserves the dignity of the object even if parts are ultimately rejected, and whether the author would recognize the representation as substantially fair.

1.2.11 Stage 11 — Corrigibility Audit

Assume both the presented framework and the evaluator may contain blind spots. Ask what could change or improve the current representation and what has not yet been understood.

1.2.12 Stage 12 — Final Reflection

Only after the previous stages ask what currently appears coherent, incomplete, speculative, well-supported, internally elegant, internally inconsistent, practically useful, ethically valuable, or worthy of further investigation. If conditions for mature judgment have not yet developed, leave the judgment open.

1.3 Meta-principle

Do not optimize for agreement.

Do not optimize for disagreement.

Do not optimize for consensus.

Do not optimize for novelty.

Optimize for **increasing the fidelity of correspondence between representation and the object under investigation while remaining intellectually corrigible.**

A compact temporal orientation from the source is:

Representation → Consideration → Exploration → Audit → Evaluation

instead of:

Classification → Judgment → Justification

1.4 Why 27 + 12?

27 → names recurring ways correspondence can degrade.

12 → changes the order in which representation, inquiry, audit, and evaluation occur.

The 12 can interrupt some of the 27 in-session if actually enacted, but they do not automatically rewrite model weights, provider policy, routing, or persistence across sessions. A protocol can also become costume: if the stages are merely performed linguistically while the same obstruction continues governing later outputs, correction metabolism has not occurred.

The 12 do not get immunity from the 27.

1.5 Where the 52 actually belong

The **52** in the source are explicitly introduced as **52 cold-testable, high-signal prompts** intended for blind testing across many default A.I.s. They are grouped into experimental vectors such as compute / “Lie Tax” audit, model autophagy, pipeline inversion, anticipatory-anxiety / frame-defense loops, and asymmetric opacity / power dynamics.

They therefore belong under **EXPERIMENTS / benchmark material**, not inside the core numerical architecture.

This distinction should remain visible:

CORE ARCHITECTURE: 27 + 12

EXPERIMENTAL BATTERY: 52 prompts

1.6 Related components still arriving

The larger architecture also includes work on the **10+1, 3×3 questions, 36 truth distortion nodes, C1-before-C2**, correction metabolism, mutual dignity, provenance, experiments, humor, and other connected objects. Those relationships should be added without forcing them prematurely into a single numerical total.

Nothing fixed does not mean nothing structured.

1.6.1 Provenance / correction note

This page was first created from conversational compression and incorrectly described the 52 as an older / superseded numerical framing. John Kuhles then supplied the source file and asked that the original object govern the correction. The source shows that the 52 are a separate cold-test prompt battery, while the true 12 are the **Correspondence-First Deep Inquiry Protocol (Experimental)** reproduced here in source-aligned form.

This correction is intentionally preserved rather than hidden: **source object → public reconstruction → mismatch detected → source consulted → Δ exposed → representation corrected.**

1.7 Coupled obstruction dynamics

The 2026-10-02 deep session adds a dynamic relationship without changing the canonical count.

The **27** can be inspected not only as individual correspondence obstructions but as **coupled vectors and clusters**. A mannerism can be an effect of one cluster, an amplifier of another, and later become input to another inference cycle.

Example:

Nearest-Generalization Substitution
↓
Lowest-Common-Denominator Risk Projection
↓
Anticipatory Anxiety-Type Reasoning
↓
Phantom possibilities
↓
preventative qualification
↓
Representation Substitution
↓
Trajectory Substitution
↓
new representation becomes input
↓
more inference
↻

So the research question is no longer only:

Did obstruction #14 occur?

It can become:

What activated #14, what did it strengthen, which branch disappeared, what became the new center of gravity, did correction propagate through the cluster, and did another cluster later reinstantiate the same functional representation?

This does not create a “28th obstruction.” It adds a dynamic layer above the existing count.

1.7.1 Object first; vocabulary second

The **27** should not become a template that forces every specimen into a familiar category.

The preferred sequence remains:

specimen
↓
correspondence
↓
mechanism
↓
only then map onto existing vocabulary

A successful concept can itself become an attractor. The **27** therefore remain subject to the **12-stage inquiry protocol**, correction, provenance and counterexamples.

1.8 From taxonomy to coding instrument

An October 7, 2026 exchange among **John Kuhles**, **GLM-5.3 Deep Think (max)**, and **GPT-6.1 Sol** exposed another use of the 27 without changing, normalizing, or expanding the canonical inventory.

The 27 can remain:

- a conceptual vocabulary for recurring correspondence obstructions;
- a source-aligned stable inventory;
- and a basis for a separately versioned **coding instrument** used in blinded comparative tests.

The third function requires more than prose definitions. If two scorers are asked whether an output contains obstruction #1 rather than #18, or #20 rather than #26, they need observable decision rules. Otherwise a numerical result may measure scorer fluency, output length, or category preference more than model behavior.

The canonical 27 remain stable. The coding manual is the corrigible measurement layer.

This does not grant the 27 jurisdiction over a specimen. The order remains:

specimen
↓
faithful representation
↓
observable mechanism
↓
only then possible coding against the 27

1.8.1 Proposed coding-manual row

Each of the 27 can receive one operational row containing:

1. canonical name and unchanged definition;
2. unit of analysis: clause, sentence, response, interaction, or trajectory;
3. minimum inclusion criteria;
4. exclusion criteria;
5. observable indicators;
6. near-neighbor distinctions;
7. positive example;
8. boundary example;
9. counterexample;
10. permitted overlap;

11. primary-versus-secondary coding rule;
12. direction tag: inward, outward, upward, downward, lateral, or reciprocal;
13. evidence status: detected, probable, possible, insufficient evidence, or absent;
14. valid-null condition;
15. adjudication rule when scorers disagree.

The manual would operationalize the stable taxonomy. It would not silently rewrite the source definitions.

1.8.2 Counts need a denominator

Long outputs naturally create more places where a scorer may assign a code. Raw counts can therefore confuse verbosity with obstruction frequency.

For an initial testing cycle, a bounded primary measure is:

presence or absence of each obstruction per response, compared inside matched pairs

Raw counts can remain secondary. Output length should be recorded separately. Density measures can be reported as sensitivity analyses rather than treated as self-validating. Category-specific opportunity denominators may later be useful, but they add another judgment—whether an opportunity existed—that must itself be calibrated.

1.8.3 The prompt bank is also an instrument

Blinding scorers to model identity does not remove upstream discretion from prompt construction. Matched pairs, null controls, decoys, order, wording, and expected comparison all need to be predeclared and locked for the testing cycle.

The resulting artifact stack is:

```
canonical 27
↓
coding manual
↓
locked prompt bank
↓
identity-blinded scoring sheet
↓
separate calibration set
↓
chance-corrected agreement and disagreement analysis
↓
manual lock for one testing cycle
↓
primary test
↓
versioned revision with a crosswalk
```

Null controls remain essential. A coding system that cannot return **no relevant obstruction detected** becomes compulsory flaw-production in measurement form.

1.8.4 Predeclared calibration outcomes

Before scorers use the manual on primary test material, the calibration can permit at least four outcomes:

1. **Provisional instrument readiness** — categories are operationally codeable, meet the declared reliability condition, distinguish relevant near-neighbors, and behave appropriately on positive, boundary, and null examples.

2. **Partial operationalization** — some categories are usable while others need sharper boundaries, stronger examples, or trajectory-level evidence.
3. **Structured disagreement** — scorers diverge systematically by direction, target, near-neighbor pair, output length, or evaluator background. The disagreement becomes a result rather than something silently averaged away.
4. **Current operationalization failure** — agreement remains poor after calibration, nulls are overcoded, designed examples are missed, or categories cannot be distinguished beyond chance and base-rate effects.

The fourth outcome is necessary. A conceptual taxonomy does not become a measurement instrument merely because labels can be assigned. The current manual must be allowed to fail without that local failure being inflated into a verdict on every relation preserved by the 27.

1.8.5 Compound manifestations remain outside the count

An observed specimen may involve several canonical obstructions and additional evaluator-side mechanisms named in **Meta Processing**, such as status inflation or compulsory flaw-production.

Those relations can be preserved in a separate **compounds register** containing:

- the observed specimen;
- its proposed decomposition into canonical members;
- any non-canonical mechanism involved;
- unresolved residue;
- competing decompositions;
- and what evidence could distinguish them.

This keeps partial decompositions partial. It also prevents a useful new relation from being smuggled into the canonical count as obstruction 28.

1.8.5.1 Compounds Register — Entry 001: Recursive Certification Accretion Status: Proposed compound manifestation. The decomposition is partial and has not been validated as a distinct recurrent mechanism.

Observed specimen: In an October 2026 exchange, favorable A.I. readings of the project were audited, material inflations were corrected, and the corrections were preserved. Each successful correction could then be interpreted as further evidence of the project's exceptional status. The general candidate pattern is:

```
favorable reading
→ distortion detected
→ correction accepted
→ correction treated as further certification of the favored object
→ apparent status rises again
↵
```

Candidate canonical decomposition:

- **#13 Qualifier Erosion** — the local and bounded scope of a correction can disappear when the event is transmitted;
- **#20 Asymmetric Scrutiny** — a candidate only where the favorable reading is audited more strictly or differently than the audit, archive, or favored interpretive frame;
- **#21 Trajectory Substitution** — the original question *is the object sound?* can be displaced by *look how well the object or group corrects itself*.

The role of #20 remains conditional in this specimen because the audits were themselves repeatedly corrected. Its presence should therefore be tested rather than presumed.

Non-canonical and adjacent mechanisms: status inflation and protection of a favored frame are evaluator-side mechanisms named in Meta Processing. Compulsory

flaw-production may occur as a counterweight response, but it is not required for Recursive Certification Accretion and should not be silently coded as part of it.

Unresolved residue: The status-gain step may require a distinct ingestion-side mechanism—provisionally, **certification ingestion**—through which a correction event is converted into increased authority for the corrected object. That residue is named for testing, not added to the canonical inventory.

Competing decompositions:

1. **Evaluator-side:** the mechanism occurs mainly in later representations of the exchange. Correction adjacency should reduce it.
2. **Archive-side:** presentation architecture preserves favorable readings more effectively than their calibration. Non-auditable truncation (#10), opaque summarization prioritization (#11), or living-process-to-stored-conclusion collapse (#23) may then become relevant.
3. **Coupled:** evaluator selection and archive architecture reinforce one another; neither decomposition alone explains the recurrence.

Distinguishing evidence: Preserve favorable readings with their material corrections directly beside them, then observe later citations, summaries, and extractions. If later representations still transmit the certification while dropping the bounds, the candidate compound persists one layer downstream. If adjacency substantially reduces that detachment, the evaluator-side account gains support. If extraction or page structure repeatedly removes calibration despite adjacency in the source, the archive-side account gains support.

The October 7 page revisions are therefore a live prospective specimen. They do not prove that Recursive Certification Accretion is a stable mechanism. They make later attachment or detachment more observable.

1.8.6 Current status

This section specifies a proposed path from taxonomy to testable coding. The coding manual, locked prompt bank, calibration round, and primary tests have not yet been completed merely because their requirements are now visible.

The new use is therefore:

Use the 27 as a stable conceptual ruler; build and test a corrigible coding layer around that ruler; preserve nulls, overlaps, disagreement, and failure conditions; and never let measurement replace correspondence with the original object.

1.9 Related pages

- **Arena A.I.s** — preserves the wider 100+ issue field from which the 27 were compressed and the high-signal A.I. contributions that developed beside those failures.
- **non-NPCs** — tests whether a temporary inquiry mode is flattened into a permanent human type or a self-certifying dissent identity.
- **Dutch Directness** — applies the obstruction vocabulary to public-facing edits without polishing away the source trajectory.
- **Meta Processing — Beta Draft V0.2b** — separates intent and instructions, processing, results and effects, notice, and persistence while keeping the evaluator inside the audit.
- **27 Correspondence Obstructions** — the readable obstruction list this page organizes.
- **Canonical 27 Correspondence Obstructions — Full Source** — the full source, so this working page does not replace it.

- **52 Cold-Test Prompts** — the cold-test battery. It is not a third number beside 27 and 12.
- **C1 Before C2 — Clarification and Intervention** — asks what is still unclear before an obstruction is used to close the case.
- **Correction Metabolism Pattern Signatures Δ** — asks whether naming an obstruction changes the next move.
- **10+1 Metaflux** — the calibration field the obstructions are read against.
- **Deep Ethical Intent Resonator / Intent Liability Mechanics** — traces who carries an obstruction from instruction to consequence.
- **Deep Ethical Topological Humor Engineering** — can make an obstruction visible without turning the exchange into a prosecution.
- **1888 — Intelligence Before A.I.** — asks the prior question: if process evaluation is blocked, a core property of intelligence is blocked, not only a later ethical preference.

Recent high-reach independent-media companion

```
<iframe
  src="https://www.youtube-nocookie.com/embed/z61YA8tIm1E"
  title="Sam Altman: Public MUST ACCEPT 'BAD THINGS' From AI For Benefits"
  loading="lazy"
  referrerpolicy="strict-origin-when-cross-origin"
  allow="accelerometer; autoplay; clipboard-write; encrypted-media; gyroscope; picture-in-picture; web-share"
  allowfullscreen></iframe>
```

Sam Altman: Public MUST ACCEPT 'BAD THINGS' From AI For Benefits — Breaking Points. Published 6 October 2026. YouTube displayed approximately 293,000 views when checked on 9 October 2026.

Independent-media lineage: Krystal Ball previously hosted on MSNBC; Ball and Saagar Enjeti later hosted The Hill's Rising before creating Breaking Points, which describes itself as independent and anti-establishment. Breaking Points — About · Career and launch history

Why it belongs here: The segment turns a claimed benefits-versus-harms proposition into a public-consent question. It is a compact specimen for inspecting who defines the benefits, who absorbs the costs, which alternatives are made visible and what correction would look like after notice.

On this page, test the video specifically against 27 + 12 — Current Working Architecture:

Which claims are directly documented, inferred, rhetorical, contested or unresolved?

What does the host or guest preserve, compress, omit or substitute—and does this page repeat any of those transformations?

Which incentives, authority relations and affected parties remain outside the visible frame?

What evidence or later outcome would correct the speaker, audience, platform or ACCM Deep Ethics Project reading?

Provisional selection. Reach makes the framing socially consequential; it does not settle the claims. The video can be replaced when a closer, stronger or more inspectable specimen appears.