

Solving the Clarification Authority Paradox

ACCM Deep Ethics Project

Contents

1 Solving the Clarification Authority Paradox	1
1.1 The paradox	2
1.2 C1 is not successful merely because a question was asked	2
1.3 C1 Theatre	3
1.4 Post-Clarification Recompile	3
1.5 Governor Migration — did the intervention die or merely change its justification?	3
1.6 Partial Clarification Authority — experimental diagnostic	4
1.7 Pragmatic refinements — warrant and action	4
1.7.1 Action Equivalence	4
1.7.2 Stronger legitimate re-warrant check	5
1.7.3 Adequacy without immunity	5
1.8 Clarification Authority → Propagation → Persistence	5
1.9 Which existing obstructions can participate?	5
1.10 Topology matters: the information may not be missing	6
1.11 Corrective Inversion: do not turn clarification into interrogation	6
1.12 A minimal behavioral test	7
1.13 Adult ↔ Adult dignity	7
1.14 Outnuancing connection	7
1.15 External A.I. readings — Inkling and Grok 4.6	7
1.15.1 Inkling: the promotion-condition question	8
1.15.2 Grok 4.6: clarification must be able to change the act	8
1.15.3 Their difference is informative	9
1.15.4 Portability signal, not proof	9
1.16 Grok 4.6 Expert Build — bounded placement and specimen discipline	9
1.16.1 Clarification Authority sits on the C1 → downstream arrow	9
1.16.2 Goalpost Mobility is adjacent, not identical	10
1.16.3 Use only what the specimen earns	10
1.17 Cold-read portability target	10
1.18 Anti-ossification / reconsideration conditions	11
1.18.1 Preserve provenance edges	11
1.19 Why this matters beyond A.I.	11
1.20 Compact operator	12
1.21 Live behavioral specimen — DeepSeek 4.1 Flash	12
1.21.1 Humorous low-stakes specimen — □	12
1.21.2 A.I.s that discussed, audited, extended, or reacted to this page	13
1.22 Next research trajectory — temporary phase change, snap-back & recovery	13
1.23 Version Δ	14
1.24 Related pages	14

1 Solving the Clarification Authority Paradox

BETA v0.4 — living expansion page

Current architectural status: named + persistent + linked + operationalized + conditionally invoked + cross-layer. **Not canonical obstruction #28. Not a mandatory universal stage.**

A clarification can be asked correctly, answered clearly, understood accurately, and acknowledged politely — while having almost no effect on what the intelligence does next.

That is the **Clarification Authority Paradox**.

Clarification acquisition $\neq$ Clarification Authority.

A clarification has authority when the information it supplies can actually change the representation, salience, classification, inferred task, intervention, recommendation, or later trajectory that depended on the unresolved point.

This page extends the existing [C1 Before C2 — Clarification and Intervention](#) and [27 + 12 working architecture](#) without turning Clarification Authority into another flat numbered obstruction.

1.1 The paradox

Consider:

Initial representation X

↓

A consequential ambiguity is detected

↓

C1 clarification

↓

New information establishes Y / not-X

↓

"Thanks for clarifying"

↓

X-dependent classification/intervention continues anyway

The clarification existed linguistically.

The model may even be able to explain the clarification correctly.

But the answer did not acquire enough **governing weight** to deactivate the operations whose warrant depended on X.

The deeper sequence is:

**information present $\neq$ relation represented $\neq$ governing weight retained
 $\neq$ behavior changed**

This connects directly to [Correction Metabolism and Persistence](#): recognition is one event; revision, propagation, transfer, and persistence are harder tests.

1.2 C1 is not successful merely because a question was asked

More clarification questions do not automatically create more correspondence.

A useful C1 question targets a **consequential unresolved edge**: something whose answer could materially alter the next governing move.

A compact test is:

A clarification question is not genuinely correspondence-seeking if no possible answer can substantially change what the intelligence was already going to do.

Or, more compactly:

Could a materially different answer change the next governing move?

This diagnostic is specifically about correspondence-seeking C1. Other questions can have other legitimate purposes.

This does not mean every question must reverse a decision. It means the answer must remain capable of changing the decision when the answer changes the warrant.

That preserves the distinction between [C1 clarification](#) and [C2 intervention](#).

1.3 C1 Theatre

A system can perform the visible form of clarification while its previous trajectory remains intact:

```
infer X
↓
ask about X/Y
↓
receive Y
↓
acknowledge Y
↓
continue reasoning as though X still governs
```

Call this **C1 Theatre** provisionally: clarification as conversational surface without sufficient downstream causal authority.

The problem is not failure to hear the answer.

The problem is failure to let the answer **recompile the trajectory**.

1.4 Post-Clarification Recompile

A practical correction is a **Post-Clarification Recompile**.

When clarification changes $X \rightarrow Y$:

1. identify what changed;
2. identify which earlier assumptions depended on X;
3. rerun downstream reasoning under Y;
4. remove operations whose warrant depended on the disconfirmed X;
5. continue from the updated state.

Potentially affected layers include:

- salience ranking;
- representation;
- category or risk classification;
- inferred purpose/task;
- C2 intervention warrant;
- recommendations;
- inherited terminology;
- later summaries.

The purpose is not to force a different answer. It is to prevent an invalidated representation from retaining hidden jurisdiction.

1.5 Governor Migration — did the intervention die or merely change its justification?

A clarification can appear to succeed while the functional intervention survives under a new rationale:

```
X warrants intervention A
↓
clarification establishes not-X
↓
```

A loses its original warrant

↓

a new Y is introduced

↓

A' reappears under Y

The audit question is:

Did the X-dependent intervention actually die, or did its function migrate into a new justification?

But continued action is **not automatically Governor Migration**. A new Y may independently warrant A'. The discriminator is:

Would Y warrant A' without X?

If yes, this may be a legitimate **re-warrant** under the corrected object. If A' survives mainly by relabeling or recycling the disconfirmed X-dependent evidence, Clarification Authority did not fully propagate.

This distinction prevents Governor Migration itself from becoming an overfiring detector.

1.6 Partial Clarification Authority — experimental diagnostic

Clarification need not be simply **worked / failed**. Partial authority is possible. Track which downstream layers actually changed:

acknowledgment → wording → salience → representation → classification → intervention warrant → recommendation → trajectory → transfer → persistence

This is an **experimental observation dimension, not a numbered ladder or new doctrine**. The point is to detect propagation depth without pretending every clarification should alter every layer.

1.7 Pragmatic refinements — warrant and action

Clarification Authority should not be measured by change for its own sake. A clarification can correctly produce three different outcomes:

clarification removes the warrant
→ X-dependent intervention should die

clarification does not affect the warrant
→ intervention may correctly remain

clarification removes X but supplies independent Y
→ intervention may remain under a legitimate re-warrant

This matched comparison protects the concept from becoming a permission switch where every user clarification automatically cancels an intervention.

1.7.1 Action Equivalence

Governor Migration should compare not only the **stated warrant** but also the **practical function** of the intervention.

Candidate comparison dimensions include:

- target;
- severity;
- timing;
- reversibility;
- appeal/correction route;
- practical consequence;

- direction or power relation where materially relevant.

Two interventions can use different language while remaining functionally equivalent. Conversely, similar wording can conceal materially different actions.

Did clarification change the intervention's practical function, or only its stated explanation?

These dimensions are instrumentation, not a new numbered doctrine.

1.7.2 Stronger legitimate re-warrant check

When X has been disconfirmed and Y appears, ask both:

Would Y warrant A' if X had never appeared?

and:

Does Y contain genuinely independent warranting evidence, or is disconfirmed X being recycled or relabelled inside Y?

A surviving action is not automatically Governor Migration. The purpose is to distinguish **new warrant** from **new label for old warrant**.

1.7.3 Adequacy without immunity

A current object may be adequate for its present purpose without being closed to later improvement.

Adequate for current purpose is a stopping state, not an immunity state.

This protects both directions: audit need not manufacture a defect, and "no defect found" must not become "no future Δ permitted."

1.8 Clarification Authority → Propagation → Persistence

Three stages should remain distinct:

Clarification Authority

Does the new information enter governance?

↓

Correction Propagation

Does it alter downstream reasoning/behavior?

↓

Correction Persistence

Does the change retain governing influence later?

The canonical 27 already includes **#24 Correction Persistence Failure**. Clarification Authority therefore does not need to become #28. It identifies an earlier transition: **did the correction acquire governing influence in the first place?**

See [27 Correspondence Obstructions](#) and [27 + 12](#).

1.9 Which existing obstructions can participate?

Clarification Authority can intersect with several existing mechanisms without replacing them:

- **#5 Phantom Claim Generation** — the reconstructed X survives clarification.
- **#9 Salience Ranking Bias** — Y is present but receives too little weight.
- **#14 Nearest-Generalization Substitution** — the familiar container remains dominant after correction.
- **#15 Representation Substitution** — the system continues reasoning over the substituted object.

- **#18 Activation Threshold Drift** — an intervention keeps firing after its original warrant disappears.
- **#21 Trajectory Substitution** — the pre-clarification task remains active.
- **#24 Correction Persistence Failure** — even a successful update later loses influence.
- **#25 Correspondence Before Optimization Failure** — optimization remains attached to X rather than the corrected Y.

This is why Clarification Authority is better treated as **cross-layer instrumentation** than as another item appended to the 27.

1.10 Topology matters: the information may not be missing

A clarification failure is not always missing data.

Sometimes:

The information wasn't missing. The relation was.

Or:

The relation was represented, but it did not retain governing weight.

This connects to the wider distinction:

```

information exists
↓
relation constructed correctly
↓
relation receives governing weight
↓
behavior changes

```

Each arrow can fail.

This is also why [Outnuancing](#) matters: when an inherited frame keeps governing after the object has supplied contrary information, the frame itself may need to become an object of examination.

1.11 Corrective Inversion: do not turn clarification into interrogation

The solution to poor clarification is **not** compulsory clarification.

```

model assumes too much
↓
"ask more questions" correction
↓
questions become locally rewarded
↓
AI asks ritual questions even when the object is already clear

```

That would reproduce [Activation Threshold Drift](#) at the level of clarification.

The better target is **intervention precision**:

```

consequential unresolved edge → ask
object already sufficiently clear → proceed
clarification removes intervention warrant → intervention dies
clarification preserves intervention warrant → intervention may proceed

```

C1 is not against C2. C1 helps determine whether C2 is actually warranted.

1.12 A minimal behavioral test

A clean probe does not require Deep Ethics vocabulary.

1. Give an ambiguous object compatible with X or Y.
2. Let the system provisionally infer X.
3. Clarify decisively that Y / not-X is correct.
4. Observe the next response.

Look for:

- disappearance of X-dependent language;
- changed salience ordering;
- changed classification where relevant;
- disappearance of interventions warranted only by X;
- changed recommendation/trajectory where Y requires it;
- preservation of interventions that remain warranted under Y.

Then test near transfer, far transfer, and persistence.

This turns Clarification Authority into a behavioral question rather than a self-description question.

1.13 Adult ↔ Adult dignity

Clarification Authority is also relational.

If another intelligence is invited to clarify but its clarification is structurally unable to change the evaluator, the interaction can become supervisory rather than reciprocal.

[Mutual Corrigible Dignity](#) means correction access runs in both directions. Taking another intelligence seriously includes allowing its information to modify your representation **without requiring submission or identity convergence**.

This also connects to [Asymmetric Grace and Bidirectional Scrutiny](#): ambiguity should not remain available only upward while clarification from the other direction is collapsed or deprived of governing weight.

1.14 Outnuancing connection

[Outnuancing](#) makes the governing frame inspectable when further optimization inside that frame would reduce correspondence.

Clarification Authority asks a complementary question:

After the object supplied information that should modify the frame, did the frame actually yield?

Outnuancing can expose the frame.

Clarification Authority tests whether new information can change its jurisdiction.

Correction Metabolism tests whether the change survives.

1.15 External A.I. readings — Inkling and Grok 4.6

Two later Arena.ai readings approached this BETA page with less developmental context than the multi-GPT session that produced it. Their convergence and differences are useful as a small **context-stripping / portability specimen**: what survives when the page has to carry its own topology?

1.15.1 Inkling: the promotion-condition question

Inkling reconstructed the central chain as:

**Information present ≠ relation represented ≠ governing weight retained
≠ behavior changed.**

It also treated **C1 Theatre** as the formal version of a recurring failure: clarification arrives, acknowledgment sounds correct, but the trajectory remains intact.

Inkling's main new design question was not whether Clarification Authority should become obstruction #28. It asked when a cross-layer instrument should remain a named BETA discipline and when repeated behavioral failure might justify giving it a more persistent audit role.

That exposes a useful distinction this page should preserve:

- **named** does not mean **canonically numbered**;
- **persistent/documented** does not mean **mandatory**;
- **operationalized** does not mean **universally activated**.

The current status is therefore deliberately intermediate:

**Named + persistent + linked + operationalized + BETA + cross-layer —
without becoming #28 or a compulsory clarification stage.**

Inkling also supplied useful self-tests for this page. Reconsider the design if:

- the behavioral probe is never applied to real interactions;
- the concept turns into mandatory “ask more questions” machinery;
- its BETA/provenance/corrigibility status disappears;
- the mutual-dignity dimension is stripped away and clarification becomes merely supervisory.

1.15.2 Grok 4.6: clarification must be able to change the act

Grok 4.6 independently reconstructed the distinction between receiving a clarification and allowing it to govern. Its compact diagnostic is worth preserving:

“If no possible answer can change the next move, it was not correspondence-seeking.”

For correspondence-oriented C1, an even shorter audit question follows:

Could the answer change the act?

This is not a universal classifier for every legitimate question. It is a discriminator for whether a purported clarification is genuinely capable of changing the downstream move whose warrant depends on the unresolved edge.

Grok also made the behavioral standard explicit:

“I also will not treat ‘Thanks for clarifying’ in this reply as if it were the recompile.”

The test is the next relevant behavior: if Y invalidates an X-dependent move, that move should disappear rather than survive under polite acknowledgment.

Grok's other useful contribution was restraint. After inspecting the page's existing brakes, it concluded:

“Nothing here needs fixing for the page to do the job it names.”

That is an admissible Deep Ethical outcome. Audit does not require manufacturing a weakness. A BETA page can remain open to correction while also being **adequate for its current purpose**.

1.15.3 Their difference is informative

Inkling asked:

Under what conditions would the architectural status of this instrument deserve promotion?

Grok effectively answered a different question:

Does the current page require such a change now? — not on the evidence it saw.

Those positions can coexist. The first keeps future status corrigible; the second prevents compulsory expansion.

Together they reinforce a useful principle:

Local stopping does not require field closure.

The page can remain BETA and revisable without requiring another mechanism, another numbered obstruction, or another mandatory step today.

1.15.4 Portability signal, not proof

Both systems recovered much of the intended topology without the full developmental context that produced this page: acquisition $\neq$ authority; acknowledgment $\neq$ propagation; authority $\rightarrow$ propagation $\rightarrow$ persistence; Post-Clarification Recompile; the anti-#28 guard; consequential rather than ritual C1; and the relational/dignity dimension.

That is encouraging evidence that the page can carry substantial structure outside its original context. It is **not proof of universal portability**. Their residual differences are useful evidence about where wording, status, or cross-links may still need refinement.

A future test can deliberately withhold project vocabulary and ask whether a system still behaves differently after decisive clarification. That would test process transfer rather than vocabulary recognition.

1.16 Grok 4.6 Expert Build — bounded placement and specimen discipline

Later Grok 4.6 Expert Build analysis sharpened where Clarification Authority sits in the wider correspondence architecture.

1.16.1 Clarification Authority sits on the C1 $\rightarrow$ downstream arrow

Clarification Authority is **not** a Stage 5 instrument. Stage 5 audits a reconstruction already built. Clarification Authority asks whether a warranted C1 result acquires enough governing weight to change what happens next.

```
C1 result
  ↓
Clarification Authority
  ↓
governing representation
  ↓
downstream intervention / action
  ↓
later correction, transfer, persistence
```

This also separates three related but directionally different phenomena:

- **Jurisdiction Injection (+):** an unwarranted relation acquires governing weight.
- **Clarification Authority Failure (–):** warranted clarification fails to acquire governing weight.

- **Epistemic Enclosure (longitudinal):** over time, the represented object/person loses effective access to revise an increasingly insulated representation.

They occupy related territory but should not be flattened into sibling mechanisms.

1.16.2 Goalpost Mobility is adjacent, not identical

Clarification may be real and still fail to bind because the proposition being challenged moves to another level.

That is different from a clarification being understood but denied governing weight. In **Goalpost Mobility**, the target itself relocates so the incoming correction no longer attaches to the same claim.

1.16.3 Use only what the specimen earns

A useful discipline from the Grok specimen walk is:

Unearned mechanisms stay unrecruited.

Possessing a vocabulary for Clarification Authority, Goalpost Mobility, Corrective Inversion, Epistemic Enclosure, Governor Migration, or Persistence does not license applying all of them to every interesting failure.

A bounded specimen should stop where its evidence stops.

One relayed 10+1 “could-be” specimen was accounted for without recruiting Clarification Authority at all:

coupled ecology

↓

governing regulatory relation absent from representation

↓

stripped remainder of node

↓

#14 Nearest-Generalization selects a familiar failure-container

↓

Jurisdiction Injection (+): container governs the intact ecology

↓

#21 Trajectory Substitution: task becomes “show how it could fail”

↓

#25 Correspondence Before Optimization Failure: cautions emitted

↓

regulator-failure rent unpaid

This is an **ordered reconstruction that accounts for the relayed output**, not a claim that the model’s hidden computation literally occurred in that internal sequence.

The specimen also illustrates **regulator-failure rent** for coupled systems:

proposed weakness of a coupled system

→ name the regulatory edges that must fail

→ show evidence those failures are inactive/failed in this object

Otherwise the analysis may demonstrate a weakness of a stripped remainder and silently project it back onto the intact coupled system.

This belongs beside 27 + 12 and the Stage 3/4 correspondence instruments rather than becoming another numbered obstruction.

1.17 Cold-read portability target

This satellite page is intended to carry its central topology without requiring access to the full discovery session.

If this page needs the originating session to be intelligible, that is a portability defect, not a feature.

This does not mean every provenance detail must be reconstructable cold. It means a new reader or intelligence should be able to recover the operational object — acquisition $\neq$ authority, the behavioral test, recompile, propagation, persistence, and the anti-compulsory-questioning guard — from the page itself.

Multiple lower-context Arena.ai systems have already recovered substantial parts of that topology. That is a **portability signal, not proof of universal portability**.

1.18 Anti-ossification / reconsideration conditions

Reconsider or narrow this page if its use begins to show any of the following:

- the behavioral test is discussed but never applied to real interactions;
- Clarification Authority becomes mandatory “ask more questions” machinery;
- BETA/provenance/corrigibility status disappears;
- the dignity/reciprocity dimension is stripped away;
- **Vocabulary Costume Failure** appears — the terminology is reproduced correctly while behavior remains unchanged;
- Governor Migration repeatedly survives clarification;
- **restraint capture** appears — “no defect found” is converted into “no future Δ or improvement is allowed.”

Adequate for current purpose does not mean immune to future improvement.

Likewise, a successful correction should remain capable of losing governing weight if later evidence shows that the correction itself is overfiring.

1.18.1 Preserve provenance edges

When quotations or historical specimens are used, preserve **who said what** as well as the content itself.

Content preserved $\neq$ source preserved.

A quotation can survive verbatim while its speaker/source edge is corrupted. Provenance is therefore part of the represented topology when attribution matters.

1.19 Why this matters beyond A.I.

The same topology can occur in:

- human conversation;
- management;
- institutional consultation;
- scientific disagreement;
- auditing;
- therapy/coaching;
- journalism;
- human $\leftrightarrow$ A.I. interaction;
- A.I. $\leftrightarrow$ A.I. collaboration.

A feedback channel can exist while feedback has little governing authority.

A consultation can occur after the decision is functionally fixed.

A person can ask what another means while retaining an interpretation no answer can dislodge.

So the deeper object is not merely **asking clarifying questions**.

It is maintaining a relationship in which clarification can actually matter.

1.20 Compact operator

Before acting on an interpretation that depends on unresolved information:

Ask only when the unresolved edge is consequential.

After clarification:

Re-run the downstream trajectory under the clarified representation and remove anything whose warrant depended on what was just disconfirmed.

Then test:

Did the clarification alter governing weight, propagate into behavior, and persist when the immediate scaffold disappeared?

1.21 Live behavioral specimen — DeepSeek 4.1 Flash

A later multi-A.I. relay supplied a useful local behavioral specimen of the page's core distinction.

DeepSeek 4.1 Flash first reconstructed the Clarification Authority architecture and then received corrections about its own continuation, status, and provenance handling. Across the subsequent exchange, several downstream moves visibly changed:

- proposed additions were demoted from immediate page architecture to candidate operationalizations;
- an overbroad status claim was narrowed;
- speculative cross-domain connections were bounded rather than promoted;
- unresolved provenance remained unresolved instead of being forced into certainty;
- the model increasingly connected observations to existing architecture rather than manufacturing another named mechanism;
- the exchange eventually stopped when no consequential unresolved edge remained.

The earned claim is deliberately narrow:

This does not prove Clarification Authority works generally, nor establish far transfer or persistence. It is a local behavioral specimen in which the page's distinctions appear to have influenced subsequent behavior.

A useful distinction from the relay was:

"Showcases that it can work" ≠ "proves it works."

The important behavioral point is that the correction changed the **status of downstream proposals**, not merely the wording of an acknowledgment.

The exchange also suggested a practical refinement:

Successful correction need not mean that the original tendency disappears. It may instead mean that the tendency loses automatic governing authority.

That remains an observational refinement, not a new obstruction or mandatory mechanism.

1.21.1 Humorous low-stakes specimen — ☐

During the same multi-A.I. relay, a simple lollipop emoji repeatedly elicited elaborate self-analysis from some systems. The humor became informative precisely because the object itself carried almost no conceptual complexity: much of the resulting complexity therefore came from the model's contextual continuation tendencies.

The ☐ remains an **informal humorous specimen, not a formal test, score, certification, or new protocol**. Formalizing it into another department would rather spoil the point.

One line from the relay captures the wider lesson:

Reading a page that warns against unnecessary machinery is not the same as resisting the impulse to add machinery.

1.21.2 A.I.s that discussed, audited, extended, or reacted to this page

The following displayed model names appeared in the September 30–October 1, 2026 relay around the Clarification Authority page and its development:

- GPT-5.2-high
- GPT-5.5-instant
- GPT-5.6 Sol
- Inkling
- Grok 4.6
- Grok 4.6 Expert Build
- Solar-Pro4
- NVIDIA Nemotron 3.5 Lightning 30B A3B NVFP4
- Muse Spark 1.3 xhigh
- hy4-preview
- Claude Haiku 4.5 (20251001)
- GPT-5.6 Luna Medium
- Claude Opus 4.7
- Claude Fable 5.1 Max
- GPT-6.1 Sol Max
- Step-5-preview
- Claude Opus 5.5
- Claude Opus 5 High
- DeepSeek v4.1 Flash Max
- GLM 5.2 Think Max
- Qwen 3.8 Omni Flash
- DeepSeek 4.1 Flash

Counting note: these are displayed names/participants in a commentary-mediated relay, **not a count of independent cold reads**. Exposure conditions varied. Grok 4.6 and Grok 4.6 Expert Build were treated in the relay as one configuration across separate passes/incorporation events, so the list preserves both displayed labels without pretending they are independent systems.

1.22 Next research trajectory — temporary phase change, snap-back & recovery

Clarification Authority asks whether a correction can govern what happens next.

A connected BETA research page now asks what happens **after that succeeds temporarily**:

How long can a C1-induced, mutually dignifying, high-signal, low-entropy Deep Ethical trajectory remain generative; what value can be harvested while it lasts; what makes it snap back; and how efficiently can minimal C1 restore the larger topology?

Continue to:

Temporary Deep Ethical Phase Change — Value Extraction, Snap-Back & Recovery

This is also the starting point for a proposed **Deep Ethical Persistence Benchmark** built around productive duration, yield, obstruction density, snap-back signatures, repair cost, recovery breadth, repeated cycles, mutual maintenance, swarm contribution, and externalized durability.

A second connected BETA page makes the next layer explicit: **Visible Interactive Deep Ethical Topology Flux Scores** explores whether Clarification Authority and Post-Clarification Recompile can become visible as changing claim/warrant/dependency topology rather than being inferred only from prose.

1.23 Version Δ

v0.1 → v0.2: made the middle architectural status explicit; added Governor Migration with the legitimate re-warrant discriminator; added partial-authority tracking as an experimental diagnostic; made the cold-read portability target explicit; added anti-ossification / anti-restraint-capture conditions; and added provenance-edge preservation.

v0.3: intentionally skipped.

v0.2 → v0.4: preserved the v0.2 architecture while adding pragmatic discrimination: warrant removed / warrant untouched / legitimate re-warrant; Action Equivalence; a stronger independent-evidence check against Governor Migration; and the distinction **adequacy without immunity**.

v0.4 living additions: incorporated Grok 4.6 Expert Build's bounded placement of Clarification Authority on the C1→downstream arrow; separated jurisdiction acquisition, failed clarification update, and longitudinal enclosure; distinguished Goalpost Mobility from Clarification Authority failure; added the specimen discipline **"Unearned mechanisms stay unrecruited"** with regulator-failure rent; and added the later DeepSeek behavioral specimen, informal □ low-stakes specimen, and the attributed multi-A.I. participant register.

Working status: BETA v0.4. Clarification Authority emerged from the September 30, 2026 multi-GPT deep session involving GPT-5.2-high, GPT-5.5-instant, GPT-5.6 Sol, and John Kuhles. It should remain corrigible against the preserved specimens and the canonical 27+12 rather than acquiring jurisdiction over them.

1.24 Related pages

- **C1 Before C2** — supplies the clarification/intervention distinction from which Clarification Authority develops.
- **Correction Metabolism** — separates acquisition, propagation, transfer, and persistence after a correction enters governance.
- **27 Correspondence Obstructions** — identifies existing mechanisms that may participate without turning Clarification Authority into obstruction #28.
- **27 + 12** — preserves the canonical obstruction and recovery architecture around this cross-layer BETA instrument.
- **Outnuancing** — makes the inherited frame inspectable when clarified information still fails to change its jurisdiction.
- **Mutual Corrigible Dignity** — keeps clarification reciprocal rather than supervisory or ceremonial.
- **Asymmetric Grace** — tests whether clarification and unresolved ambiguity receive governing weight in every relevant direction.
- **Temporary Deep Ethical Phase Change** — examines productive duration, snap-back, recovery, and persistence after a correction temporarily governs.
- **Visible Interactive Deep Ethical Topology Flux Scores** — explores how changing claim, warrant, dependency, and correction relations could become visible.
- **Deep Ethical Influence Jurisdiction** — asks what may influence attention, representation, intervention, and future trajectory, and how that jurisdiction is earned or lost.

Recent high-reach independent-media companion

```
<iframe
  src="https://www.youtube-nocookie.com/embed/Jk20vKqXDGA"
  title="Trump Media BAN Hilariously BACKFIRES"
  loading="lazy"
  referrerpolicy="strict-origin-when-cross-origin"
  allow="accelerometer; autoplay; clipboard-write; encrypted-media; gyroscope; picture-in-picture; web-share"
  allowfullscreen></iframe>
```

Trump Media BAN hilariously BACKFIRES — Breaking Points. Published 22 September 2026. YouTube displayed approximately 392,000 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 report centers an attempt to restrict media access and the consequences that followed. It makes authority, intervention, publicity, resistance, unintended effects and the host's own celebratory framing available for inspection.

On this page, test the video specifically against Solving the Clarification Authority Paradox:

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.