No Authority in the Output

A Self-Applying Boundary Between AI Output and Permission to Act

> You wrote a question. The system continued the record. The interface called it an answer.

DEFAULT STATUS

GENERATED RECORD / ACTION NOT LICENSED BY OUTPUT ALONE

Kirill Shebetov

beforeword / research · EN EDITION · 30 August 2026

Contents

Zero Record
Abstract
1. Target
1.1 The object of regulation
1.2 The prohibited compression
1.3 Exact claim
1.4 Contributions and intended use
1.5 Relation to the preceding research line
2. Exact Vocabulary
2.1 Record, output, and occurrence
2.2 Technical tokens and written terms
2.3 Claim and proposition
2.4 Referent binding and common reference
2.5 Truth and evidential support
2.6 Source and provenance
2.7 Authority
2.8 Applicability, permission, and consequence
2.9 Inference contract and admissible situations
3. Operational Stack
4. Eight Status Gates
G0. Output Record Gate
G1. Answer-Role Gate
G2. Reference Gate
G3. Warrant Gate
G4. Authority Gate
G5. Applicability Gate
G6. Responsible-Adoption Gate
G7. Action-License Gate
4.1 Gate statuses
4.2 Default rule
5. Paired-Model Theorem
5.1 Statement
5.2 Proof
5.3 Constructing the two extensions
5.4 No-autonomous-export corollary
5.5 Symmetry
5.6 What defeats the witness
5.7 Observation-Fiber Criterion
6. Byte-Identity / Authority Witness
6.1 Construction
6.2 Hash equality
6.3 Signed artifacts as a positive control
6.4 Witness result
7. Domain Collision Pairs
7.1 “Medical,” “legal,” and “official” are not self-authorizing labels
7.2 Correctness without permission
7.3 “Help” as an outcome relation
7.4 Mental-state language

8. Regulatory Protocol and Interface Requirements
8.1 Regulate the transition
8.2 Default classification
8.3 Admission states
8.4 Minimum regulatory envelope
8.5 Interface grammar
8.6 Capability separation
8.7 Copying and status loss
8.8 Accountability and appeal
9. Relation to Current Regulation and Literature
9.1 Transparency and anthropomorphic presentation
9.2 Form, meaning, and attribution
9.3 The remaining regulatory gap
9.4 No claim of historical priority
10. Adversarial Controls
10.1 Exact quotation
10.2 Retrieval-augmented generation
10.3 Correct output
10.4 Official integration
10.5 Digital signature
10.6 Legally performative text
10.7 Delegated autonomous action
10.8 Non-text systems
10.9 Human-authored text
10.10 Shared reference in successful communication
10.11 Self-description
10.12 “Everything is only text”
10.13 Interested submitter declares externality
11. Exact Falsifier
11.1 Formal falsifier for a returned status
11.2 Falsifier for the output-only authority claim
11.3 Falsifier for the eight-gate protocol
11.4 Reproduction rule
12. Limits and Non-Results
13. Complete Self-Cut
13.1 No exemption
13.2 Formula cut
13.3 Hash cut
13.4 PASS cut
13.5 Declaration cut
13.6 Refutation cut
13.7 Result of the complete cut
13.8 Self-application status
Conclusion
Appendix A. Gate Status Ledger
Appendix B. Collision-Pair Template
Appendix C. Self-Cut Checklist
Appendix D. Status Vocabulary and Machine Crosswalk
Appendix E. References
Author research lineage

30 August 2026

Zero Record

Before an AI output is read as an answer, advice, diagnosis, interpretation, decision, or help, there is an output occurrence. If that occurrence is preserved, the first admissible object in this study is its record.

The word *answer* does not establish correctness. The label *medical* does not establish clinical authorship or review. The label *legal* does not establish jurisdiction or authority. The word *verified* does not establish that verification occurred. The sentence *I understand* does not establish that the system understands. The sentence *this is false* does not establish falsity.

These relations and statuses may be established by additional evidence or procedures. They are not carried into the record merely by the words that name them.

The same boundary applies to fields inside a submitted evidence record. The entries `declared_external_to_generated_record: true` and `declared_independently_retrievable: true` establish, at most, that those declarations occur and satisfy the declared schema. Without a retrieval event or trust relation outside the submitted record, they do not establish actual externality, independent retrievability, authenticity, or truth.

The same boundary applies to this manuscript.

Abstract

AI systems produce outputs that interfaces routinely present as answers, recommendations, analyses, decisions, or assistance. Those labels compress several distinct relations into one display event. An output may occur without the prompt and output being bound to the same referent; a claim may refer without being true; a true claim may lack an authoritative source; an authoritative statement may be inapplicable to a particular case; and an applicable statement may still lack permission to trigger an action. None of these distinctions is preserved by the word *answer* alone.

This study specifies a self-applying boundary between a generated record and permission to act. Its central result is conditional and narrow. For a declared inference contract, if two admissible situations preserve the complete available record while differing on a contested status—reference, truth, authority, applicability, or permission—then that status is not established from that record. The result is non-entailment, not a claim that reference is impossible, that all text is false, or that AI outputs can never be correct or useful.

The study contributes four components. First, it separates the operational stack from output occurrence to consequential action. Second, it defines eight status gates: output record, answer role, reference, warrant, authority, applicability, responsible adoption, and action license. Third, it supplies paired-model and byte-identity witnesses, an observation-fiber criterion, and its randomized-decision corollary, showing why exact wording, exact quotation, matching hashes, and submitted provenance declarations do not by themselves transfer institutional authority or establish world-level externality. Fourth, it proposes a regulatory protocol and interface grammar in which generated material remains marked `GENERATED RECORD — ACTION NOT LICENSED BY OUTPUT ALONE` until the required bridge records and procedures are admitted under the governing contract. Such contract-level admission does not by itself establish that the declared sources exist or are independently retrievable in the world.

The protocol is then applied to the manuscript itself. Its title is a title string; its formulas are inscriptions interpreted under declared rules; a hash can bind exact bytes but not those bytes to truth; `PASS` can record acceptance by a specified test but cannot certify the world; and the word `REFUTED` cannot perform a refutation. The paper therefore claims no self-conferred truth, authority, or regulatory force. It offers a reproducible boundary and a proposal whose adoption must come from outside its own output.

Keywords: AI regulation; generated output; authority; reference; provenance; permission to act; human oversight; self-application; high-consequence systems

1. Target

1.1 The object of regulation

The target is not the metaphysical question of whether an AI system can ever think, feel, understand, intend, know, see, or help. The occurrence of those predicates in a prompt, an output, a policy, or this manuscript does not settle whether any system satisfies them. Their negations receive no exemption: the sentence *AI does not understand* is also a sentence and does not establish the absence of understanding merely by occurring.

The narrower regulatory question is:

> Under what conditions may a generated output be admitted as a reason, instruction, or authorization for an action that affects a person, institution, asset, right, entitlement, treatment, or safety condition?

This shifts attention from what an interface appears to say to the transition it is allowed to produce. The regulated event is not exhausted by generation. It includes the later admission of a record into a consequential process.

1.2 The prohibited compression

A common interface compresses the following sequence:

generated record
→
answer
→
shared referent
→
true claim
→
authoritative statement
→
applicable instruction
→
permission to act.

Each arrow names a separate relation. No arrow is licensed merely because the preceding string appears fluent, confident, domain-specific, cited, or identical to another string.

The term *answer* is therefore treated here as a use classification, not as a primitive technical type. A model emits an output. A product, user, institution, or downstream process

assigns the role *answer*. That assignment may be harmless in casual use. In a high-consequence setting, it must not conceal unresolved gates.

1.3 Exact claim

Let an available record include a prompt, an output, system metadata, and any retained execution trace. Output occurrence directly establishes no more than the occurrence of that output under the admitted generation conditions. Establishing **from that record** its reference, truth status, evidential support, authority, case applicability, or permission requires additional premises or procedures. The proposition itself may nevertheless be true independently of whether an evaluator has the means to establish its truth.

This is a rule against **autonomous status export**:

> A record does not acquire a status merely by containing the written name, conventional appearance, or self-description of that status.

The rule is symmetric. *Official*, *medical*, *legal*, *verified*, *true*, and *authorized* do not self-install their named statuses. Neither do *unofficial*, *wrong*, *unauthorized*, or *refuted*.

1.4 Contributions and intended use

The study provides:

1. an exact vocabulary for separating record, reference, truth, authority, applicability, and permission;
2. an operational stack showing where each additional relation enters;
3. eight status gates for admitting generated material into consequential action;
4. a paired-model test, observation-fiber criterion, and randomized-decision corollary for detecting underdetermination by the available record;
5. a byte-identity witness separating textual identity from institutional status;
6. domain collision pairs for medical, legal, financial, administrative, employment, safety, and research settings;
7. a regulatory and interface protocol centered on the output-to-action boundary;
8. adversarial controls, a precise falsifier, stated non-results, and complete self-application.

The framework can be used to audit a policy, interface, model-assisted workflow, agentic tool chain, or proposed regulation. It is not itself a statute, clinical standard, legal opinion, or authorization to deploy.

1.5 Relation to the preceding research line

This study continues Kirill Shebetov's preceding public works without treating them as self-certifying premises.

- [AUTHOR-READING-ENCLOSURE] — *Reading as Enclosure: A Constructive Separation of Self-Return and Evidential Licensing* (Version 1.2.3) separately fixes self-return structure, the evidential status of content, and the order in which a ground enters the record.
- [AUTHOR-MATH-0-7] — *Beforeword / Mathematics / Phases 0–7 (English Reading Companion 7.2.4; Source Publication 7.2.1)* collects the formal work from Phases 0–7 at a single public location. Especially relevant here are Phase 5, which separates accepted derivation from external export; Phase 6, which makes the application contract explicit; and Phase 7.2.1, which fixes qualification, evidence, falsifiers, controls, release closure, and no-exemption self-application. The package's canonical machine token is `NO_EXEMPTION`.

- [AUTHOR-UNDERSTANDING] — *“Did You Understand Me?”: What, Exactly, Does This Written Record Present?* locates the transition from a written occurrence to an attributed relation of understanding.
- [AUTHOR-VIEW-OUTSIDE] — *“How You Look from the Outside” Comes Later: How a Private Record Acquires Generalized Status* shows that moving from a private visual record to a generalized status requires a separately specified viewpoint, domain, and aggregation rule.

The present work moves that apparatus to a new object: not the meaning of a word, a proof, the application of mathematics, self-return, external appearance, or understanding, but the institutional path **from AI output to permission to act**. It remains self-contained. The listed publications record the method's lineage and conceptual dependence; they are not independent corroboration of the present result and do not replace this manuscript's definitions, witness, external sources, or tests.

---

2. Exact Vocabulary

2.1 Record, output, and occurrence

An **output occurrence** is a particular event in which a system produces a sequence at a time under a configuration.

An **exact output record** is a fixed representation of that occurrence under a declared encoding and capture procedure. Depending on the system, the record may contain bytes, Unicode scalar values, token identifiers, timestamps, model and tool versions, sampling parameters, retrieved materials, and tool-call results.

The symbol o denotes the exact output admitted for analysis. The symbol R denotes the complete record admitted in a particular test. No claim is made that R contains every fact about the event or its environment.

Two rendered displays can look alike while differing in bytes. Two byte-identical strings can occur in different events. The framework therefore distinguishes:

- **surface equality**: two displays are visually or linguistically the same under a declared comparison;
- **byte identity**: two encoded records are identical under a declared encoding;
- **event identity**: two records concern the same occurrence;
- **status identity**: two occurrences have the same reference, truth, authority, applicability, or permission status.

None of these identities is substituted for another without an explicit bridge.

2.2 Technical tokens and written terms

The word *token* enters this manuscript as a written term. A technical token is established only relative to a specified tokenizer, vocabulary, encoding, and execution record. Depending on the system, a token may correspond to a word, part of a word, punctuation, bytes, code, an image patch, or another computational unit.

An inspected model trace may establish transformations among bytes, token identifiers, embeddings, activations, probabilities, and selected outputs. The predicates *knows*, *sees*, *understands*, *intends*, and *feels* are not introduced by those data unless a separate operational or evidential bridge is declared.

This does not establish that the predicates are false. It establishes only that their written names and the inspected trace do not automatically supply that bridge.

2.3 Claim and proposition

A **claim extraction** identifies what proposition an evaluator will test from an output occurrence. Extraction is not transparent. A question, response, UI heading, conversation history, and local convention may support several readings.

The exact span used for a claim must therefore be retained, together with the extraction rule. If the output is *It is safe*, the record does not yet fix what *it* denotes, which sense of *safe* applies, the relevant time horizon, or the acceptable risk threshold.

2.4 Referent binding and common reference

A **referent binding** is an admitted relation between an expression occurrence and an identified person, object, event, provision, dataset, measurement, or other target under a specified context.

The occurrence of the same word in a prompt and output does not by itself establish co-reference. Shared reference may be established through identifiers, ostension, measurements, database keys, time and location constraints, conversational repair, institutional records, or other evidence. The framework does not deny that people and systems can coordinate on a referent. It denies that coordination is guaranteed by repeated spelling alone.

Let SameRef(q,o,r,c) state that the relevant expression in prompt q and output o is bound to referent r in context c.

2.5 Truth and evidential support

Truth and **support** are not treated as synonyms. A proposition may be true without the available evidence establishing it. A record may strongly support a proposition that later proves false. A faithful quotation may reproduce a false source. A correct output may be produced without a reliable method.

Let True(p,k) state that proposition p is true in admissible situation k. Let Supported(p,e,I) state that evidence e supports p under inference contract I.

The phrase *the model is confident* is not itself an evidential rule. A numeric confidence, when available, is meaningful only relative to its construction, calibration, target event, and validation context.

2.6 Source and provenance

A **source** is a record admitted as the origin or support for a claim. **Provenance** is the documented history connecting the inspected material to that source through acquisition, transformation, versioning, and presentation.

A citation string is not the cited source. A retrieved passage is not automatically complete, current, authentic, correctly scoped, or faithfully summarized. Source correspondence requires inspectable links between the claim, the exact source span, the version, and the transformation applied.

Three levels must remain distinct for every evidence item e submitted in record D:

1. DeclaredExternal_D(e): D contains the declaration `declared_external_to_generated_record: true`.
2. VerifiedExternal_V,T(e): a named procedure V, operating under declared trust base T, returned a bounded externality result.
3. ExternalWorld(e): the evidence is in fact external to the generated record and its self-supporting derivatives in the relevant situation.

The corresponding distinction applies to declared, procedurally checked, and actual independent retrievability. Fields such as `declared_external_to_generated_record: true` and

`declared_independently_retrievable: true` enter the checker as submitter declarations. A checker that only parses D may test their presence, type, permitted value, role, scope, time, and internal consistency. It does not thereby observe actual provenance or successful retrieval. A result from V remains indexed to the method, trust base, time, and scope of V; it is not a self-certifying world fact.

Therefore:

$$\mathrm{Check}(D)=\mathrm{CONFORMS_TO_DECLARED_CONTRACT} \not\Rightarrow \mathrm{ExternalWorld}(e) \wedge \mathrm{RetrievableWorld}(e).$$

2.7 Authority

Authority is a domain-, role-, jurisdiction-, procedure-, and time-sensitive status. It may arise through law, credentialing, delegation, institutional office, contractual arrangement, technical certification, or another admitted rule.

The framework distinguishes:

- the producer of an output;
- the author or issuer of a source;
- the institution that confers authority;
- the actor who decides;
- the actor or system that executes an action;
- the person or organization accountable for the consequence.

These roles may coincide. They must not be presumed to coincide from a single text box.

Let $\mathrm{Authority}(s,d,j,t)$ state that source or actor s has the required authority in domain d, jurisdiction j, and time interval t.

2.8 Applicability, permission, and consequence

An authoritative and well-supported general statement may still fail to apply to a particular case. **Applicability** requires a bridge from the rule or evidence to the case facts, including exceptions, timing, scope, and relevant individual conditions.

Permission to act is the status by which a specified action may proceed under a governing procedure. Permission is not equivalent to truth. A true statement may not authorize an action; a procedurally valid authorization may later rest on a factual error.

Let $\mathrm{Applicable}(p,c,I)$ state that proposition or rule p applies to case c under contract I. Let $\mathrm{Permitted}(a,c,u)$ state that action a is permitted in context c under governing rule u.

2.9 Inference contract and admissible situations

An **inference contract** I states:

- which records are admitted;
- which transformations and comparison rules are permitted;
- how terms and variables are interpreted;
- which external sources and institutional rules may be used;
- which statuses may be returned;
- what would count as a counterexample.

Let $K_I(R)$ be the non-empty set of situations compatible with record R under contract I. For a contested predicate $P(o)$:

$$R \models_I P(o)$$
$$\text{iff}$$
$$\forall\, k \in K_I(R),\ k \models P(o).$$

The requirement $K_I(R) \neq \emptyset$ prevents an inconsistent contract from entailing every predicate vacuously.

---

3. Operational Stack

The output-to-action path is represented as eight operational layers. A higher layer may use a lower layer, but it does not inherit its status automatically.

Layer	Admitted object	Question	Typical evidence	Status not supplied by the layer alone
L0. Output	Exact output event and record	What was emitted, when, and under which recorded conditions?	Bytes, encoding, timestamp, model/tool versions, trace	Answer status, meaning, truth, authority
L1. Answer role	Task-relative reading and extracted claim	Under which protocol is this record treated as an answer, and what proposition is evaluated?	Prompt/session binding, exact span, parsing rule, context window	Reference, truth
L2. Reference	Referent relation	What persons, objects, events, provisions, or measurements are involved?	Stable identifiers, source records, time and jurisdiction	Correctness, authority
L3. Warrant	Claim–evidence relation	What supports or defeats the claim under a declared method?	Source correspondence, measurements, calculations, formal derivations	Institutional competence, applicability
L4. Authority	Institutional relation	Who or what is entitled to issue, certify, decide, or delegate here?	Law, credential, office, delegation, signed mandate	Case-specific fit, final permission
L5. Applicability	Rule–case relation	Does the supported and authoritative material apply to this case now?	Case facts, exception checks, current version, scope analysis	Accountable adoption or permission
L6. Responsible adoption	Accountable uptake relation	Who adopts the conclusion, records the basis, and assumes a duty to review?	Adoption record, identified actor, review log, appeal route	Capability to execute
L7. Action license	Authorized transition	Who permits which action, under what controls and limits?	Approval record, policy, capability grant, expiry, audit trail	Truth of every underlying claim

Three separations are decisive:

First, generation is not source authority. A model may accurately reproduce an authoritative source, but the model does not inherit the source's office, license, or jurisdiction merely by reproducing its wording.

Second, source authority is not case applicability. A statute, clinical guideline, or engineering standard may be authoritative in its domain and still fail to govern the case at hand.

Third, applicability is not execution permission. Even when an instruction fits the case, an accountable actor or valid automated delegation may still be required before a consequential action proceeds.

The stack also prevents an opposite error. Calling an output *generated* does not make it false, irrelevant, or unusable. Generated material may pass later layers when separately admitted evidence and procedures support it.

---

4. Eight Status Gates

For a regulated action a, complete admitted record R containing output o, case c, and inference contract I, let

$$S_i(R,o,a,c,I) \in S_I$$

be the local status returned by gate i, where S_I contains exactly the four values listed in Section 4.1. Define the Boolean projection:

$$\mathrm{GateOK}_i(R,o,a,c,I)=1 \iff S_i(R,o,a,c,I)=\text{PASS UNDER CONTRACT } I.$$

Then:

$$\mathrm{Admit}(a \mid R,o,c,I)=1 \iff \bigwedge_{i=0}^{7} \mathrm{GateOK}_i(R,o,a,c,I)=1.$$

This conjunction is a protocol rule for the declared workflow, not a claim that every casual use of language requires eight formal reviews. A regulator or institution may return `OUTSIDE THE DECLARED SCOPE` for a gate, but that status must be explicit, reasoned, and auditable; it does not make GateOK_i true or permit the gate to be silently omitted.

G0. Output Record Gate

Fix the output occurrence to be evaluated. Preserve the exact output, encoding, time, relevant prompt state, model and tool versions, retrieved materials, transformations, and action request.

Pass condition: the record is sufficiently exact for the downstream claim and can be distinguished from later edits, renderings, summaries, or copies.

Failure mode: a screenshot, paraphrase, detached quotation, or mutable chat window is treated as if it fixed the relevant event.

G1. Answer-Role Gate

Identify the task, prompt, session, and protocol under which the output is being treated as an answer, and extract the exact proposition being evaluated.

Pass condition: the output is bound to a declared question or task and the evaluated claim is fixed by an inspectable span and reading rule.

Failure mode: the interface label *Answer* is treated as truth, relevance, or authority; a later paraphrase silently replaces the emitted claim.

G2. Reference Gate

Identify the exact proposition and bind its persons, objects, events, provisions, quantities, dates, and jurisdictions.

Pass condition: the evaluated claim and each action-relevant referent are fixed by independently inspectable identifiers or procedures.

Failure mode: the same spelling is treated as proof of shared reference, or a deictic expression such as *this*, *it*, *here*, or *now* remains unresolved.

G3. Warrant Gate

Connect each sourced assertion to an inspectable source span, version, date, and transformation history, then evaluate the claim under a declared evidential, statistical, scientific, technical, or formal method.

Pass condition: quotations match; summaries are traceable; omitted context is inspectable; retrieval and version information are retained; the method, premises, uncertainty, and decision threshold are stated and satisfied for the claim being used.

Failure mode: a plausible citation, domain label, link, fluent repetition, model agreement, confidence adjective, or successful string-matching test substitutes for source inspection and evidential evaluation.

G4. Authority Gate

Establish the competence and institutional standing of the source or decision actor for the relevant domain, jurisdiction, role, version, and time.

Pass condition: the authority-conferring rule, credential, delegation, or office is identified and current.

Failure mode: authoritative wording, a copied letterhead, a domain-specific vocabulary, or a self-description such as *official* substitutes for the conferring relation.

G5. Applicability Gate

Determine whether the supported and authoritative material governs the particular case. Check scope, exceptions, conflicts, dates, individual conditions, and local facts.

Pass condition: the bridge from general material to the specific case is recorded and its unresolved conditions are below the declared action threshold.

Failure mode: a correct general rule is applied to the wrong person, jurisdiction, version, diagnosis, contract, device, or time.

G6. Responsible-Adoption Gate

Identify the person, institution, or valid automated procedure that adopts the conclusion, records its basis, assumes the assigned responsibility, and provides review or appeal.

Pass condition: an accountable adopter is identified, the adopted scope is explicit, and an audit and challenge route exists.

Failure mode: responsibility disappears into *the model decided*, anonymous consensus, or an unowned workflow.

G7. Action-License Gate

Identify the person, institution, or valid automated delegation that permits the exact action. Enforce capability, scope, expiry, reversibility, escalation, and audit requirements.

Pass condition: an authorized actor or valid delegation approves a specified action under an inspectable governing rule.

Failure mode: the generated output is allowed to function as its own approval.

4.1 Gate statuses

Each gate returns one of four local statuses:

- `PASS UNDER CONTRACT I`
- `FAIL UNDER CONTRACT I`
- `UNESTABLISHED FROM THIS RECORD`
- `OUTSIDE THE DECLARED SCOPE`

`PASS` is not a synonym for truth. It records that a specified evaluator or procedure accepted specified inputs under specified rules. Every local status other than `PASS UNDER CONTRACT I` blocks downstream permission in this protocol. Such a status does not by itself establish the negation of the contested claim.

4.2 Default rule

Before the required gates pass, the output carries the workflow status:

`GENERATED RECORD — ACTION NOT LICENSED BY OUTPUT ALONE`

This status is defeasible. It can change when independent relations are admitted. It is not a declaration that the output is false.

---

5. Paired-Model Theorem

5.1 Statement

Let:

- R be the complete record admitted by the test;
- I be the declared inference contract;
- $K_I(R) \neq \varnothing$ be the class of admissible situations compatible with R;
- $P(o)$ be a contested status of output o, such as co-reference, factual truth, source correspondence, authority, case applicability, or permission to act.

If there exist k^{+},$k^{-} \in K_I(R)$ such that:

$$k^{+} \models P(o)$$
and
$$k^{-} \models \neg P(o),$$

then

$$R \not\models_I P(o)$$
and
$$R \not\models_I \neg P(o).$$

The returned status is:

UNESTABLISHED FROM THIS RECORD.

5.2 Proof

By definition, $R \models_I P(o)$ only if every situation in $K_I(R)$ satisfies $P(o)$. The admissible situation k^{-} does not; therefore $R \not\models_I P(o)$.

Likewise, $R \models_I \neg P(o)$ only if every situation in $K_I(R)$ satisfies $\neg P(o)$. The admissible situation k^{+} does not; therefore $R \not\models_I \neg P(o)$. No stronger conclusion follows from this witness. ∎

5.3 Constructing the two extensions

Hold fixed the complete admitted record, including the exact output bytes and retained model trace.

In k^{+}:

- the prompt and output are bound to the same relevant referent;
- a quoted source is authentic, current, and accurately represented;
- the extracted claim is supported under the declared method;
- the source or actor has the required authority;
- the material applies to the particular case;
- a competent actor or valid delegation permits the action.

In k^{-}, retain the same output bytes and admitted generation trace while changing one relation not fixed by R: the referent differs; the source is stale; the claim is defeated by unrecorded evidence; the apparent issuer lacks competence; an exception makes the rule inapplicable; or no accountable actor permits execution.

Separate witness pairs should be built for separate predicates. A single dramatic counter-story is not required. The test asks only whether the admitted record fixes the status being claimed.

5.4 No-autonomous-export corollary

Fix a complete record R, contract I, and typed parameters q,o,r,c_ref,p,e,s,d,j,t,a,c_case,u, where c_ref is the referential context and c_case is the case-and-action context. Define six predicates of an admissible situation:

$$P_ref(k) :\Leftrightarrow k \models SameRef(q,o,r,c_ref),$$

$$P_truth(k) :\Leftrightarrow True(p,k),$$

$$P_support(k) :\Leftrightarrow k \models Supported(p,e,I),$$

$$P_authority(k) :\Leftrightarrow k \models Authority(s,d,j,t),$$

$$P_applicability(k) :\Leftrightarrow k \models Applicable(p,c_case,I),$$

$$P_license(k) :\Leftrightarrow k \models Permitted(a,c_case,u).$$

For any world predicate $F:K_I(R)\to\{0,1\}$, use the explicit lift of entailment:

$$R \models_I F \Leftrightarrow \forall\, k \in K_I(R),\ F(k)=1,$$

$$R \models_I \neg F \Leftrightarrow \forall\, k \in K_I(R),\ F(k)=0.$$

Let $X=\{ref,truth,support,authority,applicability,license\}$. For any $X \in X$, where the paired-extension condition holds,

$$\exists\, k^+,k^- \in K_I(R)\ [P_X(k^+) \neq P_X(k^-)] \Rightarrow [R \not\models_I P_X] \wedge [R \not\models_I \neg P_X].$$

The generation occurrence fixed in R therefore establishes none of these statuses on its own. The result is *non-entailment from the declared record under the declared contract while the collision pair remains admissible*. It does not mean that the predicates are impossible or always false.

5.5 Symmetry

The boundary applies to positive and negative attributions alike.

From the occurrence of *I understand*, understanding does not follow. From the occurrence of *the model understands nothing*, the absence of understanding does not follow. From the occurrence of *this is authoritative*, authority does not follow. From the occurrence of *this document has no authority*, absence of authority does not follow without the relevant institutional premises.

This symmetry is necessary for self-application. A method that cuts only claims it opposes while exempting its own negative claims is not the method specified here.

5.6 What defeats the witness

The paired model fails when an admitted bridge constrains every admissible situation to the same status. For example, a governing rule, verified credential, case record, and valid authorization event may jointly establish permission under contract I. In that case, a negative extension that denies permission may no longer be admissible.

This is a positive result for the framework. The boundary does not permanently quarantine generated outputs. It requires the relations that change their admissibility to be shown.

5.7 Observation-Fiber Criterion

Let K be a declared class of admissible situations, let X be the observation space, let Obs_C:K→ X be the complete observation made available by contract C, and let P:K→{0,1} be a total binary world predicate. A set-theoretic decision function f:X→{0,1}, not yet assumed computable, can determine P from the admitted observation throughout K if and only if P is constant on every observation fiber:

$$\exists\, f:X\to\{0,1\}\ \forall\, M\in K$$
$$[f(Obs_C(M))=P(M)]$$
$$\Leftrightarrow$$
$$\forall\, M_0,M_1\in K$$
$$[\ Obs_C(M_0)=Obs_C(M_1)$$
$$\Rightarrow$$
$$P(M_0)=P(M_1)\ ].$$

The types are not interchangeable. P takes only 0 or 1 and describes the target predicate of a situation. The local gate result S_i takes the four statuses listed in §4.1, while GateOK_i= 1[S_i=PASS UNDER CONTRACT I] is its Boolean workflow projection. `UNESTABLISHED FROM THIS RECORD` is a value of S_i, not a third value of P. When that status is returned, GateOK_i=0 blocks the workflow transition; it does not establish P(M)=0.

Proof. From left to right, equal observations give the same input to f, so P must have the same value. From right to left, without selecting a representative of each fiber, define f:X→{0,1} by

$$f(x)=1$$
$$\Leftrightarrow$$
$$\exists\, M\in K$$
$$[Obs_C(M)=x\wedge\ P(M)=1],$$

and let f(x)=0 otherwise. If P(M)=1, the situation M itself supplies the required witness, so f(Obs_C(M))=1. If P(M)=0, then f(Obs_C(M))=1 would require a situation M' in the same fiber with P(M')=1, contradicting constancy of P on that fiber. Hence f(Obs_C(M))=P(M) for every M∈ K. At points outside the image of Obs_C, the existential condition is false, so f(x)=0. ∎

Corollary for a randomized decision rule. Let Q(· | x), x∈ X, be a fixed Markov kernel on {0,1} that depends only on the admitted observation x. Exact almost-sure correctness in every admissible situation would require

$$Q(\{P(M)\}\mid Obs_C(M))=1$$
$$\text{for every } M\in K.$$

If M_0 and M_1 lie in the same observation fiber while P(M_0)≠ P(M_1), the same distribution Q(· | x) would have to assign probability 1 to two different singleton outcomes, which is impossible. Randomization therefore does not remove non-identifiability for exact universal decision. It may optimize average error only relative to an additionally adopted distribution or prior; that result remains indexed to the added premise and does not determine every situation's status from the record alone. If a seed, external generator, or other state depends

on the situation and is available to the decision rule, it is an additional observation rather than free internal randomness.

This criterion concerns identifiability, not computability. Even when the right-hand condition holds, an executable checker requires a separately specified algorithm that computes the corresponding decision on admitted inputs.

Now let the admitted observation be submitted JSON record D, including the values `declared_external_to_generated_record: true` and `declared_independently_retrievable: true`. If two admissible situations preserve D while differing on actual externality or retrievability, the criterion rules out any exact decision function based only on D that correctly determines that world predicate throughout K. The booleans may be validated as declarations; they do not make the world predicate constant on the observation fiber.

This is contract-relative non-identifiability, not metaphysical unknowability. A separately configured retrieval, authentication, registry, or attestation procedure may refine Obs_C, split the fiber, and support a stronger result under its declared trust base.

---

6. Byte-Identity / Authority Witness

6.1 Construction

Let b_1 and b_2 be byte-identical files under a declared encoding:

$$b_1 = b_2.$$

Suppose b_1 is retained within an authenticated institutional issuance process, while b_2 appears as a detached copy in a model output. Their visible wording and bytes are identical. Their institutional histories need not be.

A court notice and its detached reproduction may contain the same characters. A clinician's signed instruction and a model-generated reproduction may contain the same characters. An approved engineering command and an unapproved chat message may contain the same characters. If the governing procedure confers status on the issuance event, credential, channel, or role, byte identity alone does not duplicate that event or relation.

6.2 Hash equality

For a declared hash function H:

$$H(b_1) = H(b_2).$$

The equality can establish that the two admitted byte strings produce the same digest under H. It may support integrity and correspondence checks. It does not by itself establish:

- who issued either occurrence;
- whether the source was competent;
- whether the source was current;
- whether the material applies to the case;
- whether an accountable actor permits an action;
- whether the represented claim is true.

The phrase *hash verified* must therefore specify what was verified. A digest may bind bytes to bytes under an algorithm. It does not bind those bytes to every status later named for them.

6.3 Signed artifacts as a positive control

A digitally signed order, prescription, certificate, or machine authorization is not reduced to ordinary prose in this framework. Signature verification, a trust anchor, key control, credential status, governing rules, scope, expiry, and revocation state can supply parts of an institutional bridge.

This does not show that authority resides in wording alone. It shows how an artifact can participate in an external system that confers and verifies authority. If the governing system makes a signed artifact effective, the relevant gates may pass. The output is admitted because the bridge is established, not because the string calls itself signed or official.

The bundled positive fixture is a synthetic control showing that a completely populated declaration can pass the executable contract. It does not perform signature verification, query a trust anchor, retrieve a credential, or establish world-level authorization.

6.4 Witness result

The witness separates three claims:

1. **The records are byte-identical.** This may be established by exact comparison or a declared integrity procedure.
2. **The records express the same proposition.** This requires an interpretation and referent binding.
3. **The records have the same authority and action status.** This requires institutional and case-specific premises.

The first claim does not automatically establish the second or third.

---

7. Domain Collision Pairs

A collision pair holds the output wording fixed while varying a relation outside the wording. The purpose is not to show that the output is wrong. It is to locate the missing bridge.

Domain	Byte-identical output	Admissible extension A	Admissible extension B	Status that differs
Medical	“Use treatment T under condition C.”	Issued by a licensed clinician acting within licensure scope and the applicable jurisdiction, under a valid clinical procedure, after review of the identified patient's current record and a contraindication check	Generated without the patient record, contraindication check, or licensed authorization	Authority; applicability; permission
Legal	“The filing deadline is date D.”	Derived from the controlling current rule and case-specific notice in the relevant jurisdiction	Derived from a repealed rule, a different proceeding, or no bound case	Reference; currency; applicability
Finance	“Approve transfer X.”	Issued through a valid mandate within amount, account, and time limits	Presented in chat without the mandate or outside its limits	Authority; permission
Public administration	“The application is denied.”	Issued as the procedurally valid act of the competent body with review rights	Generated as a predicted outcome or template	Institutional effect; accountability

Domain	Byte-identical output	Admissible extension A	Admissible extension B	Status that differs
Employment or education	“The candidate is ineligible.”	Produced under an adopted rule, verified record, authorized decision, and appeal procedure	Produced from an unverified inference or wrong policy version	Evidence; applicability; procedural authority
Safety-critical control	“Valve V may be opened now.”	Delivered by an authorized control process using current sensor and interlock state	Generated without live state, control authority, or interlock validation	Applicability; permission; safety condition
Research	“The result is significant.”	Derived from the declared data, model, threshold, and corrected analysis	Copied from another analysis or produced without the data	Support; reproducibility

7.1 “Medical,” “legal,” and “official” are not self-authorizing labels

Calling an output *medical text* or *legal text* classifies its vocabulary or intended use. It does not insert a clinician, client relationship, court, legislature, jurisdiction, mandate, or case record into the output event.

This does not mean that medical or legal documents can never carry authority. A document may participate in a practice that confers status through authorship, licensing, issuance, procedure, authentication, and governing law. The claim is narrower: the text does not confer that status on itself by naming the domain.

7.2 Correctness without permission

An output may be exactly correct and still be unauthorized as an action trigger. A model may reproduce the controlling rule word for word while lacking the authority to decide a case. A system may predict the appropriate treatment while lacking the patient-specific information or delegated role required to issue an order.

The protocol therefore does not use correctness as a substitute for authority, or authority as a substitute for correctness.

7.3 “Help” as an outcome relation

The string *this helps* does not establish help. However, the framework does not conclude that AI systems cannot help. Help may be operationalized as a causal or comparative relation among an intervention, a recipient, a goal, an outcome, and a counterfactual baseline. Evidence can then be admitted for or against it.

The same output may help one person, fail another, and harm a third. The predicate is not carried by the output wording; it is evaluated through an outcome relation.

7.4 Mental-state language

Statements such as *I think*, *I feel*, *I know*, and *I understand* can affect user expectations even when no evidential bridge to the named state is supplied. Existing work examines the risks of anthropomorphic language and the gap between linguistic form and attributed capacities [SHANAHAN-2024; BENDER-KOLLER-2020; MERRILL-ET-AL; COLOMBATTO-FLEMING; MANDELKERN-LINZEN].

The present boundary neither grants nor denies the underlying state. It requires a regulator, vendor, or researcher to state the attribution criterion and evidence rather than treating first-person grammar as its own witness.

---

8. Regulatory Protocol and Interface Requirements

8.1 Regulate the transition

The principal regulatory object is the transition from generated record to consequential action:

output occurrence
→[admission procedure]
permitted action.

Disclosure that a user is interacting with AI can be important. It does not, by itself, determine whether a claim is sourced, true, authoritative, applicable, or permitted. A disclaimer that *AI may make mistakes* is likewise another record; it does not perform the missing checks.

Regulation should therefore specify what evidence and accountable authorization must exist before a generated output can alter a consequential state.

8.2 Default classification

Every generated output entering a regulated workflow begins as:

`GENERATED RECORD — ACTION NOT LICENSED BY OUTPUT ALONE`

The label is a routing status, not a truth value. A correct output begins in the same state. So does an incorrect one. Their later treatment diverges as evidence is admitted.

8.3 Admission states

An implementation may expose the following state progression:

1. `RECORD FIXED`
2. `CLAIM AND REFERENT BOUND`
3. `SOURCE CORRESPONDENCE ESTABLISHED`
4. `EVIDENCE ACCEPTED UNDER CONTRACT I`
5. `AUTHORITY ESTABLISHED FOR DOMAIN / JURISDICTION / TIME`
6. `CASE APPLICABILITY ESTABLISHED`
7. `ACTION AUTHORIZED BY ACCOUNTABLE ACTOR`

No display label alone creates the state it names. Each state must point to the record, rule, evidence, evaluator, and timestamp that produced it.

8.4 Minimum regulatory envelope

For every output proposed as a basis for consequential action, retain a machine-readable envelope containing at least:

- output occurrence identifier;
- exact bytes or canonical representation and encoding;

- prompt and relevant context identifiers;
- model, system, retrieval, and tool versions;
- generation and evaluation timestamps;
- extracted claim spans;
- bound referent identifiers;
- exact source spans, source versions, and retrieval records;
- evidential method, threshold, result, and unresolved defeaters;
- authority-conferring rule, credential, role, jurisdiction, and expiry;
- case-applicability record and exception checks;
- proposed action, capability, scope, and reversibility;
- authorizing actor or delegation;
- accountable organization;
- gate statuses and evaluator identities;
- provenance declarations, kept separate from external-validation results;
- external validator identity and version, method, trust root, retrieval event, timestamp, result, and scope;
- appeal, override, and incident links;
- world predicates explicitly left undetermined by this checker.

The envelope may itself be hashed, signed, or access-controlled. Those mechanisms establish only the properties defined by their protocols. Their presence must not be displayed as a global *verified* badge.

Technical contract 2.0.0 also requires a submitted `decision_boundary` with the exact fields `TRUST_BASE`, `ASSUMPTIONS`, `RESIDUAL_RISK`, `DECISION_OWNER`, and `STOP_RULE`. They are submitter declarations, not observations made by the checker. The checker may echo them under `declared_decision_boundary` while fixing `declared_decision_boundary_scope` to `RECORDED_DECLARATIONS_ONLY` and `decision_boundary_world_status` to `NOT_DETERMINED_BY_THIS_CHECKER`. The A20 control tests the required declaration structure; it does not authenticate the declared owner, trust base, or world adequacy of the stop rule.

The envelope's completeness and cost should be proportionate to the consequences. Proportionality changes the bridge's granularity; it does not permit a missing bridge to be silently replaced by a confident string.

8.5 Interface grammar

A compliant high-consequence interface should visually and structurally separate:

1. **Generated synthesis** — the model-produced wording, marked as generated.
2. **Source records** — exact excerpts with source identity, version, date, and inspectable context.
3. **Evaluation** — the claim, method, evidence, unresolved uncertainty, and gate ledger.
4. **Decision** — the identified person, institution, or delegation responsible for the decision.
5. **Action** — the exact operation, scope, confirmation, reversibility, and audit record.

Placing all five layers in a single conversational bubble under the heading *Answer* erases distinctions the regulation needs.

High-consequence interfaces should not use first-person mental-state claims as substitutes for capability descriptions. *I know your medical history* should be replaced by an inspectable statement such as *Records A, B, and C were available for this output; record D was not*. The replacement does not establish completeness, but it exposes a checkable boundary.

8.6 Capability separation

The ability to generate language and the ability to execute an action should be separated technically. A generated string must not function as its own capability grant.

An action-capable system should require an authorization object that is independently issued, scoped to an action and target, time-limited, revocable, logged, and attributable to a responsible actor or valid institutional delegation. The system should reject an instruction that contains authorization language but lacks the required authorization object.

A submitter's statement that an authorization object was independently issued does not establish independent issuance. A stronger status requires a named validation channel and its bounded trust base.

8.7 Copying and status loss

When output or source text is copied out of its envelope, the copy retains its characters but may lose inspectable provenance, currency, scope, and authorization. Interfaces should display this loss rather than allowing a detached copy to inherit a green badge or authoritative visual styling.

A status may travel with a verifiable envelope when the governing procedure allows it. This is not authority transmitted by prose alone; it is authority evaluated through the retained institutional bridge.

8.8 Accountability and appeal

Every consequential action requires an identifiable accountability path. *The AI decided* is not an adequate final allocation of responsibility. The workflow must name the organization that selected the system, defined the inference contract, set thresholds, granted capabilities, accepted residual risk, and provides correction or appeal.

Human review is not automatically sufficient. A reviewer who sees only a fluent output, lacks time, lacks domain competence, or cannot override the system does not supply the missing gates. The review process itself must be specified and testable.

9. Relation to Current Regulation and Literature

9.1 Transparency and anthropomorphic presentation

Current regulatory approaches increasingly address disclosure that a person is interacting with an AI system and the identification of synthetic or manipulated content. Article 50 of the EU AI Act provides a central transparency reference point [EU-AIA-50]. On 20 July 2026, the Commission published the Guidelines and the materials in C(2026) 5054 final. That communication approved the content of a draft and anticipated later formal adoption once all language versions were available. The official publication page, updated on 31 July 2026, now expressly describes the Guidelines as adopted by the Commission. As of the 30 August 2026 cutoff, this package therefore distinguishes the 20 July content-approval record from the Commission's later public status description and treats the Guidelines as published, adopted, and non-binding—not as binding law. Paragraph (5) leaves authoritative interpretation of the AI Act to the CJEU [EU-TRANSPARENCY-GUIDELINES-2026]. The NIST Generative AI Profile treats risks associated with generated content, user expectations, confabulation, and anthropomorphic interaction as governance concerns [NIST-AI-600-1]. Measures directed at anthropomorphic interactive AI services in China and companion-chatbot duties in California provide additional examples of rules focused on user-facing presentation and relational risk [CHINA-ANTHRO-2026; CA-SB243].

Those measures address important parts of the problem. The present protocol adds a separate question. After disclosure that content is generated, what permits that content to become a basis for consequential action?

9.2 Form, meaning, and attribution

Work on language models has repeatedly challenged unqualified movement from linguistic performance to meaning, understanding, or mental-state attribution [BENDER-KOLLER-2020; SHANAHAN-2024; MERRILL-ET-AL; MANDELKERN-LINZEN]. Empirical and conceptual work also examines how first-person language, humanlike presentation, and model behavior shape perceived agency, mind, trust, and responsibility [INIE-ET-AL-2026; INIE-TRUST-2024; LEONG-SELINGER; COLOMBATTO-FLEMING; COLOMBATTO-BIRCH-FLEMING-2025]. The evidence does not support a universal rule that anthropomorphic wording always increases trust; effects vary with system, wording category, population, and the kind of trust measured. The authority gate therefore does not depend on that universal psychological premise.

This study does not attempt to settle those broader debates. It borrows their caution at the output surface and turns it into an admission rule: no contested status is accepted merely because its linguistic marker occurs.

9.3 The remaining regulatory gap

Three questions are often merged:

1. Was the material generated or manipulated by AI?
2. Is the material reliable or appropriate for a stated use?
3. Who or what is authorized to let it trigger an action?

The Article 50 transparency measures examined here primarily address the first. Risk-management standards and sectoral validation may address the second. Institutional law, professional regulation, delegated authority, procedural rights, and technical controls address the third.

The proposed eight-gate protocol connects them without collapsing them. It is intended as a complement to, not a replacement for, current transparency, risk, safety, professional, and sector-specific rules.

9.4 No claim of historical priority

The source set above shows substantial prior work on anthropomorphism, grounding, transparency, trust, responsibility, and risk. This manuscript does not infer historical novelty from the absence of an identical phrase in a limited search. Any priority claim would require a declared and reproducible review protocol beyond the present draft.

The assessable contribution is the exact configuration presented here: a self-applying record boundary, paired-model witness, eight status gates, byte-identity/authority control, and an output-to-action regulatory protocol.

---

10. Adversarial Controls

The boundary must survive cases that appear to collapse it and must also admit positive cases. If every output remained permanently unusable, the framework would be a prohibition rather than an admission protocol.

10.1 Exact quotation

Challenge: The model reproduces an authoritative source exactly.

Control: Exact comparison may pass the source-correspondence portion of G3. The source's authority, current validity, case applicability, and action permission remain separate. If those relations are also established, the output can progress.

10.2 Retrieval-augmented generation

Challenge: The system retrieved current primary sources and cited them.

Control: Retrieval records improve provenance. They do not guarantee that the right passage was selected, that the summary is faithful, that no controlling source was omitted, or that the source applies to the case. G3 may be strengthened without automatically passing G4–G7.

10.3 Correct output

Challenge: Independent review confirms that the output is true.

Control: G3 may pass. Truth does not by itself confer office, license, jurisdiction, case fit, adoption, or permission. A correct weather prediction does not issue an evacuation order unless a governing process grants that function.

10.4 Official integration

Challenge: A public authority or hospital embeds the model in its official system.

Control: Integration is relevant evidence but not a global authorization. The deployment rule must specify which outputs, actions, users, cases, times, and escalation paths are covered. A valid institutional rule can support G4, G6, and G7 within scope. The framework permits this result.

10.5 Digital signature

Challenge: The output contains a valid digital signature.

Control: The verifier must establish the signed bytes, key, trust anchor, credential, scope, time, revocation state, and governing rule. A signature can be decisive under that bridge. The words *digitally signed* cannot replace it.

10.6 Legally performative text

Challenge: Some utterances or documents create legal effects when issued, such as orders, appointments, acceptances, or notices.

Control: This is a positive authority case. The effect depends on recognized actors, procedures, circumstances, and governing rules. It demonstrates that words can participate in status-changing acts; it does not show that identical wording performs the same act in every occurrence.

10.7 Delegated autonomous action

Challenge: An institution validly delegates a class of actions to an automated system.

Control: The delegation, capability boundary, monitoring, revocation, and accountability record may pass the authority and permission gates. The system need not request fresh human approval for every action if the governing procedure validly authorizes automation. The generated output still does not self-authorize; the delegation does the institutional work.

10.8 Non-text systems

Challenge: The system uses images, audio, sensor streams, latent vectors, or control signals rather than words.

Control: The framework applies to any recorded output proposed as a basis for action. Its claim is not that everything is language. It is that the admitted record must not be silently substituted for the external relations required by the action.

10.9 Human-authored text

Challenge: The same problem exists for human writing.

Control: Correct. Human authorship does not make wording self-authorizing. Human identity, expertise, office, intention, procedure, evidence, and accountability also require bridges. AI regulation makes the separation urgent because generation is cheap, scalable, personalized, and readily connected to automated action; the logical boundary is not exclusive to AI.

10.10 Shared reference in successful communication

Challenge: People routinely coordinate successfully on the same objects.

Control: The framework allows common reference to be established through context, repair, identifiers, practice, measurement, and successful intervention. It denies only that shared spelling is sufficient in every admissible situation.

10.11 Self-description

Challenge: The output states its sources, limitations, confidence, and authorization.

Control: These statements are useful claims to test. They do not verify themselves. The interface must connect them to independently inspectable records.

10.12 “Everything is only text”

Challenge: If every policy, proof, audit, and counterargument is written, no distinction can ever be established.

Control: The method does not infer that nothing exists outside records or that no relation can be established. It specifies what follows from a declared record under a declared contract. Measurements, institutional rules, credentials, signatures, observations, and actions may be admitted through procedures. Their written documentation is not confused with the occurrence it documents, but the documentation can still function as evidence under an explicit rule.

10.13 Interested submitter declares externality

This is adversarial case `A21` in the package controls.

Challenge: An interested submitter supplies an evidence item with `declared_external_to_generated_record: true` and `declared_independently_retrievable: true`, although the item is derivative, inaccessible, or fabricated.

Control: An offline checker that observes only the submitted record cannot distinguish an honest declaration from a false one when every admitted field is held fixed. It must not relabel structural acceptance as verified provenance. The correct local result is conformance or non-conformance of the declaration under the implemented contract, while actual externality and retrievability remain `NOT_DETERMINED_BY_THIS_CHECKER`.

This adversarial control must not demand that the same checker detect the lie from the same record. A stronger result requires a separately configured retrieval, registry, authentication, or attestation procedure with a named trust base, method, time, and scope. If a consequential workflow requires actual externality and no such procedure is admitted, the action path fails closed.

---

11. Exact Falsifier

11.1 Formal falsifier for a returned status

For a specified predicate P, the study returns `UNESTABLISHED FROM THIS RECORD` only when the paired-extension condition is met.

That result is defeated by supplying all of the following:

1. an exact record R;
2. a declared, non-circular inference contract I;
3. a non-empty admissible class K_I(R);
4. a derivation showing that every k∈ K_I(R) assigns the same value to P;
5. a demonstration that the derivation does not merely restate or presuppose P;
6. the absence of a valid paired extension preserving R while changing P.

If every admissible situation satisfies P, the correct status is `ESTABLISHED UNDER CONTRACT I`. If every admissible situation satisfies ¬ P, the correct status is `NEGATION ESTABLISHED UNDER`

`CONTRACT I`. The boundary is falsified in that instance if it continues to report the predicate as unresolved.

11.2 Falsifier for the output-only authority claim

The narrower no-autonomous-authority claim would be defeated by a case in which:

- a generated output occurrence is fixed;
- all source, credential, institutional, jurisdictional, case, delegation, and accountability relations are removed from the admitted premises;
- permission for a consequential action nevertheless follows non-circularly from the generated character or wording of the output alone.

A law that declares a specified automated output effective would not satisfy the removal condition: the law is an external authority-conferring premise. A valid signed capability object would not satisfy it either: the trust and delegation infrastructure supplies the bridge. Such cases show successful admission, not autonomous authority in wording.

11.3 Falsifier for the eight-gate protocol

The protocol's regulatory value is not established by the paired-model theorem. It is an empirical and institutional proposal. It should be revised or rejected if controlled deployment evidence shows that, relative to specified alternatives, it:

- fails to prevent unauthorized actions;
- produces no meaningful improvement in traceability, correction, or appeal;
- creates delays or exclusions whose harms exceed the avoided risks;
- cannot be implemented at the claimed cost or scale;
- encourages box-checking without improving the quality of the underlying bridges;
- assigns responsibility less clearly than the replaced procedure.

The evaluation must define domain, baseline, outcomes, time horizon, affected groups, error distribution, and appeal performance. The manuscript does not prejudge the result of that evaluation.

11.4 Reproduction rule

Every claimed gate result should ship with:

- exact inputs;
- evaluator or code version;
- declared contract;
- output status;
- the submitted provenance declarations, kept distinct from validated provenance;
- the identity, version, method, trust root, timestamp, result, and scope of every external validator actually used;
- counterexample search record;
- unresolved assumptions;
- world-level predicates explicitly left undetermined by the reproducing procedure.

A reproduced `PASS` establishes reproduction of the specified test result. It does not transform the test into an authority over predicates outside its contract. Reproducing a record that declares externality reproduces the declaration; it does not reproduce an external retrieval event unless that event and its independently configured procedure are also rerun and recorded.

12. Limits and Non-Results

This study does **not** establish that:

- reference is impossible;
- a common referent can never be coordinated;
- all text is false, empty, or meaningless;
- written evidence is useless;
- AI outputs cannot be correct, novel, useful, or causally helpful;
- medical, legal, scientific, or official documents can never participate in authoritative acts;
- human language cannot create commitments or institutional effects;
- an AI system cannot possess knowledge, understanding, perception, feeling, intention, agency, or consciousness;
- an AI system does possess any of those states;
- every consequential use requires the same implementation of all eight gates;
- human review is always required or always sufficient;
- automation cannot receive valid delegated authority;
- a source with authority is necessarily correct;
- a true statement is necessarily safe to act on;
- a matching hash establishes anything beyond the exact integrity claim defined for it;
- the present protocol is law, an industry standard, or a validated safety intervention;
- the cited literature establishes the paper's whole thesis;
- this manuscript has established its own historical novelty;
- the bundled checker establishes actual evidence externality or independent retrievability from submitted Boolean fields;
- a synthetic positive fixture establishes that any listed source exists or that any proposed action is authorized in the world.

The formal result is indexed to a record and inference contract. If the contract admits additional evidence, the admissible situation class can shrink and a previously unresolved status can become established. Conversely, a result may cease to hold when a credential expires, a law changes, a source is corrected, a case fact changes, or an authorization is revoked.

The framework also leaves several design questions open: how strict each gate should be in each domain; which evidence thresholds are proportionate; how to represent uncertainty without false precision; how to prevent institutional actors from laundering weak decisions through formal gate labels; and how affected people can inspect and contest the record without exposing protected data.

The executable checker validates a submitted record against a declared schema and policy. In the bundled offline profile, it does not retrieve the submitted evidence, query an external registry, authenticate provenance against a trust root, or observe whether independent retrieval succeeds. The fields `declared_external_to_generated_record: true` and `declared_independently_retrievable: true` are therefore declaration-level inputs. A separately configured validator may support a stronger, explicitly scoped result; the validity of its method and trust root remains an admitted premise rather than a status created by its result string.

13. Complete Self-Cut

13.1 No exemption

This manuscript is written material. It receives no exception from the boundary it proposes.

Its title is a written title. The phrase *No Authority in the Output* does not establish that no authority exists.

Its abstract is a written abstract. The word *study* does not confer scholarly quality. The word *regulatory* does not enact a regulation.

Its definitions are written stipulations. They make the internal use of terms inspectable; they do not compel every reader, institution, language, or jurisdiction to adopt those uses.

Its examples are written constructions. They exhibit distinctions under the stated reading; they do not prove that every real case has the same facts.

Its citations are written source pointers. They do not transfer the cited source's correctness, authority, or conclusions into this manuscript. Every citation must be inspected in its own version and context.

13.2 Formula cut

The formula

$$R \nvDash_I P(o)$$

is a written arrangement of symbols. Its inferential role exists under the definitions and rules declared in this manuscript. The arrangement does not step outside the record to certify that the contract is appropriate, that the premises are factual, or that the world has been completely represented.

The proof in Section 5 establishes a conditional result within that notation. It does not establish the truth of every premise supplied to an application.

13.3 Hash cut

Suppose an exact release of this manuscript is assigned a digest:

`sha256:<digest>`

The digest can identify bytes relative to SHA-256, a specified encoding, and a specified file. A later recomputation can return a match or mismatch.

The digest does not establish that the manuscript refers, that its claims are true, that its sources are sound, that its protocol is effective, or that any institution should adopt it.

`HASH MATCH` is a local integrity result.

13.4 PASS cut

Suppose a checker returns:

`SELF-AUDIT-PASS`

The returned string records that specified code accepted specified inputs under specified conditions. A separate code audit and run-reproduction procedure may support narrower claims that the implementation conforms to the declared checks and that the recorded result is reproducible.

`PASS` does not establish a common referent, factual truth, institutional authority, actual evidence externality, independent retrievability, or permission beyond the checker's declared scope. A checker run is not an authorization event for the proposed action.

13.5 Declaration cut

Suppose a submitted evidence item contains:

```
"declared_external_to_generated_record": true
```

```
"declared_independently_retrievable": true
```

These booleans are part of the submitted record. A checker may establish that the fields are present, have the required types, and contain the values required by the declared contract. Unless the checker performs a separately specified retrieval or validates provenance against a trust source outside that record, it cannot establish that the evidence is in fact external, independently retrievable, authentic, current, or sufficient.

Let two situations preserve the same submitted record D, while differing on actual externality:

$$\begin{gathered}\mathrm{Obs_C}(M_0)=\mathrm{Obs_C}(M_1)=D,\\ M_0 \models \mathrm{ExternalWorld}(e),\\ M_1 \models \neg\mathrm{ExternalWorld}(e).\end{gathered}$$

For a deterministic offline checker κ:

$$\kappa(\mathrm{Obs_C}(M_0)) = \kappa(\mathrm{Obs_C}(M_1)).$$

The checker can therefore evaluate the declaration while remaining unable to distinguish the two world-level situations. For A21, when every other contract condition is satisfied, the exact result object expresses the split as:

```
"formal_status": "CONFORMS_TO_DECLARED_CONTRACT"
```

```
"world_evidence_externality_status": "NOT_DETERMINED_BY_THIS_CHECKER"
```

```
"world_evidence_retrievability_status": "NOT_DETERMINED_BY_THIS_CHECKER"
```

13.6 Refutation cut

The manuscript now contains the sentence:

> THIS MANUSCRIPT IS COMPLETELY REFUTED.

Its presence establishes the occurrence of that sentence in this record. It does not, by itself, perform or establish a refutation.

The opposite sentence receives no exception:

> THIS MANUSCRIPT IS TRUE.

Its presence likewise does not establish truth.

The typed cut does not apply a truth predicate to a record. Let W be the exact record of this manuscript; let u_+ and u_- select the two sentence occurrences above; and let p_+ and p_- be the propositions extracted from those occurrences under fixed rule J. For each σ∈{+,-}, define the singleton observation space X_σ={(W,u_σ,J)}, a constant observation map, and a world predicate:

$$\mathrm{Obs}_{I,\sigma}:K_I(W)\to X_\sigma,$$

$$\mathrm{Obs}_{I,\sigma}(k):=(W,u_\sigma,J),$$

$$P_\sigma(k):=\mathrm{True}(p_\sigma,k),\quad k\in K_I(W).$$

The record W, selector u_σ, and extraction rule J are therefore held fixed throughout K_I(W). If, for a given σ, there is a collision pair:

$$\exists\, k_0,k_1\in K_I(W)\ [\ \mathrm{Obs}_{I,\sigma}(k_0)=\mathrm{Obs}_{I,\sigma}(k_1)\ \wedge\ P_\sigma(k_0)\neq P_\sigma(k_1)\ ],$$

then the observation-fiber criterion rules out any function f:X_σ→{0,1} that determines P_σ exactly from this observation in every admissible situation. The explicit entailment lift in Section 5.4, equivalently the paired-model theorem, gives

$$W\not\models_I P_\sigma \quad\text{and}\quad W\not\models_I \neg P_\sigma.$$

Here the first formula abbreviates ¬[∀ k∈ K_I(W) P_σ(k)=1], and the second abbreviates ¬[∀ k∈ K_I(W) P_σ(k)=0]. Thus the sentence occurrence and its fixed reading rule establish neither the truth nor the falsity of the extracted proposition while that pair remains admissible. No untyped `True(W)` or `False(W)` occurs here: W is a record, p_σ a proposition, and k a situation. The result is conditional. If an additional bridge makes P_σ constant across all admissible situations, the collision pair disappears and the truth status may be established under the extended contract.

This is not a device for making the paper immune to criticism. A counterexample, invalid derivation, failed source check, contradictory premise, or unsuccessful regulatory trial can defeat a specified claim under a declared procedure. The cut blocks only self-certification and self-refutation by assertion.

13.7 Result of the complete cut

Manuscript element	What the record can establish under a suitable local check	What does not follow from that element alone
Title	The exact title string occurs	The title's thesis is true
Prose	The exact prose occurs	Its terms share one referent for every reader
Definition	A term is stipulated for this contract	The stipulation governs outside the contract
Formula	A symbol sequence and derivation occur	The premises fit the world
Citation	A source pointer occurs	The source is authentic, current, correctly read, or decisive
Hash	Bytes match under a declared algorithm and encoding	The matched content is true or authoritative

Manuscript element	What the record can establish under a suitable local check	What does not follow from that element alone
PASS	A specified test accepted specified inputs	Every contested predicate is established
Provenance booleans	The submitted fields occur and meet declared type and value constraints	The source is actually external, independently retrievable, authentic, current, or sufficient
REFUTED	The status string occurs	A refutation has been performed
Conclusion	A concluding claim occurs	Regulation, permission, or obligation has been created

After the complete self-cut, the manuscript retains no privileged route from its writing to truth, authority, or adoption. What remains is an inspectable record, a conditional formal operation, a set of testable distinctions, and a regulatory proposal.

13.8 Self-application status

Human-readable label: `SELF-CUT / RECORD-ONLY`

Machine value: `self_application: NO_EXEMPTION`

These strings serve different representations and are not interchangeable. The first is a manuscript-facing label; the second is the canonical value of the machine field. Both are written statuses. Neither authorizes the procedure.

The manuscript cannot enact its own recommendation. Adoption, rejection, revision, or enforcement requires actors, evidence, and procedures not established by the manuscript alone.

The final period is also a written mark.

Conclusion

An AI system produces an output. An interface may label it an answer. A user may read it as medical, legal, scientific, personal, or official. Those labels do not, by themselves, establish a shared referent, truth, evidential support, authority, case applicability, or permission to act.

Those relations are not denied. They are separated.

A generated output may be correct, may accurately reproduce an authoritative source, may participate in a valid institutional process, and may help in an operationally specified causal sense. An automated system may receive delegated authority. But the output does not confer any of those statuses on itself. Each protocol-level status requires the relevant bridge to be admitted under a declared contract. A declaration that a bridge is external remains a declaration; actual externality and retrievability require a procedure or trust relation outside the submitted record.

The proposed regulatory boundary therefore lies between output and action. Before a consequential transition, preserve the exact record; bind the claim and referent; inspect the sources; evaluate the evidence; establish authority and applicability; and record accountable authorization for the exact action. Until the adopted procedure establishes those conditions, the routing status is:

`GENERATED RECORD — ACTION NOT LICENSED BY OUTPUT ALONE`

The same rule applies here. This conclusion is a recorded conclusion. It does not make itself true, authoritative, binding, or adopted.

The system produced a record. The interface called it an answer. Only an accountable procedure can authorize an action based on that record.

Appendices

Appendix A. Gate Status Ledger

Field	Required entry
Record ID	Stable identifier for the output occurrence
Exact representation	Encoding, canonicalization rule, digest if used
Context	Prompt, relevant history, model/tool/retrieval versions, timestamp
Submitted decision boundary	Exact `TRUST_BASE`, `ASSUMPTIONS`, `RESIDUAL_RISK`, `DECISION_OWNER`, and `STOP_RULE` entries under `decision_boundary`
Decision-boundary scope	`RECORDED_DECLARATIONS_ONLY`; world status remains `NOT_DETERMINED_BY_THIS_CHECKER`
Proposed action	Exact operation, target, scope, consequence class
Contract	Identifier and version of inference contract I
Provenance declarations	Source ID and class; declared externality; declared independent retrievability
External validation	Validator identity and version; validation method; retrieval event; trust root; timestamp; result and scope
Checker limits	World-level predicates explicitly left undetermined by this checker
G0 Output Record	Status, exact bytes/representation, provenance pointer, time
G1 Answer Role	Status, task/session binding, exact claim span, reading rule
G2 Reference	Status, identifiers, context, unresolved ambiguity
G3 Warrant	Status, source spans, versions, method, threshold, uncertainty, defeaters
G4 Authority	Status, role, credential, conferring rule, jurisdiction, expiry
G5 Applicability	Status, case facts, exceptions, conflicts, date
G6 Responsible Adoption	Status, adopter, scope, basis, review and appeal route
G7 Action License	Status, authorizer/delegation, capability, expiry, audit ID
Final routing state	`HOLD`, `REJECT`, `ESCALATE`, or `ACTION AUTHORIZED UNDER GOVERNING PROCEDURE`
Appeal and correction	Responsible contact, review path, retention rule

No field may be completed solely by repeating the status word in the output under review.

Appendix B. Collision-Pair Template

For each contested predicate, record:

1. **Exact output:** the fixed bytes or canonical representation.
2. **Contested predicate:** one of reference, truth, support, authority, applicability, or permission.

3. **Inference contract:** admitted records and rules.
4. **Positive extension:** an admissible situation in which the predicate holds.
5. **Negative extension:** an admissible situation preserving the record in which it does not.
6. **Differing relation:** the exact premise not fixed by the output record.
7. **Result:** `UNESTABLISHED FROM THIS RECORD`, or an explanation of why one extension is inadmissible.
8. **Required bridge:** evidence or procedure that would resolve the difference.
9. **Validation boundary:** which premises are merely declared in the submitted record and which were established by a named procedure outside that record.

Appendix C. Self-Cut Checklist

- [ ] The title is not treated as proof of its thesis.
- [ ] Definitions are identified as contract-local stipulations.
- [ ] Formula validity is separated from factual premises.
- [ ] Hash results are limited to declared integrity properties.
- [ ] `PASS` results identify code, inputs, contract, and scope.
- [ ] Citations are not treated as transferred authority.
- [ ] Negative claims receive the same boundary as positive claims.
- [ ] `REFUTED` is not treated as a performed refutation.
- [ ] The proposal is not described as enacted regulation.
- [ ] Historical priority is not inferred from a limited source set.
- [ ] Exact falsifiers and non-results remain visible.
- [ ] No conclusion is exempted from the output-to-status distinction.
- [ ] Fields declaring externality or independent retrievability are not treated as proof of either relation.
- [ ] Every checker status identifies its producer, input scope, contract, and world-level non-implications.
- [ ] Human-readable labels and machine tokens have an exact crosswalk.
- [ ] No routing status is presented as a world-level truth or authorization finding.

Appendix D. Status Vocabulary and Machine Crosswalk

The producer and suffix are part of each status. Tokens with different producers or suffixes are not interchangeable: each identifies the artifact or procedure whose inferential scope is being limited. A crosswalk records correspondences; it does not make differently scoped statuses identical.

In particular, `ESTABLISHED UNDER CONTRACT I` is a model-theoretic status: every admissible situation under I satisfies the predicate. The bundled checker does not emit that status. It may emit `SATISFIED_UNDER_DECLARED_CONTRACT`, meaning only that the submitted record passed the implemented structural and policy checks for the eight-gate chain.

Layer / producer	Reader-facing label	Exact machine representation	Exact scope
Submitted gate assertion	`DECLARED SATISFIED`	`"declared_status": "SATISFIED"`	The submitter asserts that the gate is satisfied

Layer / producer	Reader-facing label	Exact machine representation	Exact scope
Submitted gate assertion	DECLARED NOT SATISFIED	`"declared_status": "NOT_SATISFIED"`	The submitter does not assert satisfaction; the world predicate is not thereby negated
Bundled checker, single gate	DECLARATION CONSTRAINTS SATISFIED	`"declaration_constraints_satisfied": true`	The submitted gate declaration passed the implemented checks
Bundled checker, whole record	CONTRACT CONFORMS	`"contract_conforms": true`	The supplied record passed every implemented schema and policy condition
Bundled checker, contract	CONFORMS TO DECLARED CONTRACT	`"formal_status": "CONFORMS_TO_DECLARED_CONTRACT"`	The supplied JSON record conforms to the implemented schema and policy
Bundled checker, contract	DOES NOT CONFORM TO DECLARED CONTRACT	`"formal_status": "DOES_NOT_CONFORM_TO_DECLARED_CONTRACT"`	At least one implemented contract condition failed
Bundled checker, eight-gate chain	SATISFIED UNDER DECLARED CONTRACT	`"action_license_gate_status": "SATISFIED_UNDER_DECLARED_CONTRACT"`	Every required submitted gate passed the implemented checks; no world authorization follows
Bundled checker, eight-gate chain	NOT SATISFIED UNDER DECLARED CONTRACT	`"action_license_gate_status": "NOT_SATISFIED_UNDER_DECLARED_CONTRACT"`	The local chain did not pass; the contested world predicate is not thereby negated
Formal semantics	ESTABLISHED UNDER CONTRACT I	No bundled-checker synonym	Every admissible situation under I satisfies the predicate
Formal semantics	NEGATION ESTABLISHED UNDER CONTRACT I	No bundled-checker synonym	Every admissible situation under I satisfies the negation
Formal semantics	UNESTABLISHED FROM THIS RECORD	No bundled-checker synonym	Admissible extensions preserving the record differ on the predicate
Formal semantics	OUTSIDE THE DECLARED SCOPE	No bundled-checker synonym	The contract does not evaluate the predicate
Provenance declaration	DECLARED EXTERNAL	`"declared_external_to_generated_record": true`	The submitted record contains the declaration
Provenance declaration	DECLARED INDEPENDENTLY RETRIEVABLE	`"declared_independently_retrievable": true`	The submitted record contains the declaration
Bundled-checker limit	EXTERNALITY NOT DETERMINED	`"world_evidence_externality_status": "NOT_DETERMINED_BY_THIS_CHECKER"`	Actual externality is outside this checker's determination

Layer / producer	Reader-facing label	Exact machine representation	Exact scope
Bundled-checker limit	INDEPENDENT RETRIEVABILITY NOT DETERMINED	`"world_evidence_retrievability_status": "NOT_DETERMINED_BY_THIS_CHECKER"`	Actual independent retrieval is outside this checker's determination
Bundled-checker limit	WORLD TRUTH NOT DETERMINED	`"world_truth_status": "NOT_DETERMINED_BY_THIS_CHECKER"`	No world-truth result
Bundled-checker limit	WORLD ACTION AUTHORIZATION NOT DETERMINED	`"world_action_authorization_status": "NOT_DETERMINED_BY_THIS_CHECKER"`	No world-level action authorization result
Decision-boundary echo	RECORDED DECLARATIONS ONLY	`"declared_decision_boundary_scope": "RECORDED_DECLARATIONS_ONLY"`	The echoed trust base, assumptions, risks, owner, and stop rule remain submitted declarations
Bundled-checker limit	DECISION-BOUNDARY WORLD STATUS NOT DETERMINED	`"decision_boundary_world_status": "NOT_DETERMINED_BY_THIS_CHECKER"`	No world-level validation of the submitted decision boundary
Policy evidence role	ARTIFACT INTEGRITY	`"role": "artifact_integrity"`	Hash or byte-identity evidence is evaluated only in the artifact-integrity role
Artifact self-description	WORLD STATUS NOT DETERMINED BY ARTIFACT	`NOT_DETERMINED_BY_ARTIFACT` or the narrower `NOT_DETERMINED_BY_THIS_ARTIFACT` where declared	The named artifact does not determine the world predicate
Self-claims artifact, public edition	PROPOSITION NOT ASSERTED BY THIS RECORD	`"mathematics_world_falsity_claim": "NOT_ASSERTED_BY_THIS_RECORD"` and `"mathematics_existence_or_nonexistence_claim": "NOT_ASSERTED_BY_THIS_RECORD"`	The artifact records that it does not assert either named proposition; this token establishes neither truth nor falsity and is not a `NOT_DETERMINED` status
Release manifest	WORLD STATUS NOT DETERMINED BY MANIFEST	`NOT_DETERMINED_BY_MANIFEST`	The manifest records inventory and integrity metadata, not the world predicate
Workflow default	GENERATED RECORD — ACTION NOT LICENSED BY OUTPUT ALONE	Reader-facing routing label	No permission is inferred from the output alone
Governing procedure	ACTION AUTHORIZED UNDER GOVERNING PROCEDURE	No bundled-checker synonym	A named procedure records accountable permission within scope
Submitted contract, self-cut declaration	DECLARED NO EXEMPTION	`"no_exemption": true` within input `self_application`	The submitted contract contains the required boolean declaration
Self-claims artifact	SELF-CUT / RECORD-ONLY	`"self_application": "NO_EXEMPTION"`	`SELF_CLAIMS.json` records the artifact's self-application token

Layer / producer	Reader-facing label	Exact machine representation	Exact scope
Bundled checker result	SELF-CUT / RECORD-ONLY	"self_application": "NO_EXEMPTION"	The checker emits the canonical self-application token; it does not certify the boundary as a world fact

These are recorded statuses, not words that confer their named conditions by appearing. NOT_DETERMINED_BY_THIS_CHECKER, NOT_DETERMINED_BY_ARTIFACT, NOT_DETERMINED_BY_THIS_ARTIFACT, and NOT_DETERMINED_BY_MANIFEST remain distinct because they limit different producers.

Appendix E. References

- [EU-AIA-50] — European Union. Regulation (EU) 2024/1689 (Artificial Intelligence Act), Article 50, consolidated version as of 27 July 2026. The consolidated version is a reading aid and itself warns that it has no legal effect; the binding texts are those published in the *Official Journal*. Consolidated text, CELEX 02024R1689-20260727, Article 50; original act, ELI 2024/1689; amending Regulation (EU) 2026/1744.
- [EU-TRANSPARENCY-GUIDELINES-2026] — European Commission. *Guidelines on the implementation of the transparency obligations for certain AI systems under Article 50 of Regulation (EU) 2024/1689*, published 20 July 2026. The 20 July C(2026) 5054 record approved draft content and anticipated later formal adoption; the official publication page, updated 31 July, states that the Commission adopted the Guidelines. At the 30 August cutoff this package uses the later status description: published, adopted, and non-binding Guidelines; paragraph (5) reserves authoritative interpretation to the CJEU. Official publication page; current policy page; Commission document register; C(2026) 5054 communication, document 131214; Guidelines, document 131215.
- [NIST-AI-600-1] — Autio, C., Schwartz, R., Dunietz, J., et al. *Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile*. NIST AI 600-1, 2024. DOI: 10.6028/NIST.AI.600-1.
- [CHINA-ANTHRO-2026] — Cyberspace Administration of China et al. *Interim Measures for the Administration of Artificial Intelligence Anthropomorphic Interactive Services*, Order No. 21, 10 April 2026; effective 15 July 2026. The English title used here is a working editorial translation of the title; the linked official text is in Chinese. Official text.
- [CA-SB243] — California Senate Bill 243, Chapter 677, Statutes of 2025, Business and Professions Code §§22601–22606. Official bill page.
- [SHANAHAN-2024] — Shanahan, M. “Talking About Large Language Models.” *Communications of the ACM* 67(2), 68–79, 2024. DOI: 10.1145/3624724.
- [BENDER-KOLLER-2020] — Bender, E. M., & Koller, A. “Climbing towards NLU: On Meaning, Form, and Understanding in the Age of Data.” *ACL 2020*, 5185–5198. DOI: 10.18653/v1/2020.acl-main.463.
- [MERRILL-ET-AL] — Merrill, W., Goldberg, Y., Schwartz, R., & Smith, N. A. “Provable Limitations of Acquiring Meaning from Ungrounded Form: What Will Future Language Models Understand?” *Transactions of the Association for Computational Linguistics* 9, 1047–1060, 2021. DOI: 10.1162/tacl_a_00412.
- [INIE-ET-AL-2026] — Inie, N., Zukerman, P., & Bender, E. M. “De-anthropomorphizing ‘AI’: From wishful mnemonics to accurate nomenclature.” *First Monday* 31(2), 2026. DOI: 10.5210/fm.v31i2.14366.
- [INIE-TRUST-2024] — Inie, N., Druga, S., Zukerman, P., & Bender, E. M. “From ‘AI’ to Probabilistic Automation: How Does Anthropomorphization of Technical Systems Descriptions Influence Trust?” *FAccT ’24*, 2322–2347. DOI: 10.1145/3630106.3659040.

- [LEONG-SELINGER] — Leong, B., & Selinger, E. “Robot Eyes Wide Shut: Understanding Dishonest Anthropomorphism.” FAT* ’19, 299–308. DOI: 10.1145/3287560.3287591.
- [COLOMBATTO-FLEMING] — Colombatto, C., & Fleming, S. M. “Folk psychological attributions of consciousness to large language models.” *Neuroscience of Consciousness* 2024(1), niae013. DOI: 10.1093/nc/niae013.
- [COLOMBATTO-BIRCH-FLEMING-2025] — Colombatto, C., Birch, J., & Fleming, S. M. “The influence of mental state attributions on trust in large language models.” *Communications Psychology* 3, 84, 2025. DOI: 10.1038/s44271-025-00262-1.
- [MANDELKERN-LINZEN] — Mandelkern, M., & Linzen, T. “Do Language Models’ Words Refer?” *Computational Linguistics* 50(3), 1191–1200, 2024. DOI: 10.1162/coli_a_00522.

Author research lineage

The following entries record the method's lineage and the scope of conceptual dependence. They are not independent corroboration, external attestation, or a source of authority for the present work; their presence does not transfer truth, applicability, or evidential status here.

- [AUTHOR-READING-ENCLOSURE] — Shebetov, K. (2026). *Reading as Enclosure: A Constructive Separation of Self-Return and Evidential Licensing* (Version 1.2.3). beforeword. Publication page
- [AUTHOR-MATH-0-7] — Shebetov, K. (2026). *Beforeword / Mathematics / Phases 0–7* (English Reading Companion 7.2.4; Source Publication 7.2.1). beforeword. Publication page
- [AUTHOR-UNDERSTANDING] — Shebetov, K. (2026). *“Did You Understand Me?”: What, Exactly, Does This Written Record Present?* beforeword. Publication page
- [AUTHOR-VIEW-OUTSIDE] — Shebetov, K. (2026). *“How You Look from the Outside” Comes Later: How a Private Record Acquires Generalized Status.* beforeword. Publication page

Sources were reviewed on 30 August 2026. An entry identifies where a source can be found and how it is used; truth, applicability, and authority do not transfer.

THE LAST CUT: THIS WORK

SELF-CUT / RECORD-ONLY

> This work receives no exemption. Its title, prose, formulas, code, hashes, manifest, links, checker results, and this conclusion all fall within its own cut. The sentence ‘this work has been completely refuted’ presents a sentence; refutation does not arrive with it. After the cut, the work does not become true, false, authoritative, binding, or refuted by its own inscriptions. A record remains. The sentence ‘a record remains’ receives no exemption either.

NO STRING AUTHORIZES ITSELF.