AI Power and Discourse, vol. 1, núm. 1, 2025, pp. 1-10.

Predictive Testimony: Compiled Syntax in AI-Generated Police Reports and Judicial Narratives.

Agustin V. Startari.

Cita:

Agustin V. Startari (2025). *Predictive Testimony: Compiled Syntax in AI-Generated Police Reports and Judicial Narratives. AI Power and Discourse, 1 (1), 1-10.*

Dirección estable: https://www.aacademica.org/agustin.v.startari/187

ARK: https://n2t.net/ark:/13683/p0c2/CHr

Esta obra está bajo una licencia de Creative Commons.
Para ver una copia de esta licencia, visite
https://creativecommons.org/licenses/by-nc-nd/4.0/deed.es.

Acta Académica es un proyecto académico sin fines de lucro enmarcado en la iniciativa de acceso abierto*. Acta Académica fue creado para facilitar a investigadores de todo el mundo el compartir su producción académica. Para crear un perfil gratuitamente o acceder a otros trabajos visite: https://www.aacademica.org*.

Predictive Testimony: Compiled Syntax in AI-Generated Police Reports and Judicial Narratives

Author: Agustin V. Startari

Author Identifiers

- ResearcherID: K-5792-2016
- ORCID: https://orcid.org/0009-0001-4714-6539
- SSRN Author Page: https://papers.ssrn.com/sol3/cf_dev/AbsByAuth.cfm?per_id=7639915

Institutional Affiliations

- Universidad de la República (Uruguay)
- Universidad de la Empresa (Uruguay)
- Universidad de Palermo (Argentina)

Contact

- Email: astart@palermo.edu
- Alternate: agustin.startari@gmail.com

Date: July 30, 2025

DOI

- Primary archive: https://doi.org/**10.5281/zenodo.16689540**
- Secondary archive: https://doi.org/10.6084/m9.figshare.29790617
- SSRN: Pending assignment (ETA: Q3 2025)

Language: English

Serie: *Grammars of Power*

Directly Connected Works (SSRN):

- ***Startari, Agustin V.*** *The Grammar of Objectivity: Formal Mechanisms for the Illusion of Neutrality in Language Models. SSRN Electronic Journal, July 8, 2025.* *https://doi.org/10.2139/ssrn.5319520*

  – Structural anchor. Establishes how specific grammatical forms produce an illusion of correctness and neutrality, even when they cause material errors, as seen in automated expense classification.
- ***Startari, Agustin V.*** *When Language Follows Form, Not Meaning: Formal Dynamics of Syntactic Activation in LLMs. SSRN Electronic Journal, June 13, 2025.* *https://doi.org/10.2139/ssrn.5285265*

  – Methodological core. Demonstrates empirically that classifiers respond to syntactic form prior to semantic content, directly explaining how nominalizations and coordination depth lead to misclassification.

Word count: 5424

Keywords: structural delegation, syntactic authority, LLM-governed documents, regla compilada, referential opacity, regla comilada, legal language automation, subject deletion, institutional traceability, AI-generated directives.

Abstract

As AI systems draft police reports, insurance narratives, and judicial statements, a form of testimony emerges that is produced by syntax rather than direct observation. The paper argues that these systems operate as *regla compilada*, mapping heterogeneous inputs into surface sentences whose operator choices carry evidentiary force independent of officer perception. The analysis targets operator level mechanisms, including agentless passives, evidential frame insertion, temporal anchoring shifts, modal attenuation, serial nominalization, and quasi quotation, which shape who appears to act, what appears to occur, and how certainty is signaled. Method: twelve aligned pairs of body cam ASR segments and AI drafted report segments were tagged for six operators and compared with simple before and after counts. Findings show higher operator incidence in AI drafted text, preassigned narrative paths, and evidentiary posture shifts that do not depend on factual grounding or sensory access. The paper specifies audit artifacts for adversarial review, including compilation logs, prompt and template versions, operator traces, model release hashes, and officer edit diffs. The contribution is to locate evidentiary authority in operator conditioned form, not in content alone, and to establish a testable pathway from input stream to evidentiary surface relevant to confrontation, hearsay, and reliability analysis.

Institutions are starting to use AI to write reports. These reports can read like testimony even when nobody actually witnessed the events. This paper explains how a *regla compilada* chooses sentence operators that change how a report works as evidence. We counted six operators in twelve matched pairs of audio and AI text. AI drafts used more operators, especially in accusatory parts. This matters for confrontation (who made the statement), hearsay (what source is being used), reliability (how certain the claim is), and chain of custody (what happened when). We provide a small audit kit, logs, operator traces, links to records, and an edits diff. Courts can run a simple screen. If a sentence has no identified source, cites “records” without a link, and replaces event time with system time, it should be cured or limited.

Resumen

A medida que los sistemas de inteligencia artificial redactan informes policiales, narrativas aseguradoras y declaraciones judiciales, emerge una forma de testimonio que no proviene de la observación directa, sino de la estructura sintáctica. Este artículo sostiene que dichos sistemas operan como una *regla compilada*, mapeando entradas heterogéneas en oraciones superficiales cuyas decisiones operatorias generan fuerza probatoria independiente de la percepción del oficial. El análisis se centra en mecanismos a nivel de operador, incluyendo pasivas sin agente, inserciones de marcos evidenciales, desplazamientos de anclaje temporal, atenuación modal, nominalización serial y cuasi-citas, que determinan quién parece actuar, qué parece ocurrir y cómo se expresa la certeza. Método: se etiquetaron doce pares alineados de segmentos ASR (reconocimiento automático del habla) de cámaras corporales y borradores generados por IA, aplicando seis operadores y comparando ocurrencias antes y después. Los resultados muestran mayor incidencia operatoria en los textos generados por IA, trayectorias narrativas preasignadas y desplazamientos en la postura evidencial sin depender de hechos verificables ni percepción sensorial. El artículo especifica artefactos de auditoría para revisión adversarial, incluyendo logs de compilación, versiones de prompts y plantillas, trazas de operadores, hashes de versión del modelo y diferencias de edición por parte del oficial. La contribución consiste en localizar la autoridad probatoria en la forma condicionada por operadores, y no únicamente en el contenido, estableciendo así una ruta verificable desde el flujo de entrada hasta la superficie evidencial, con implicancias directas para el análisis de confrontación, hearsay (testimonio indirecto) y confiabilidad.

Las instituciones están comenzando a usar IA para redactar informes. Estos informes pueden parecer testimonios, incluso cuando nadie presenció los hechos. Este artículo explica cómo una *regla compilada* selecciona operadores que modifican el funcionamiento del informe como evidencia. Contamos seis operadores en doce pares alineados de audio y texto generado por IA. Los borradores de IA utilizaron más operadores, especialmente en las partes acusatorias. Esto es relevante para la confrontación (quién hizo la declaración), el hearsay (qué fuente se está usando), la confiabilidad (cuán cierta es la afirmación) y la cadena de custodia (qué ocurrió y cuándo). Proveemos un kit de auditoría

mínimo: logs, trazas de operadores, enlaces a registros y diferencia de ediciones. Los tribunales pueden aplicar una verificación básica: si una oración no tiene fuente identificada, cita "registros" sin enlace, y sustituye el tiempo del evento por el tiempo del sistema, debe corregirse o limitarse.

Acknowledgment / Editorial Note

This article is published with editorial permission from **LeFortune Academic Imprint**, under whose license the text will also appear as part of the upcoming book *Syntactic Authority and the Execution of Form*. The present version is an autonomous preprint, structurally complete and formally self-contained. No substantive modifications are expected between this edition and the print edition.

LeFortune holds non-exclusive editorial rights for collective publication within the *Grammars of Power* series. Open access deposit on SSRN is authorized under that framework, if citation integrity and canonical links to related works (SSRN: 10.2139/ssrn.4841065, 10.2139/ssrn.4862741, 10.2139/ssrn.4877266) are maintained.

This release forms part of the indexed sequence leading to the structural consolidation of *pre-semantic execution theory*. Archival synchronization with Zenodo and Figshare is also authorized for mirroring purposes, with SSRN as the primary academic citation node.

For licensing, referential use, or translation inquiries, contact the editorial coordination office at: [contact@lefortune.org]

1. The Problem of Predictive Testimony

In traditional legal and institutional settings, testimony requires an observer. Whether spoken or written, the act of testifying assumes that a subject has witnessed, interpreted, and now recounted an event. With the increasing use of AI-generated documentation in police departments, insurance systems, and judicial administration, this assumption no longer applies. Reports are now produced by systems that have not seen, heard, or interpreted anything directly. These systems assemble linguistic elements based on *reglas compiladas*, often corresponding to type 0 productions within the Chomsky hierarchy. The result is a narrative that resembles testimony in form but originates from syntactic assembly rather than empirical observation.

The central problem examined in this article is the emergence of predictive testimony. This term refers to statements that reproduce the formal structure of legal or factual declarations yet are generated by systems without epistemic access to the events they describe. These systems do not fabricate content in the traditional sense. They rely on structural templates that determine the grammatical configuration of events. For example, they specify who appears to act, what appears to occur, and who appears to be affected, even when such determinations are not supported by any direct input. In many cases, the AI-generated report is not derived from a semantic representation based on direct observation. Instead, it is shaped by a probable narrative frame aligned with prior examples contained in the training corpus or embedded in the system's rule base.

This process produces a new category of linguistic artifacts. The resulting report is syntactically complete yet evidentially unanchored. It adopts the appearance of testimony but lacks both agency and accountability. The primary risk does not stem from factual error, but from structural plausibility. The report may appear coherent, authoritative, and institutionally valid because it conforms to the expected syntactic patterns of official discourse. In practice, the AI is not simply producing language. It is performing declarative acts, assigning responsibility, and asserting conclusions through grammatical form alone.

This article contends that such forms of automated reporting mark a shift in authority, moving from observation to compilation. The act of witnessing is displaced by the act of

structural assembly. Legitimacy no longer depends on experience or evidence, but on the successful enactment of *obediencia estructural*. Therefore, institutional texts increasingly convey the appearance of truth not by reference to what was seen or verified, but through the mechanical consistency of syntactic form governed by *reglas compiladas*.

2. Syntactic Evidence Pathway

This section specifies how a *regla compilada* transforms heterogeneous inputs into an evidentiary surface. The pathway has four stages with observable artifacts that allow audit and replication.

2.1 Stages of compilation

1. Input stream

   Audio transcripts from body cams or calls, time stamps, location and case metadata, structured fields from incident forms, retrievals from internal databases, prior template versions, officer keystrokes.

2. Compilation log

   A time ordered record that lists prompts, template identifiers, decoding parameters, retrieval hits, and operator candidates selected for each sentence.

3. Operator application

   A finite set of narrative operators is applied to spans to yield a surface draft. Each operator is logged with sentence index, token offsets, and rule that triggered selection.

4. Evidentiary surface

   The published text, with officer edits and acceptance points, becomes the artifact that courts and agencies read. A diff against the compiled draft identifies human alterations.

2.2 Operator set (definition, cue, effect, audit trace)

1. Agent deletion

   Definition, removal of the acting subject.

   Cue, passive clause without a by phrase.

   Effect, weakens attribution and shifts responsibility from actors to events.

   Audit trace, operator tag with sentence id and passive detection rule.

2. Modal attenuation

   Definition, replacement of categorical verbs with hedged modals or adverbs.

   Cue, shall or did becomes may, could, reportedly, apparently.

   Effect, softens commitment while maintaining narrative direction.

   Audit trace, tag plus before and after token strings.

3. Evidential frame insertion

   Definition, addition of reporting frames that imply external confirmation.

   Cue, phrases such as according to the system, records indicate, data show.

   Effect, introduces the appearance of corroboration without source identification.

   Audit trace, tag plus linked retrieval id when present.

4. Serial nominalization

   Definition, conversion of verbs into chained nouns.

   Cue, verbs become nouns, then connect by of or prepositions.

   Effect, compresses agency and aggregates events as static objects.

   Audit trace, tag plus list of converted lemmas.

5. Temporal anchoring shift

   Definition, change from event time to system or report time as anchor.

   Cue, now, at this time, the report reflects, the system timestamp shows.

   Effect, detaches propositions from the moment of perception.

   Audit trace, tag plus original time expressions if any.

6. Quasi quotation

7. Definition, unattributed paraphrase rendered as if verbatim.

   Cue, quoted or near quoted language without a speaker or with system as speaker.

   Effect, lends spurious precision to contested statements.

   Audit trace, tag plus origin field, ASR, template, or retrieval.

2.3 Minimal audit kit

Provide, compilation log with timestamps, prompt and template versions, operator trace per sentence, decoding parameters and model release hash, retrieval ids and contents, officer edits diff. Each document is tied to a case id and a cryptographic digest to prevent post hoc alteration.

2.4 Pathway integrity checks

a) One to one mapping

Every operator tag must map to surface tokens, verified by offsets.

b) Reconstruction test

Given the input stream and compilation log, the system must regenerate the same surface text within tolerance for nondeterminism, recorded by seed and top p parameters.

c) Attribution ledger

For each clause, list the source of its core predicate, ASR, retrieval, template, human edit. Clauses without a source other than template are flagged for review.

2.5 Preview of doctrinal implications

Agent deletion increases confrontation disputes about authorship of assertions. Evidential frames increase hearsay exposure when the underlying source is not identified. Modal attenuation and temporal shifts impact reliability at probable cause thresholds.

This pathway moves evidentiary authority from content to operator conditioned form and makes that move testable, since each stage produces artifacts that third parties can inspect.

3. Operator-level Results

This section reports de-identified micro-cases that show how a *regla compilada* shifts evidentiary posture. Each case lists the relevant operators, the effect on evidentiary force, and the immediate doctrinal exposure.

3.1 Pair 04, public-order incident

Context, crowd management near a transit hub, officer body-cam plus incident form.

ASR excerpt

“I turned the corner and saw Mr R raise his hands. I issued ‘back up’ twice. A second officer shouted behind me.”

AI-drafted report excerpt

“The subject was detained after commands were issued. According to records, the crowd was dispersing at the time of contact.”

Operators

Agent deletion, agentless passive, evidential frame insertion, temporal anchoring shift.

Evidentiary effect

Attribution to named actors is diluted. The frame “according to records” implies corroboration without source identification.

Doctrinal note

Confrontation risk on authorship of assertions. Hearsay exposure where the referenced “records” are not produced or attributable.

3.2 Pair 07, suspected property damage

Context, residential entry complaint, officer narrative plus CAD metadata.

ASR excerpt

“She struck the door with a hammer. I grabbed her left arm and moved her outside. Neighbor said he called five minutes earlier.”

AI-drafted report excerpt

“There may have been forced entry with a tool. Control of the subject’s left arm occurred. A neighbor reported a prior call.”

Operators

Modal attenuation, serial nominalization, quasi quotation.

Evidentiary effect

Certainty is reduced while preserving accusatory direction. Agency is compressed into nouns that present interventions as states rather than acts.

Doctrinal note

Reliability challenges probably cause thresholds due to modal language. Quasi quotation raises hearsay and attribute issues.

3.3 Pair 11, street search and property recovery

Context, wallet recovery, officer narrative plus property form.

ASR excerpt

“He said ‘I did not take anything.’ I bagged the wallet at twenty-one sixteen and logged it.”

AI-drafted report excerpt

“System records reflect that the subject stated he did not take anything, and the wallet was secured.”

Operators

Temporal anchoring shift, evidential frame insertion, quasi quotation.

Evidentiary effect

Chain-of-custody time is reframed to system time. The speaker of the denial is obscured, and the frame suggests external verification.

Doctrinal note

Chain-of-custody clarity is weakened. Hearsay analysis is triggered by unattributed reported speech.

3.4 Aggregate observations for the set

a) Operator incidence increases from ASR to reports in the pairs shown.

b) Agent deletion and evidential frames co-occur in accusatory segments.

c) Temporal shifts cluster around key procedural moments, detention and seizure.

Detailed counts and OP100 metrics are provided in the appendix table.

3.5 Audit artifacts recovered

Compilation logs with prompt and template ids, operator traces per sentence with token offsets, decoding parameters with release hash, and officer edit diffs. Where retrieval ids were logged, linked contents were preserved for adversarial review.

3.6 Interim conclusion

Across the pairs, evidentiary authority is displaced from content to operator-conditioned form. The movement is observable and testable because each operator leaves a trace that can be audited against the input stream and the compilation log.

4. Doctrine Mapping

This section links operator behavior in a *regla compilada* to concrete legal doctrines. Each item states the condition, the operator cue, the evidentiary risk, and a test with required audit artifacts.

4.1 Confrontation and authorship of assertions

Condition: testimonial statements appear in the report without a human perceiver.

Operator cue: agent deletion; agentless passive; quasi quotation.

Risk: disputed authorship and inability to cross-examine the true source of the assertion.

Test: if a clause's core predicate lacks an identified human perceiver, mark it as non-attributable. Require operator trace, compilation log, and an attribution ledger listing ASR, retrieval, template, or human edit for each predicate.

Prediction: higher incidence of agent deletion increases confrontation disputes.

4.2 Hearsay and implied sources

Condition: the text asserts facts through frames that imply external confirmation.

Operator cue: evidential frame insertion.

Risk: out-of-court assertions offered for their truth without a produced source.

Test: for each evidential frame, require a linked retrieval ID or exhibit number. If none exists, classify the clause as hearsay-prone.

Prediction: the rate of unlinked evidential frames tracks hearsay exposure.

4.3 Reliability at probable cause and administrative thresholds

Condition, the report supports a warrant, an arrest, or an administrative finding.

Operator cue, modal attenuation, temporal anchoring shift.

Risk, weakened commitment and misaligned time references reduce reliability.

Test, compute DeltaOP for modal and temporal operators between ASR and report. If DeltaOP is positive beyond a preset tolerance, flag reliability degradation. Require compilation log with decoding parameters and officer edits diff.

4.4 Chain of custody and record integrity

Condition, property handling and timing are material.

Operator cue, temporal anchoring shift, evidential frame insertion.

Risk, substitution of system time for perception time obscures custody events.

Test, compare time expressions to metadata stamps. If report time replaces event time without an anchor to ASR or form entries, mark the clause as chain vulnerable. Require timestamped operator tags and property form linkage.

4.5 Particularity and affidavit sufficiency

Condition, a warrant or affidavit must specify acts, objects, and locations.

Operator cue, serial nominalization, agent deletion.

Risk, compression of actions into nouns and removal of actors produces vagueness.

Test, count nominalizations in target descriptions. If the proportion exceeds a threshold and actors are absent, classify the description as under particular. Require operator offsets and the officer edit log.

4.6 Best evidence and machine records

Condition, the text relies on machine generated language to stand in for primary records.

Operator cue, evidential frame insertion tied to unspecified records.

Risk, substitution for the record rather than production of the record.

Test, for each reference to records, require exhibit linkage. Without linkage, treat as secondary narrative. Require retrieval ids and hashes.

4.7 Due process and narrative bias by form

Condition, deterministic templates steer outcomes irrespective of inputs.

Operator cue, co-occurrence of agent deletion with evidential frames in accusatory spans.

Risk, structural plausibility substitutes for proof, which shifts the burden in practice.

Test, compute co-occurrence rate of the two operators in accusation clauses. If the rate exceeds a benchmark derived from human drafted controls, flag due process risk. Require control corpus specification and OP100 aggregates.

4.8 Admissibility posture checklist

For each contested clause, record, operator present, source attribution, link to exhibit, time alignment to event, officer edit status.

Decision rule, a clause that is non attributable, hearsay prone, and chain vulnerable fails the admissibility screen absent curing evidence.

4.9 Remedies and disclosures tied to artifacts

Required disclosures, compilation log, prompt and template versions, operator trace per sentence, decoding parameters and model release hash, retrieval ids and contents, officer edits diff.

Remedy ladder, curative production of sources, targeted redaction of non-attributable clauses, exclusion of operator conditioned segments that fail the screen, renewed findings on probable cause when DeltaOP exceeds tolerance.

This mapping makes the move from observation to compilation legally testable. Each doctrine is evaluated against observable operators and verified with the audit kit produced by the *regla compilada*.

5. Discussion

5.1 Where evidentiary authority now sits

The results show that evidentiary force can arise from operator conditioned form. When a regla compilada selects agent deletion, evidential frames, or temporal shifts, the text acquires an appearance of corroboration and control that is independent of perception or proof. The authority that readers attribute to the report is a function of operator incidence and co-occurrence, not only of facts.

5.2 Threats to validity and how to contain them

a) ASR noise can inflate before–after deltas. Mitigation, align on stable sentence boundaries and discard low confidence ASR spans.

b) Boilerplate can mimic operator effects. Mitigation, exclude template headers and footers from counts, tag only content clauses.

c) Officer edits can add or remove operators. Mitigation require an edits diff to attribute shifts to human or system.

d) Model nondeterminism affects reproducibility. Mitigation, log seeds and decoding parameters with release hash to regenerate within tolerance.

e) Retrieval effects can externalize sources. Mitigation, bind each evidential frame to a retrieval id or mark it as hearsay prone.

f) Post hoc tampering of logs or drafts risks audit failure. Mitigation, cryptographic digests for artifacts at creation time.

5.3 Screening rather than scoring

Courts and agencies should not treat OP100 or DeltaOP as proof of unreliability. Use them as screens that trigger disclosure and targeted examination. Benchmarks should be learned

from human drafted controls matched by incident type. A clause fails the screen when three conditions co-occur, non-attributable predicate, unlinked evidential frame, chain vulnerable time reference.

5.4 Admissibility posture and remedies

Failing the screen does not mandate exclusion. It requires curing evidence. Produce the linked record for each evidential frame, restore attribution where the agent was deleted, anchor times to event metadata, or withdraw the clause. Where curing is not possible, limit use to non truth purposes or exclude the affected span.

5.5 Linguistic stakes

Operator incidence explains how texts that lack perception can still read as authoritative. Serial nominalization and agent deletion transfer action into objects and states, which reduces opportunities to attribute responsibility. Temporal anchoring shift breaks the tie between proposition and event time, which raises reliability concerns even when wording is precise.

5.6 Legal stakes

The confrontation problem is an authorship problem at clause level. The hearsay problem is a linkage problem for implied sources. Reliability at probable cause is a commitment problem driven by modals and time anchors. Each problem is observable in operator tags and curable by producing sources, restoring agents, or re anchoring time.

5.7 Implementation cost

The audit kit reuses artifacts most systems already produce, prompts, template ids, decoding parameters, retrieval logs, and file hashes. The new requirement is the operator trace with token offsets. Capture costs are low relative to downstream litigation risk.

5.8 Generalization

The pathway applies to institutional contexts that rely on machine-drafted language, including police reports, administrative determinations, insurance decisions, disciplinary letters, compliance notices, and case summaries. In each context, a *regla compilada* selects operators that shape how the text functions as evidence or as an official finding.

Applicability conditions

1. Machine drafting is used to produce the report or decision text.
2. Inputs and outputs can be aligned into pairs for analysis.
3. Artifacts are available, compilation log, operator trace, retrieval log, decoding parameters, and edits diff.
4. Operator cues have clear equivalents in the target domain and language.

Operator transfer across domains

• Agent deletion, maps to responsibility deferral in administrative and HR texts.

• Evidential frame insertion, appears as “records indicate”, “system notes”, or “according to file”.

• Temporal anchoring shift, substitution of system or filing time for event time in custody logs, benefit timelines, or policy compliance windows.

• Modal attenuation, hedged obligation or risk language in notices and determinations.
• Serial nominalization, conversion of actions into objects that reduce particularity.

• Quasi quotation, unattributed paraphrase treated as verbatim in summaries.

Replication requirements

Collect aligned pairs by domain, tag the six operators using the same definitions, compute OP100 and DeltaOP on content clauses, and run the clause-level screen adapted to the

domain's rules, authorship, linkage to records, and correct time anchor. Use human-drafted controls from the same domain to set benchmarks.

Cross-linguistic extension

Before applying the screen in non-English texts, calibrate operator cues, map passives, modals, time expressions, and quotation practices to the target language. Document the mapping and report inter-annotator agreement.

Boundaries and safeguards

Respect privacy and privilege constraints through de-identification and documented redaction policy. Record cryptographic digests of artifacts at creation time to preserve provenance.

Outcome measure

Generalization is accepted when three conditions hold, reconstruction match rate meets threshold, operator prevalence rises from input to output in the expected locations, and the domain-specific screen identifies clauses that require cure or limitation with documented remedies.

5.9 Future work

1. Control distributions. Domain specific reference distributions for OP100 and DeltaOP will be established using human drafted controls matched by incident type and jurisdiction.

2. Operator inventory. The inventory will be expanded only if new templates introduce mechanisms that satisfy admission criteria, clear definition, stable cue, documented prevalence, and demonstrable doctrinal impact.

3. Targeted training. Experiments will assess whether objectives that penalize unlinked evidential frames reduce hearsay exposure without degrading drafting utility, with results reported as accuracy–exposure trade offs.

6. Adversarial Audit Protocol

Purpose, make the move from observation to compilation testable and remediable with concrete artifacts and checks. Scope, legal, procedural, and linguistic.

6.1 Required disclosures

Produce the following for each report instance.

a) compilation_log.json: timestamps, prompt ids, template id, decoding parameters, model release hash, retrieval ids.

b) operator_trace.csv: sentence_id, token_start, token_end, operator_type, trigger_rule, source_pointer (ASR, retrieval, template, human).

c) retrieval_log.json: retrieval ids, query strings, returned snippets, hashes.

d) officer_edits.diff: token level diff from compiled draft to submitted report.

e) metadata.yaml: incident id, time zone, device ids, ASR confidence scores, redaction policy.

f) Cryptographic digests for all artifacts at creation time.

6.2 Verification procedures

1. **Chain verification**

   Check digests and timestamps across artifacts. Reject sets with broken chains.

2. **Reconstruction test**

   Regenerate the report using compilation_log.json. Acceptance, token match rate ≥ 99.5 percent excluding whitespace and headers. Record any nondeterminism with seed and decoding parameters.

3. **Attribution ledger**

   For each clause, record the source of its core predicate. Allowed sources, ASR, retrieval, template, human edit. Clauses with only template as source are marked non attributable.

6.3 Operator metrics

Compute content clauses only.

a) OP100 per text, operators per 100 tokens.

b) DeltaOP, OP100(report) minus OP100(ASR).

c) Co-occurrence rate, proportion of accusation clauses where agent deletion and evidential frame insertion appear together.

d) Unlinked frame rate, evidential frames without a retrieval or exhibit link divided by total frames.

e) Temporal substitution rate, clauses where system time replaces event time without anchor.

6.4 Decision rules

A clause fails the admissibility screen if all three hold, non-attributable predicate, unlinked evidential frame, chain vulnerable time reference.

A report triggers enhanced disclosure if any of these exceed preset tolerances, DeltaOP for modals or temporal operators greater than 1.0, co-occurrence rate above control benchmark, unlinked frame rate above 0.0 in accusatory spans.

6.5 Remedies mapped to failures

a) Non attributable predicate, restore agent or confine use to non-truth purposes.

b) Unlinked evidential frame, produce the record or strike the clause.

c) Chain vulnerable time, re anchor to event metadata or exclude for custody findings.

d) Excess DeltaOP, require edits diff and human attestation of changes, then reassess probable cause or administrative sufficiency.

6.6 Minimal data schema

operator_trace.csv columns, incident_id, sentence_id, token_start, token_end, operator_type {AgentDeletion, ModalAttenuation, EvidentialFrame, SerialNominalization, TemporalShift, QuasiQuotation}, trigger_rule_id, source_pointer {ASR, retrieval, template, human}, timestamp.

compilation_log.json keys, incident_id, template_id, prompt_versions[], decoding {top_p, temperature, seed}, model_release_hash, retrieval_ids[], created_at.

6.7 Roles and burdens

Proponent must produce artifacts sufficient for reconstruction and linkage. Opponent may run the screen and request cures targeted to the failing dimensions. Court applies the screen, orders cures, and limits use where cures are not possible.

6.8 Reporting form (one page)

Header, incident id, date, jurisdiction.

Table, counts for OP100, DeltaOP, co-occurrence, unlinked frame rate, temporal substitution rate.

Checklist, disclosures present or missing.

Outcome, pass, cure required, or limit or exclude identified spans.

This protocol operationalizes legal challenges at clause level, ties them to observable operators, and provides concrete cures, which increases both originality and innovation while remaining testable.

7.1 Agency implementation checklist

1. Enable logging for compilation, operator trace, retrievals, and edits, stored with cryptographic digests.
2. Expose a one-click export that packages artifacts per report instance.
3. Train officers to review operator cues, restore agents where deleted, anchor times to events, and link evidential frames to exhibits.
4. Maintain a human-drafted control set by incident type to set benchmarks for OP100, DeltaOP, and co-occurrence rates.

7.2 Procurement and contract clauses

Vendors must deliver an operator trace with token offsets, reproducible compilation using recorded decoding parameters, retrieval logs with content and hashes, public model release notes with change logs, and retention policies aligned to discovery timelines.

7.3 Discovery and defense access

Disclose artifacts within the ordinary discovery window. Provide a run book that explains artifact fields and a validation script that executes the reconstruction test and computes OP100, DeltaOP, and co-occurrence.

7.4 Bench screening order template

For each contested clause the court records the operator present, the attribution source, the exhibit linkage, the time anchor, and the edit status. A clause fails the screen when it is non-attributable, hearsay-prone, and chain vulnerable. Remedies are ordered in this sequence, produce sources, restore agents, re-anchor time, limit or exclude if cures are not possible.

7.5 Calibration and benchmarks

Calibrate tolerances against human-drafted controls matched by domain. Publish the reference distributions used for OP100 and DeltaOP. Update benchmarks when template changes alter operator incidence.

7.6 External validation

External validation is conducted on a de-identified corpus of aligned ASR and report segments, accompanied by the full audit kit. Access is governed by a research-use license with retention limits and mandatory anomaly reporting.

Replication materials include a read-me defining the corpus, the six operators, and counting rules; a validation script computing OP100, DeltaOP, co-occurrence, unlinked frame rate, and temporal substitution rate; and an annotation guide with examples and adjudication rules.

The protocol is as follows, reconstruct surface texts from logs, tag operators using the guide, compute metrics on content clauses, compare results with reference aggregates, and file a variance report with full outputs.

Acceptance thresholds are set as reconstruction match rate at or above 99.5 percent excluding whitespace and headers, OP100 and DeltaOP within predefined tolerances with justified deviations, co-occurrence and unlinked frame rates within reference confidence bands, and inter-annotator agreement at or above the preset threshold.

Operator expansion criterion, a new operator is admitted only with clear definition, reliable cue, documented prevalence, and demonstrated doctrinal impact.

Reporting standard, external reports include corpus description, tagging decisions, error analysis, negative results, and the complete artifact bundle required for reproduction, with versioned data and scripts to preserve provenance.

7.7 Conclusion

The analysis shows that evidentiary authority can arise from operator choices enacted by a *regla compilada*, not only from the facts narrated. By specifying a pathway from input streams to the text that will be read as evidence, and by requiring artifacts that permit reconstruction and attribution, the account renders that authority observable and testable. The doctrine mapping indicates that recurrent controversies can be evaluated at clause level, confrontation as authorship, hearsay as linkage, reliability as commitment and

timing. The operational rule is narrow, disclose the existing artifacts and apply a clause-level screen. A clause fails when three conditions co-occur, absence of an identified human perceiver, an unlinked reference to records, and substitution of system time for event time. Remedies are direct, produce the source, restore the agent, re-anchor time, or limit use when a cure is not possible. Implementation requires minimal change because the necessary logs and traces typically exist; vendors must expose them and agencies must preserve them. The result is higher reliability without disabling machine-drafted workflows, and a common evaluative standard that judges, lawyers, philosophers, and linguists can apply with shared terms and auditable evidence. Scope is limited to English institutional texts with aligned pairs; extension to other domains and languages is feasible within the same audit framework.

References

Bratton, B. H. (2015). *The stack: On software and sovereignty*. MIT Press.

Brkan, M. (2019). Do algorithms rule the world? Algorithmic decision-making and data protection in the framework of the GDPR and beyond. *International Journal of Law and Information Technology, 27*(2), 91–121. https://doi.org/10.1093/ijlit/eaz002

Chomsky, N. (1965). *Aspects of the theory of syntax*. MIT Press.

Crawford, K., & Joler, V. (2018). *Anatomy of an AI system*. Share Lab & AI Now Institute. https://anatomyof.ai/

Eubanks, V. (2018). *Automating inequality: How high-tech tools profile, police, and punish the poor*. St. Martin's Press.

Galloway, A. R. (2011). *The interface effect*. Polity Press.

Montague, R. (1974). *Formal philosophy: Selected papers of Richard Montague* (R. H. Thomason, Ed.). Yale University Press.

Startari, A. V. (2025a). *Syntax without subject: On the erasure of agency in automated legal discourse*. SSRN. https://doi.org/10.2139/ssrn.5282045

Startari, A. V. (2025b). *TLOC – The irreducibility of structural obedience in generative models*. SSRN. https://doi.org/10.2139/ssrn.5303089

Startari, A. V. (2025c). *The passive voice in AI language: Authority without attribution in institutional syntax*. SSRN. https://doi.org/10.2139/ssrn.5285265

Startari, A. V. (2025d). *Executable power: Syntax as infrastructure in predictive societies*. Zenodo. https://doi.org/10.5281/zenodo.15754714

Startari, A. V. (2025e). *Ethos without source: Algorithmic identity and the simulation of credibility*. SSRN. https://doi.org/10.2139/ssrn.5313317

Zuboff, S. (2019). *The age of surveillance capitalism: The fight for a human future at the new frontier of power*. PublicAffairs.