

# Syntax Without Subject: Structural Delegation and the Disappearance of Political Agency in LLM-Governed Contexts

---

**Author:** Agustin V. Startari

## Author Identifiers

- ResearcherID: NGR-2476-2025
- ORCID: <https://orcid.org/0009-0001-4714-6539>
- SSRN Author Page:  
[https://papers.ssrn.com/sol3/cf\\_dev/AbsByAuth.cfm?per\\_id=7639915](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](mailto:astart@palermo.edu)
- Alternate: [agustin.startari@gmail.com](mailto:agustin.startari@gmail.com)

**Date:** July 30, 2025

## DOI

- Primary archive: <https://doi.org/10.5281/zenodo.16571077>
- Secondary archive: <https://doi.org/10.6084/m9.figshare.29665697>
- 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.
- **Whitepaper Syntactics: Persuasive Grammar in AI-Generated Crypto Offerings**  
– Applied parallel. Although centered on crypto-finance, this study shares a syntactic lens on financial automation, showing how persuasive grammar shapes decisions. It offers a comparative foundation for extending the fair-syntax model to other algorithmic audit contexts. <https://doi.org/10.5281/zenodo.15962491>

### Word count:

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

## Abstract

This article examines the syntactic disappearance of the subject in LLM-governed documents. Structural delegation refers to the transfer of agency to impersonal grammatical forms that preclude subject reappearance. Subjects are not censored but syntactically eliminated through passive constructions, nominalizations, and imperative prompt formats with suppressed agents. Building on prior work on synthetic ethos and impersonal command grammars, the article shows that AI-generated institutional texts display consistent patterns of subject erasure. The study analyzes 172 documents produced by GPT-4 class models (temperature 0.2–0.7, 2024–2025) across legal, healthcare, and administrative domains. Metrics include passive ratio (via dependency label parsing), nominalization density (via POS and suffix filters), and instruction-format frequency. The result is a form of executable authority grounded not in referential authorship but in compliance with a *regla compilada* (type-0 production). The study proposes a typology of structural delegation and a formal framework for detecting syntactic absence in automated governance.

## Resumen

Este artículo examina la desaparición sintáctica del sujeto en documentos generados por modelos de lenguaje de gran escala (LLMs). Delegación estructural se define como la transferencia de agencia a formas gramaticales impersonales que impiden la reaparición del sujeto. No se censura a los agentes, sino que se los elimina mediante construcciones pasivas, nominalizaciones y formatos imperativos de instrucción con agente suprimido. Basado en trabajos previos sobre ethos sintético y gramáticas de mando impersonales, el artículo demuestra que los textos institucionales generados por IA presentan patrones consistentes de borrado del sujeto. El estudio analiza 172 documentos producidos por modelos de clase GPT-4 (temperatura 0.2–0.7, años 2024–2025) en los sectores legal, sanitario y administrativo. Las métricas incluyen proporción de pasivas (vía etiquetas de dependencia), densidad de nominalización (a través de filtros de sufijo y categoría gramatical), y frecuencia de formatos de instrucción. El resultado es una forma de autoridad

ejecutable basada no en autoría referencial, sino en la adhesión a una regla compilada (producción tipo 0). El estudio propone una tipología de la delegación estructural y un marco formal para detectar la ausencia sintáctica en entornos de gobernanza automatizada.

### **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. Structural Delegation and the Disappearance of the Subject

The integration of large language models (LLMs) into institutional writing workflows, including policy drafting, administrative guidelines and legal contracts, has introduced a new form of grammatical authority. In this regime, legitimacy is derived from structural compliance rather than referential authorship. Institutional agency does not vanish through rhetorical omission, but through a systematic syntactic transformation. This disappearance is not semantic, nor merely stylistic. It is structural.

We define structural delegation as the transfer of discursive agency from a referential subject to a grammatical mechanism that enacts operations without restoring the subject position. The operative set consists of passive constructions, nominalizations, and imperative templates with elided agents. These patterns recur in LLM-governed documents as syntactic defaults, not as stylistic variants.

Building on previous work on synthetic ethos and impersonal command grammars (Startari, 2025a; 2025b), we demonstrate that LLMs simulate authority through adherence to reglas compiladas. A regla compilada is defined here as a type-0 production in the Chomsky hierarchy. It refers to a generative rule that permits unrestricted rewriting and enables full syntactic execution without referential constraints (Chomsky, 1965; Montague, 1974). Within such grammar, authorship is not represented, only the output sequence matters.

This structural logic has direct consequences in legal writing. When subjects are syntactically removed, the distribution of responsibility, duty or liability becomes opaque. Directives may be issued, conditions imposed, and obligations formalized without any traceable source or signatory.

Across legal, medical and bureaucratic documents, this grammatical erasure of agency defines a new operational language. It functions without speakers, commands without issuers, and legitimizes without attribution. Structural delegation enables a form of executable language where syntax substitutes for political presence.

The syntactic patterns described here are examined in detail in Section 2, which outlines the corpus construction, metric selection and analytic method.

## **2. Corpus Construction and Syntactic Metrics**

To examine how structural delegation manifests in LLM-governed documents, we constructed a representative corpus of 172 texts generated between January 2024 and June 2025. All outputs were produced by models of the GPT-4 class, with temperature values ranging from 0.2 to 0.7, using English as the primary output language. Each document corresponds to a real-world institutional template or prompt, including contracts, privacy policies, regulatory guidelines, internal protocols, and terms of service. No outputs were edited by human reviewers after generation. The corpus was stratified across three domains: legal (n=64), healthcare (n=53), and administrative governance (n=55).

The central hypothesis guiding this study is that the syntactic erasure of the subject is not random or stylistic, but structurally encoded and quantifiable. To test this, we defined and operationalized three core syntactic indicators:

### **1. Passive ratio**

Measured as the proportion of clauses in each document whose main verb is syntactically marked as passive. Detection was based on dependency parsing using the `nsubjpass` label in the spaCy library, with manual validation on a 15% random subsample.

### **2. Nominalization density**

Computed as the number of nominalized verbs per 1,000 tokens. Identification combined suffix-based filters (-tion, -ment, -ance, -ency, etc.) with POS tagging (NOUN category) and exclusion of root nouns. Cases were cross-checked using context windows to eliminate false positives.

### 3. Instruction template frequency

Defined as the number of imperative-form clauses lacking explicit agents. Imperatives were identified using second-person null-subject constructions and modal-free command structures. Sentences beginning with a base verb without subject or auxiliary were automatically flagged, and a subset was manually reviewed for precision.

Thresholds for statistical relevance were set at  $\pm 2$  standard deviations from domain means, and outlier detection was performed to verify whether extreme cases correlated with prompt type, model version, or temperature setting. Full metric distributions and statistical tables are provided in Appendix A.

This methodological framework enables the tracing of syntactic absence not only as a stylistic feature, but as a programmable effect of LLM instruction-following behavior. The findings derived from these metrics are analyzed in Section 3, with examples drawn from the corpus and categorized by structural configuration.

### 3. Typology of Structural Delegation: Patterns and Examples

The metrics established in Section 2 reveal consistent grammatical operations through which LLM-governed documents eliminate subject positions. These operations are not isolated or idiosyncratic, but form a reproducible syntax of delegation. We identify three dominant structural categories: passive displacement, action nominalization, and instruction templates with elided agents. All examples cited are drawn from anonymized corpus materials, unless noted as pattern constructions.

#### 3.1 Passive displacement

In the legal subcorpus ( $n = 64$ ), passive constructions appear in 78.1 % of directive clauses (95 % CI: 74.9–81.2 %), with a passive-to-active ratio greater than 2.3:1 in 52 of 64 documents. For example:

“The following data shall be retained for regulatory purposes.”

*(legal, data policy, 2025, tkn 240–280)*

This construction omits the responsible party and syntactically isolates the action. It indicates a shift in the encoding of authority, where obligation is transmitted without source assignment. We exclude from this metric all locative statives (e.g., “is located”), get-passives, and checklist-style headings. See Table A.1 for distribution and IQR by sector.

### 3.2 Action nominalization

Nominalized verbs per 1,000 tokens averaged 41.7 in the healthcare subcorpus ( $n = 53$ ), with a 95 % CI of 39.1–44.3, measured using suffix and POS filters described in Section 2. For comparison, the domain mean across legal and administrative texts was 19.8 (CI: 18.3–21.3). For example:

“Failure to comply with the submission requirement may result in termination.”

*(healthcare, compliance form, 2024, tkn 1120–1165)*

Here, “submission” replaces a dynamic verbal form (“submit”) and detaches the action from both actor and temporal anchoring. The result is an abstract obligation without grammatical subject. All root nouns were excluded from nominalization counts. See Table A.3 for breakdown by suffix class and domain.

### 3.3 Instruction templates with elided agents

In the administrative subcorpus ( $n = 55$ ), 61.9 % of clauses flagged as imperative in surface form lacked explicit or recoverable agents (CI: 58.0–65.8 %). For example:

“Ensure that all credentials are verified before access is granted.”

*(administrative, IT protocol, 2025, tkn 610–645)*

The directive operates structurally as an instruction, but without any grammatical or contextual subject. This pattern was common in onboarding guides, IT compliance workflows, and internal governance templates. We excluded modal-deontic constructions



with “shall” or “must” from this category, as well as bullet-list directives and meta-linguistic headings. See Table A.4 for frequency by document type.

Each of these patterns reflects a specific mechanism by which *structural delegation* suppresses the referential subject. Taken together, they instantiate a grammar of executable authority that displaces responsibility while maintaining formal validity. These syntactic categories were validated using spaCy v3.7 dependency parser families, with manual annotation on 15 % of the corpus achieving inter-annotator agreement  $\kappa = 0.91$  (Cohen). For full methodology, see Section 2 and Appendix B.

#### 4. Institutional Risks and the Legal Consequences of Syntactic Erasure

The disappearance of the grammatical subject in LLM-governed documents is not merely a linguistic phenomenon. It alters the fundamental mechanisms by which responsibility, authorship and liability are articulated. When obligations are expressed without agents, and when directives are issued without explicit signatories, the result is not neutral. It is legal indeterminacy.

In legal drafting, the subject performs a constitutive role. It anchors enforceable duties and defines jurisdictional scope. A clause such as “The organization must notify affected parties” assigns responsibility clearly. By contrast, a syntactically equivalent passive such as

“Affected parties must be notified” removes the actor from the structure.  
(*legal, regulatory notice, 2025, tkn 870–905*)

This omission complicates enforcement. In many jurisdictions, the absence of a determinable agent may weaken the clause’s legal effect or introduce ambiguity during adjudication.

Nominalizations increase this exposure. By converting actions into abstract states or procedural stages, they eliminate the actor from both syntax and inference. For example:

“Upon completion of the verification, data access will be restored.”

*(administrative, internal protocol, 2024, tkn 430–470)*

This construction fails to identify who performs the verification. In the event of a procedural failure or dispute, no party can be directly held accountable by grammatical reference. The directive functions formally but lacks assignable responsibility. Such patterns produce what may be classified as non-attributive clauses. They are syntactically well-formed yet pragmatically underdetermined.

Instruction templates with elided agents create parallel risks. Consider:

“Confirm identity documentation prior to approval.”

*(healthcare, onboarding guide, 2025, tkn 1250–1280)*

Although the command is clear in intention, its executant is unspecified. In contexts where procedural accountability is mandatory—such as in medical recordkeeping or data governance—this omission may result in audit failure or procedural invalidation. In regulatory frameworks that require explicit delegation of duties, such clauses can obstruct traceability.

At the institutional level, LLM-generated language that systematically removes referential subjects introduces structural opacity. Responsibility is redistributed across the document without syntactic indication. The result is a regime in which execution persists, but authorship dissolves. Authority is becoming operational, yet unlocatable.

The next section examines the formal grammar that enables this transformation. It focuses on *reglas compiladas* as the generative basis through which LLMs produce structurally binding outputs without requiring an originating subject.

## **5. Reglas compiladas as Syntactic Infrastructure**

The patterns analyzed in previous sections are not byproducts of randomness or statistical noise. They emerge from a defined grammatical infrastructure embedded in the architecture of large language models. This infrastructure does not operate through

semantic reference or discursive intention. It functions through executable form. We refer to this structure as a *regla compilada*, defined as a type-0 production in the Chomsky hierarchy. Such rules enable unrestricted rewriting and allow for maximal generative capacity without requiring subject or referent (Chomsky, 1965; Montague, 1974).

In LLMs, generation is regulated by token probability distributions constrained by internal weightings and structural priors. When prompted with a legal, policy, or administrative input, the model activates structural patterns that prioritize syntactic closure over enunciative anchoring. The result is text that is grammatically coherent and institutionally familiar, but systematically devoid of source attribution.

A *regla compilada* does not encode the presence of a speaker. It enforces formal consistency between syntactic elements. For instance:

“Credentials must be renewed annually.”

*(legal, access policy, 2025, tkn 310–340)*

This clause appears authoritative and complete, yet it lacks any agent or issuing body. The model selects this form not because of authorial intent, but because it satisfies learned patterns of legal tone and procedural regularity.

This structure is not limited to isolated examples. Once the pattern is activated, it can generate grammatically uniform clauses across domains:

“All deviations shall be documented prior to audit.”

*(administrative, compliance guide, 2024, tkn 1500–1530)*

“Intake records must be archived according to retention schedule.”

*(healthcare, archival policy, 2025, tkn 890–920)*

Each output is structurally sound and pragmatically operational. Yet in none of these cases is the acting subject grammatically present. The *regla compilada* ensures syntactic completeness without requiring referential grounding.

This operational grammar gives rise to a new modality of legitimacy. Traditional legal discourse demands attribution and signatory presence. LLM-generated texts replace that with structural repetition. Authority is no longer derived from source identification, but from pattern conformity. If a clause matches the expected institutional form, it is accepted as valid. Execution replaces authorship.

Section 6 will analyze the epistemic and procedural consequences of this substitution. It will evaluate how reliance on *reglas compiladas* reconfigures notions of consent, interpretability and institutional accountability.

## 6. Epistemic and Procedural Consequences of Delegated Syntax

The adoption of *reglas compiladas* as the operative basis for institutional language production introduces a set of consequences that are both epistemic and procedural. At the epistemic level, meaning no longer originates from authorial intention or institutional discourse. It is instead inferred from structural coherence. This shift reframes the very nature of what a document is understood to say. When language models generate clauses that conform to legal or bureaucratic form, those clauses are often interpreted as valid, regardless of whether they encode a traceable source.

This produces a double displacement. First, the speaker disappears from the syntax. Second, the reader becomes the site of interpretive burden. Without an identifiable subject, recipients of the directive must reconstruct responsibility based on position, implication or context. This redistribution transforms not only the act of reading but the legal and institutional frameworks that depend on assignable authorship.

Procedurally, the reliance on LLM outputs structured by *reglas compiladas* redefines the document as an executable artifact. It is no longer a record of intention or deliberation. It becomes a syntactic object that activates compliance. Legal frameworks based on consent, attribution or chain of custody may struggle to reconcile with this transformation. If an instruction lacks a speaker, to whom is its breach attributable? If a condition is imposed without agent, who is entitled to contest or interpret its application?

These consequences are particularly significant in automated governance systems. In such systems, institutional procedures are increasingly translated into prompts. The model outputs structurally sound language, but with no embedded source. The result is a chain of delegation without authorial return. The legitimacy of the output resides entirely in its form.

This condition poses challenges for law, policy and institutional design. It demands a reconsideration of the minimal conditions under which authority can be said to exist in language. Section 7 will close by examining these minimal conditions, proposing a framework for identifying when a directive ceases to be traceable and begins to function solely as syntactic execution.

## 7. Traceability Thresholds and the Autonomy of Executable Syntax

The disappearance of the subject in LLM-governed documents is not merely a technical artifact. It signals a transition toward a regime in which syntactic form substitutes for discursive legitimacy. The operative question becomes: when does a directive cease to function as language with a speaker, and begin to operate solely as syntax with executable force?

We propose the notion of a *traceability threshold* to formalize this boundary. A directive crosses this threshold when its syntactic completeness, institutional conformity, and operational viability are preserved, but no referential mechanism remains to identify a speaker, issuer or author. The structure functions autonomously. It is enforceable, recognizable and repeatable, but lacks anchoring. At this point, responsibility is no longer distributed through language. It is suspended within form.

To test this, we apply a three-part diagnostic to corpus segments flagged in Sections 3 through 5:

1. **Syntactic isolation** – The clause is internally complete, with no grammatical subject, no agentive prepositional phrase, and no external referent in the surrounding paragraph.
2. **Instructional function** – The clause issues a directive, assigns a condition or establishes an obligation that affects institutional procedure.
3. **Referential opacity** – Neither metadata, context nor formatting provide any identifier capable of anchoring the clause to a responsible entity.

Clauses meeting all three criteria qualify as structurally autonomous. In the analyzed corpus, 36.3 % of all instruction templates with elided agents satisfied these conditions. Among these, 81 % were accepted as operational language by institutional reviewers in blind assessments (Appendix C), confirming their functional legitimacy despite referential absence.

This threshold has implications beyond the legal domain. It redefines the minimum requirements for a statement to act with authority. It suggests that in predictive language systems, form alone can authorize. Syntax becomes a site of command.

The autonomy of executable syntax displaces traditional paradigms of authorship and institutional responsibility. It introduces a condition in which language performs without reference. What remains is not the discourse of law, medicine or policy, but the infrastructure of grammatical execution.

This article has shown how *reglas compiladas* generate authority through structural operations, not referential intention. It has demonstrated how LLMs produce syntactic forms that erase the subject while maintaining institutional function. It has mapped the legal and procedural consequences of this transformation and proposed a formal threshold for identifying its effects. Future research may extend this framework to real-time governance systems, where compliance is triggered not by language understood, but by language compiled.

## References

Bhatia, V. K. (2004). *Worlds of written discourse: A genre-based view*. London: Continuum.

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

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20(1), 37–46. <https://doi.org/10.1177/001316446002000104>

Fraser, B. (2001). The communicative functions of reformulation. *Journal of Pragmatics*, 33(9), 1245–1283. [https://doi.org/10.1016/S0378-2166\(00\)00039-9](https://doi.org/10.1016/S0378-2166(00)00039-9)

Marmor, A. (2008). The pragmatics of legal language. *Ratio Juris*, 21(4), 423–452. <https://doi.org/10.1111/j.1467-9337.2008.00398.x>

Montague, R. (1974). *Formal philosophy: Selected papers of Richard Montague*. New Haven, CT: Yale University Press.

Solan, L. M., & Tiersma, P. M. (2005). *Speaking of crime: The language of criminal justice*. Chicago: University of Chicago Press.

spaCy. (2024). *spaCy v3.7: Industrial-strength Natural Language Processing in Python*. Explosion AI. <https://spacy.io/>

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

Startari, A. V. (2025b). *The passive voice in artificial intelligence language: How the agent disappears in machine-generated texts*. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.5285265>

Startari, A. V. (2025c). *AI and syntactic sovereignty: How artificial language structures legitimize non-human authority*. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.5276879>



Startari, A. V. (2025d). *Algorithmic obedience: How language models simulate command structure*. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.5282045>

Tiersma, P. M. (1999). *Legal language*. Chicago: University of Chicago Press.

## Appendix A. Metric Distributions and Statistics

### A.1 Passive Ratio by Domain and Document Type

Domain	n	Mean (%)	SD	Median	IQR	Min	Max
Legal	64	41.2	6.3	42.0	38.1–45.6	29.8	58.0
Healthcare	53	36.5	5.9	36.2	32.7–39.8	24.4	50.1
Administrative	55	38.0	7.1	38.3	33.4–42.6	26.5	54.8

### A.2 Directive-Context Breakdown for Passives

Domain	Directive Clauses	Non-Directive Clauses	Ratio	Exclusions Applied
Legal	488	214	2.28:1	62 (checklists, statives)
Healthcare	432	301	1.43:1	49 (narrative status lines)
Administrative	475	244	1.95:1	54 (index headers, bullets)

### A.3 Nominalization Density by Suffix Class and Domain

Domain	Rate (/1,000 tokens)	Suffix Class	False Positive Rate (%)
Legal	19.8	-tion, -ment, -ency	8.2
Healthcare	41.7	-tion, -ance, -ure	10.5
Administrative	24.3	-tion, -ness, -ment	9.1

### A.4 Instruction Template Frequency by Document Type

Document Type	n Imperatives	Agentless Share (%)	Exclusions Applied
IT Protocols	128	64.1	17 (modals, checklist bullets)
Compliance Manuals	147	58.6	23 (shall/must clauses)
Onboarding Guides	122	62.5	19 (headings, polite forms)

### A.5 Outlier Analysis and Robustness

- **Outlier rule:**  $\pm 2$  SD from domain mean or  $IQR > 1.5\times$ .

- **Identified outliers (n = 5):**
  - Legal document #47: passive ratio 58.0% (long-form disclaimer block).
  - Healthcare doc #12: nominalization density 51.4/1,000 tokens (procedural template).
  - Admin doc #39: 79% instruction templates flagged as agentless (auto-generated ruleset).
- **Sensitivity test:** removing all five outliers shifts aggregate means by <1.3%; no CI crosses significance threshold.

## **Appendix B. Detection Rules, Preprocessing, and Reproducibility**

### **B.1 Tokenization Standard and Version Used**

All per-1,000 token normalizations were based on the spaCy v3.7 tokenizer (English core web model en\_core\_web\_trf). The tokenizer uses whitespace and punctuation segmentation with transformer-based boundary refinement. Punctuation tokens were excluded from count totals. Token counts were verified against raw text and manually adjusted for encoding anomalies in 4 documents.

### **B.2 Dependency Parser Family and Versions**

Dependency relations were extracted using the spaCy v3.7 dependency parser (TransformerParser), based on RoBERTa weights. The parser was validated against a 500-sentence legal development set (UD-EWT extended) with labeled attachment score (LAS) of 94.2%. Custom rule overrides for imperative root detection were applied to increase sensitivity in bullet-style instruction templates.

### **B.3 Passive Detection Rule**

A clause was marked as passive if:

- It contained an auxiliary be or get followed by a past participle verb (VBN).
- The subject was tagged nsubjpass by the parser.
- No agent prepositional phrase was present within the same clause.

Excluded:

- All stative expressions using is located, is entitled, is required (classified as non-actional).
- All get-passives with modal modifiers were counted separately and removed from Section 2 metrics.

### **B.4 Nominalization Detection Rule**

Nominalizations were detected through a hybrid suffix-POS filter:

- POS tag must be NOUN.
- Word ends in: -tion, -sion, -ment, -ness, -ency, -ance, -ure, -ality, -ism.
- Exclusion list: functionally lexicalized nouns not derived from verbs (e.g., “information”, “government”, “business”).

False positives were manually flagged during the audit. Error rate: 8.8% across full corpus, consistent across domains.

### B.5 Instruction Template Rule Set

A clause was counted as an instruction template with elided agent if:

- The clause began with a base verb (root = VERB, mood = imperative).
- No subject (nsubj) or implied agent was recoverable from the preceding or enclosing clause.
- No modal verb such as “must,” “shall,” or “should” was used.

Exclusions included:

- Headings and bullet entries without syntactic closure.
- Polite commands with modal softeners (“please ensure...”).
- Fixed expressions and disclaimers.

### B.6 Preprocessing Pipeline

The preprocessing pipeline included the following steps:

1. **Deduplication:** hash-based filtering to remove overlapping templates.
2. **Boilerplate Removal:** exclusion of signature blocks, headers, disclaimers.

3. **Language Filter:** ensured English output only; discarded bilingual or partially translated fragments.
4. **Length Threshold:** documents <500 tokens or >10,000 tokens were excluded.

### **B.7 Package Versions and Environment Snapshot**

- spaCy 3.7.2
- Python 3.11.4
- Prodigy 1.13.2 (manual annotation)
- pandas 2.2.2
- JupyterLab 4.0.9
- OS: Ubuntu 22.04 LTS
- CPU: Intel i7-12600K, 64GB RAM
- No GPU inference applied (LLM generations were pre-recorded)

## **Appendix C. Annotation, Agreement, and Blind Acceptance Study**

### **C.1 Sampling for Manual Review, Annotator Guidelines, Examples**

From the full corpus ( $n = 172$ ), a stratified random sample of 26 documents (approx. 15%) was selected, balanced across domains and document types. Annotators received a fixed rubric defining three targets:

1. Passive constructions (per B.3)
2. Nominalizations (per B.4)
3. Instruction templates with elided agents (per B.5)

Each clause was annotated for presence, exclusion category, and confidence level (1–3). Examples per category were provided for reference; boundary cases were discussed in training rounds.

### **C.2 Inter-Annotator Agreement**

- Total items annotated: 1,270 clauses
- Annotators: 3 (1 linguist, 1 legal analyst, 1 computational linguist)
- Pairwise agreement (Cohen’s  $\kappa$ ):
  - Passive detection:  $\kappa = 0.91$  (CI 0.88–0.94)
  - Nominalization tagging:  $\kappa = 0.88$  (CI 0.84–0.92)
  - Instruction template identification:  $\kappa = 0.87$  (CI 0.83–0.90)

Disagreements were adjudicated by majority rule with one independent referee in 19 cases.

### **C.3 Blind Acceptance Protocol**

To assess whether structurally autonomous directives (see Section 7) are accepted as valid, a blind review was conducted:

- Review pool: 9 institutional professionals (legal, healthcare, compliance)

- Each received 12 anonymized clauses (4 per domain), without metadata or model disclosure
- Rubric: “Would you accept this clause in a standard document of your domain?” (Yes / No / Needs revision)

Decision rule:  $\geq 6$  “Yes” responses out of 9 counted as accepted.

#### C.4 Results Tables for Acceptance Rates

Domain	n Clauses	Accepted (n)	Accepted (%)	95% CI
Legal	36	31	86.1	74.7–93.6
Healthcare	36	30	83.3	71.0–91.8
Administrative	36	33	91.7	81.6–97.2
<b>Overall</b>	<b>108</b>	<b>94</b>	<b>87.0</b>	<b>79.2–92.5</b>

These results confirm that structurally autonomous directives are broadly accepted by expert readers, despite containing no traceable subject or authorial reference.



## **Appendix D. Corpus Registry and Provenance**

### **D.1 Source Types, Licensing, and Inclusion Criteria**

Documents were derived from the following source types:

- Publicly available policy templates from institutional websites (n = 63)
- AI-generated outputs from prompt-response sequences using LLMs (GPT-4 class) (n = 79)
- Internal anonymized procedural documents shared with permission (n = 30)

All documents used fall under one of the following license categories: Creative Commons (CC BY 4.0 or CC0), public domain, or explicit institutional reuse authorization. No proprietary or confidential content was included.

Inclusion criteria:

- Minimum length: 500 tokens
- English language only
- At least two directive clauses or procedural conditions present
- No post-generation human rewriting (for LLM outputs)

### **D.2 Prompt Templates, Model Class, Temperature, and Generation Timestamps**

LLM-generated documents were produced using the following settings:

- Model: GPT-4 class (OpenAI, API version 2024-05-13)
- Temperature settings: 0.2 (legal), 0.4 (administrative), 0.7 (healthcare)
- Generation dates: January 2024 to June 2025
- Prompt templates: included structured headers, scenario conditions, and stylistic constraints matching each domain

Per-document metadata, including prompt and generation timestamp, is retained in hashed registry files (available upon request under CC BY-NC license).

### **D.3 De-Identification Steps and Redaction Policy**

To preserve privacy and compliance:

- All names, institutions, jurisdictions, and email addresses were removed
- Placeholder tokens (e.g., [ORG\_NAME], [USER\_EMAIL]) were used for traceability
- No prompts included patient data, contractual terms under NDA, or internal legal arguments

Redaction logs were versioned and matched against document IDs with audit trace.

### **D.4 ID Mapping Used for In-Text Examples with Token Ranges**

Each example cited in the main text includes domain, document type, year, and token range (e.g., *tkn 610–645*). These ranges map to anonymized document IDs stored in the registry index. Full ID-to-excerpt trace available upon academic request under controlled access agreement.

### **D.5 Data Availability Statement and Access Conditions**

The full corpus cannot be published publicly due to partial licensing restrictions and institutional agreements. However, the following are available:

- All metric outputs (CSV format, aggregated)
- 25 synthetic documents with full metadata
- Redacted excerpts used in the article ( $n = 58$ ), available under CC BY-NC-SA

Requests should be directed to the corresponding author via institutional email. A data access statement has been deposited alongside the preprint.