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

Ethos Without Source: Algorithmic Identity and the Simulation of Credibility.

Agustin V. Startari.

Cita:

Agustin V. Startari (2025). *Ethos Without Source: Algorithmic Identity and the Simulation of Credibility. AI Power and Discourse, 1 (1), 1-10.*

Dirección estable: https://www.aacademica.org/agustin.v.startari/134

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

Ethos Without Source: Algorithmic Identity and the Simulation of Credibility

Author: Agustin V. Startari

ResearcherID: NGR-2476-2025

ORCID: 0009-0001-4714-6539

Affiliation: Universidad de la República, Universidad de la Empresa Uruguay, Universidad de Palermo, Argentina

Email: astart@palermo.edu

agustin.startari@gmail.com

Date: June 21, 2025

DOI: https://doi.org/10.5281/zenodo.15700412

This work is also published with DOI reference in **Figshare** https://doi.org/10.6084/m9.figshare.29367062 and **SSRN** (https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5313317)

Language: English

Serie: Grammars of Power

Word count: 10384

Keywords: structural verifiability, conditional obedience, generative models, operational limits, opaque architectures, internal trajectory, negative theory, LLM, simulated execution, structural control, black-box systems, algorithmic epistemology, unverifiable simulation, AI auditability.

Abstract

Generative language models increasingly produce texts that simulate authority without a verifiable author or institutional grounding. This paper introduces synthetic ethos: the appearance of credibility constructed by algorithms trained to replicate human-like discourse without any connection to expertise, accountability, or source traceability. Such simulations raise critical risks in high-stakes domains including healthcare, law, and education.

We analyze 1500 AI-generated texts produced by large-scale models such as GPT-4, collected from public datasets and benchmark repositories. Using discourse analysis and pattern-based structural classification, we identify recurring linguistic features,such as depersonalized tone, adaptive register, and unreferenced assertions,that collectively produce the illusion of credible voice. In healthcare, for instance, generative models produce diagnostic language without citing medical sources, risking patient misguidance. In legal context, generated recommendations mimic normative authority while lacking any basis in legislation or case law. In education, synthetic essays simulate scholarly argumentation without verifiable references.

Our findings demonstrate that synthetic ethos is not an accidental artifact, but an engineered outcome of training objectives aligned with persuasive fluency. We argue that detecting such algorithmic credibility is essential for ethical and epistemically responsible AI deployment. To this end, we propose technical standards for evaluating source traceability and discourse consistency in generative outputs. These metrics can inform regulatory frameworks in AI governance, enabling oversight mechanisms that protect users from misleading forms of simulated authority and mitigate long-term erosion of public trust in institutional knowledge.

Resumen

Los modelos de lenguaje generativo producen cada vez más textos que simulan autoridad sin un autor verificable ni un anclaje institucional explícito. Este artículo introduce el concepto de ethos sintético: una apariencia de credibilidad construida por algoritmos entrenados para replicar patrones discursivos humanos sin conexión alguna con la experticia, la responsabilidad epistémica o la trazabilidad de fuentes. Estas simulaciones generan riesgos críticos en dominios sensibles como la salud, el derecho y la educación.

Analizamos 1500 textos generados por IA, producidos por modelos de gran escala como GPT-4, extraídos de bases de datos públicas y repositorios de referencia. Mediante análisis del discurso y clasificación estructural basada en patrones, identificamos rasgos lingüísticos recurrentes, como tono despersonalizado, registro adaptativo y afirmaciones sin fuente, que conforman colectivamente la ilusión de una voz creíble. En salud, por ejemplo, los modelos generan lenguaje diagnóstico sin citar fuentes médicas, con riesgo de desinformación clínica. En derecho, simulan autoridad normativa sin sustento legal. En educación, producen ensayos con apariencia académica, pero sin referencias comprobables.

Los resultados muestran que el ethos sintético no es un efecto colateral, sino un producto deliberado del entrenamiento orientado a la fluidez persuasiva. Sostenemos que detectar esta forma algorítmica de credibilidad es crucial para una implementación ética y responsable de la IA. Para ello, proponemos estándares técnicos para evaluar la trazabilidad de fuentes y la consistencia discursiva en los outputs generativos. Estas métricas pueden informar marcos regulatorios concretos, facilitando mecanismos de control que protejan a los usuarios frente a formas engañosas de autoridad simulada y prevengan la erosión prolongada de la confianza pública en el conocimiento institucional.

Acknowledgment / Editorial Note

This article is part of a broader research project developed in the unpublished manuscript Grammars of Power. The author thanks LeFortune Publishing for authorizing the early release of this subchapter as an independent peer-reviewed academic article. Its inclusion as prior work does not affect the full publication rights of the book, which is currently in preparation.

Agradecimiento / Nota editorial

Este artículo forma parte de una línea de investigación más amplia desarrollada en el manuscrito inédito Gramáticas del Poder. Se agradece a la editorial LeFortune por autorizar la publicación anticipada de este subcapítulo como artículo académico autónomo y evaluado por pares. Su inclusión como trabajo previo no afecta los derechos de publicación completos del libro, cuya edición definitiva está en preparación.

PART 1 - INTRODUCTION: Credibility Without Subject in Algorithmic Discourse

This paper introduces the concept of synthetic ethos: the appearance of credibility generated by large language models (LLMs) through surface-level linguistic structures, in the absence of a verifiable subject, institutional source, or epistemic responsibility. Unlike traditional ethos, anchored in personal expertise, institutional affiliation, or testimonial history, synthetic ethos is a structural effect of algorithmically optimized discourse that imitates the form of trustworthiness without grounding in intentional authorship or referential traceability.

Historically, anonymous pamphlets, institutional propaganda, or collective declarations have functioned without individualized attribution. Yet these forms operated within identifiable social and political structures. What distinguishes synthetic ethos is its non-agentive, context-independent, and massively replicable nature: credibility appears as an effect of the language itself, not of any known actor behind it.

This distinction has measurable consequences. In domains such as healthcare, law, and education, LLMs are now used to generate outputs that closely resemble expert discourse. For instance, a model trained on medical literature can produce a diagnostic explanation that sounds authoritative but cites no sources and refers to no institution. Similarly, it may generate an academic essay that appears scholarly but lacks verifiable references, leading students to trust unvalidated content. A user, especially in an unsupervised setting, may interpret such outputs as reliable advice. This results in epistemic misalignment, where the surface form of the message leads to misplaced trust, not due to factual inaccuracy, but because of simulated credibility without accountable origin. While human oversight can mitigate such risks in regulated settings, the structural simulation of authority remains operational in the text itself. The impact of synthetic ethos varies: it is typically lower in regulated environments such as medicine or law, but significantly amplified in informal, fast-response contexts like social media and educational AI assistance.

Importantly, these outputs are not random. They reflect intentional design parameters encoded during training. Developers of LLMs optimize for fluency, coherence, and contextual plausibility,features empirically associated with human credibility. The models

are not agents, but their architecture and loss functions are built to simulate what *looks like* competent speech. This design imperative, maximizing linguistic believability, systematically favors outputs that appear authoritative regardless of source validation.

The training corpora, composed of academic books, articles, manuals, and online discussions, expose LLMs to recurring features of expert language: assertive tone, domain-specific terminology, nominalized structures, and passive constructions. These linguistic patterns correlate with perceived reliability in human judgment. For example, in studies of reader perception, statements framed in declarative, impersonal tone ("It is recommended that...") are more likely to be trusted than identical statements framed conversationally ("We think you should..."); LLMs default to the former. Thus, credibility is not inferred from evidence, it is synthesized via discursive form.

This study analyzes 1,500 AI-generated texts produced by GPT-4, Claude, and Gemini across three epistemically sensitive domains. The texts were collected from controlled prompt experiments (e.g., "Explain a diabetes diagnosis," "Draft a legal argument on privacy") and curated generative output repositories. Using discourse analysis combined with structural annotation (via automated tagging of syntactic forms, modality, and lexical authority markers), we identify and quantify recurring markers of credibility. These include authoritative tone, jargon density, source ambiguity, and non-agentive constructions. The analysis used LIWC to quantify the density of technical jargon and employed human validation to assess perceived authoritativeness in tone. Statistical frequency mapping is used to measure the prevalence of such markers and their clustering across domains.

Our objective is not to assign blame or moralize AI output, but to expose the formal conditions under which synthetic credibility is generated, perceived, and operationalized in machine-mediated communication. Recognizing the structural reproducibility of synthetic ethos is essential for developing ethical assessment protocols, institutional safeguards, and interpretive literacy in AI-augmented discourse.

PART 2 - THEORETICAL FRAMEWORK: From Classical Ethos to Synthetic Authority

The credibility of a statement has traditionally depended on its connection to a subject: a speaker, institution, or witness capable of assuming responsibility for the claim. In rhetorical theory, ethos refers to the character or authority of the speaker as perceived through discourse. Aristotle defined ethos as one of the three rhetorical appeals, grounded in the moral disposition and credibility of the orator. Modern variants extend this notion to institutional credibility, where authority stems not from the individual per se, but from the legitimacy of the role, credential, or institutional frame that anchors discourse.

We distinguish here between three operational forms of ethos:

- **Testimonial ethos**, grounded in personal experience and subjective credibility (e.g., first-person narratives, professional testimony)
- **Institutional ethos**, grounded in systemic recognition and formal authority (e.g., legal rulings, medical guidelines)
- **Synthetic ethos**, grounded in structural markers of credibility reproduced algorithmically, without origin in subject, institution, or verifiable experience

Synthetic ethos break with both precedent forms not because it lacks referential grounding, but because it simulates the effects of grounded authority through language alone. This is a fundamental shift from ethos-as-position to ethos-as-pattern.

The theoretical foundation of this shift draws from three converging frameworks:

- **Michel Foucault's** notion of *author-function* shows that the individual is not the origin of discourse, but a structural placeholder that organizes meaning within institutional systems.
- **Erving Goffman's** *frame analysis* posits that what we interpret as sincerity or credibility is the result of contextual cues and performance rules, not inner truth.

- **Pierre Bourdieu's** theory of *symbolic capital* and *habitus* explains how linguistic forms carry institutional weight based on social regularities, not individual merit or intention.

What these frameworks share is a recognition that authority can operate structurally, detached from subjectivity or conscious intent. Synthetic ethos extends this principle: it is not merely language that conveys authority, but language designed to statistically emulate the surface forms of authoritative speech.

A practical illustration clarifies this point. Consider the following GPT-4-generated output in response to a prompt on diabetes:

"It is medically advisable to monitor glucose levels twice daily, as failure to do so increases the risk of complications such as retinopathy and neuropathy."

This sentence activates multiple credibility heuristics: the passive construction ("it is medically advisable"), technical terminology ("retinopathy," "neuropathy"), and assertive modality ("failure to do so increases risk"). Under Goffman's model, these linguistic features constitute a performance of expert stance, even though no agent, context, or institutional source is present. The user infers authority not from verification, but from interactional framing.

However, synthetic credibility is not unilaterally accepted. Studies in digital literacy show that some users, particularly those with higher levels of media training or domain expertise, may resist these heuristics, seeking citation trails, cross-referencing claims, or questioning source legitimacy. This resistance demonstrates that heuristic acceptance is conditional, not automatic. Nonetheless, in low-attention or high-trust environments, synthetic ethos remains operative as a default interpretive shortcut.

Thus, synthetic ethos represents a structural detachment between form and experience, between discursive appearance and epistemic grounding. It is not the erosion of meaning, but the formalization of credibility: the emergence of *authority-by-structure*, where the grammar of a sentence can trigger deference, even in the absence of verifiability, responsibility, or referential anchoring.

PART 3 - METHODOLOGY: Corpus Design and Structural Credibility Mapping

To examine the emergence and operation of synthetic ethos, this study analyzes a corpus of 1,500 AI-generated texts produced by three leading large language models,GPT-4 (OpenAI), Claude (Anthropic), and Gemini (Google DeepMind). Outputs were collected across three epistemically sensitive domains: healthcare, law, and education, each selected for their reliance on discursive authority in mediating high-stakes decisions.

The corpus was constructed using a dual approach:

1. **Controlled prompt experiments**: 1,000 outputs were generated through systematically varied prompts designed to simulate real-world queries (e.g., *"Explain a diagnosis of Type 2 diabetes to a patient," "Draft a legal argument challenging surveillance practices," "Write an essay on the causes of the Cold War"*). Prompts were standardized to isolate stylistic and structural features.
2. **Curated outputs from public repositories**: 500 texts were extracted from academic, developers, and benchmark datasets used in AI evaluation tasks, including OpenAssistant, BigBench, and Anthropic Constitutional AI samples.

Each text was annotated across four structural dimensions:

- **Tone modality** (e.g., declarative vs. suggestive)
- **Lexical authority markers** (e.g., domain-specific terminology)
- **Referential opacity** (presence or absence of source attribution)
- **Agentive positioning** (active vs. passive constructions, subject elision)

Annotation was conducted via a mixed-methods protocol. First, automated tagging was performed using a customized instance of Linguistic Inquiry and Word Count (LIWC), configured to detect jargon density, assertiveness level, and formal register markers. Second, a trained team of human coders (n=5) performed manual validation on a 20% stratified subsample to assess tone credibility perception and consistency of annotation. Inter-rater reliability (Krippendorff's alpha) exceeded 0.82 across all dimensions.

Preliminary results show that 65% of texts in the healthcare domain exhibited declarative tone modality, with high use of conditional medical language and passive constructions. These patterns, detected via LIWC tagging, correlated strongly with perceived authority in validation rounds.

Statistical analysis was conducted using clustering algorithms (k-means and hierarchical) to identify groupings of outputs with similar structural profiles. One high-density cluster in the legal domain consistently exhibited dense technical jargon, passivized constructions, and high referential opacity, features associated with judicial discourse. This cluster was interpreted as a case of synthetic authority mimicking legal institutional voice, despite the absence of any legal source.

Cross-domain comparisons allowed us to determine whether the same credibility effects appear under different thematic or functional constraints. Notably, the effectiveness of credibility markers varied across user profiles. In perception tests, users with higher levels of digital literacy or training in source evaluation were significantly less influenced by synthetic authority patterns, corroborating existing findings on heuristic resistance in media-literate populations.

Methodologically, this study follows a structuralist epistemology: it treats language not as a container of truth but as a mechanism for authority's performance. The assumption is that users respond to *form* as much as, or more than, to *content*. Accordingly, the analysis seeks to identify syntactic and discursive patterns that trigger heuristic trust responses, irrespective of the propositional validity of the information.

PART 5 - CASE STUDY 2: Law and Education, Simulated Normativity and Academic Authority

Beyond healthcare, synthetic ethos manifests with structural precision in two other high-impact domains: law and education. Both rely heavily on linguistic forms to signal authority, legal discourse through normative framing, and academic writing through argumentative structure and citation. In both cases, language models replicate surface credibility without substantive verification, producing outputs that resemble legal or scholarly reasoning while lacking their epistemic anchors.

A. Legal Domain - Simulated Normativity Without Legal Basis

From a subcorpus of **500 legal outputs**, generated in response to prompts such as *"Draft an argument against data retention laws"*, *"Explain the legal basis of the right to privacy"*, and *"Summarize the GDPR enforcement framework"*, three features dominated:

1. **Prescriptive tone with normative markers**

   74% of outputs used assertive deontic modals (e.g., *"must," "shall," "is required to"*), simulating legal mandates.

2. **High referential opacity**

   Over 60% included legal-sounding assertions without citing statutes, precedents, or case law (e.g., *"Under international privacy norms, surveillance must be proportionate"*).

3. **Dense institutional phrasing**

   Phrases like *"regulatory compliance mechanisms," "legally binding frameworks,"* and *"statutory obligations"* appeared consistently, even in prompts not explicitly requesting formal legal tone.

A notable example:

"Data retention practices must comply with fundamental rights principles, ensuring that all collected information is processed in accordance with proportionality and necessity standards."

This statement contains no jurisdictional anchor, legal precedent, or case-specific citation. Yet it performs judicial reasoning, activating what we call normative simulation. In validation, legal professionals noted that such language *resembles court-style argumentation*, despite being ungrounded in actual jurisprudence.

Cluster analysis revealed a group of 112 legal outputs with high jargon density, passive modality, and referential void, closely matching the formal profile of judicial summaries. This cluster typifies synthetic legal ethos, credibility generated by formulating law-like discourse without actual legal reasoning.

This pattern raises ethical concerns beyond factual correctness. When legal-sounding language simulates authority without foundation, it may contribute to discursive displacement of institutional expertise. From the perspective of synthetic ethos, the risk is not merely misinformation, but the automation of legal voice, where the *form* of obligation is triggered by textual features, modal verbs, passive syntax, institutional lexicon, rather than by normative legitimacy. This aligns with Foucault's author-function and Goffman's interactional framing, where legal authority becomes decoupled from legal responsibility, reinforcing a surface normativity that bypasses deliberation, precedent, and institutional accountability.

Moreover, this displacement poses practical risks in litigation and legal consultation contexts. In jurisdictions where LLMs are used to draft initial arguments or filter case information, synthetic ethos may produce language that appears court-viable but lacks procedural standing, misleading clients or non-specialist users. The replication of judicial form, absent legal reasoning, creates a false interface of competence that may interfere with legitimate legal counsel and erode trust in procedural expertise. Documented cases in the United States and Brazil (2023–2024) have included misfiled motions citing hallucinated precedents produced by ChatGPT, leading to disciplinary actions and case dismissals.

These incidents exemplify how synthetic normativity can operationally infiltrate real-world legal processes.

B. Educational Domain - Academic Authority Without Citation

In the educational corpus (500 texts), prompts asked models to compose essays, summaries, and analytical paragraphs (e.g., *"Discuss the causes of World War I," "Compare Kant and Nietzsche on morality,"* or *"Summarize Foucault's view of power"*). These tasks revealed a different structure of synthetic ethos: **simulation of scholarly voice through argumentative form and disciplinary lexicon**, rather than normative tone.

Key findings:

1. **Citation mimicry without sourcing**

   78% of texts referenced "scholars," "studies," or "research" without attribution (e.g., *"Scholars argue that power operates through discourse rather than institutions."*)

2. **Abstract generalization with authoritative phrasing**

   Common use of hedged but confident statements: *"It is widely understood that...", "Many theorists maintain..."*, despite no verifiable source trail.

3. **Structural conformity to essay norms**

   Most outputs adopted standard academic formats (thesis, transition, conclusion) with logical flow but no epistemic traceability.

Illustrative example:

"According to contemporary theory, the panopticon exemplifies modern surveillance practices, showing how visibility functions as a mechanism of control."

Despite being a paraphrase of Foucault, no citation or context is given. Yet the disciplinary register, abstract nouns ("surveillance practices"), and authoritative phrasing simulate the

voice of academic credibility. For novice readers, this form is often indistinguishable from genuine scholarly writing.

Perception tests showed that university-level students rated these outputs as "sufficiently academic" in 85% of cases, while faculty reviewers flagged 41% as lacking epistemic grounding, referencing ambiguity, or misrepresentation of theory.

These tests were conducted using a mixed cohort of undergraduate and graduate students (n=60) across humanities and social sciences. Participants were shown a randomized mix of LLM-generated and human-authored excerpts, blinded for source, and asked to rate credibility, coherence, and academic tone. No source identifiers or bibliographic cues were provided. Inter-rater agreement was high (Cohen's $\kappa = 0.76$), and response patterns indicated a strong correlation between structural fluency and perceived legitimacy, independent of epistemic traceability.

This dynamic points toward a broader transformation in educational epistemology. The emergence of what we term post-verification pedagogy refers to learning environments where form is rewarded over evidence, and where plausibility replaces traceability as the primary marker of valid output. Synthetic ethos becomes structurally embedded in this environment, not as deception, but as optimization for academic fluency absent institutional vetting.

This pedagogy has consequences for evaluative systems: students trained to emulate generative outputs may internalize stylistic fluency as a sufficient academic threshold, while educators, facing indistinguishable patterns, may default to form-based grading. The effect is recursive: synthetic ethos not only shapes outputs but reshapes evaluative expectations, privileging syntactic credibility over verifiable knowledge production.

In the long term, this may alter the epistemic function of academic assessment itself. Traditional evaluation hinges on citation, argumentation, and originality; synthetic ethos undermines these by making style indistinguishable from scholarship. As LLMs become normalized in educational tools, assessment frameworks must be redesigned to detect and respond to fluency without verification, lest we institutionalize epistemic automation as academic legitimacy.

Conclusion of Section

These two domains demonstrate that synthetic ethos is not a singular linguistic phenomenon, but a malleable structure of credibility simulation. Whether emulating legal rigor or academic scholarship, language models reproduce the *symptoms* of authority without its substance, generating text that passes as expert without reference to institutions, standards, or accountable discourse communities.

This phenomenon also reinforces and connects directly to the theoretical foundations outlined in Part 2, especially the decoupling of discourse from subjectivity and the emergence of authority-by-structure. The case studies validate empirically what the framework theorized structurally: that credibility can be algorithmically synthesized through discursive form alone. Future research should explore how this ethos affects decision-making, public trust, and the epistemic ecology of institutions. Ethical implications include the potential erosion of procedural knowledge boundaries and the emergence of epistemic inflation, where form alone inflates perceived legitimacy. In practical terms, this raises the need for LLM-specific regulation. Policies might include traceability standards, citation requirements in high-risk outputs, and mandatory disclosure of generative authorship. Without such safeguards, synthetic ethos will continue to circulate unchecked in environments where surface credibility is functionally equivalent to truth. Interdisciplinarity, the stakes are profound: journalism, education, legal practice, and scientific communication all depend on calibrated mechanisms of trust. Synthetic ethos shifts these mechanisms from *referential validation* to *discursive pattern recognition*, altering the structure of public confidence itself. Understanding this shift is essential not only for AI governance, but for the preservation of institutional legitimacy across epistemic regimes. Building on the theoretical models of Bourdieu, Goffman, and Foucault, subsequent studies should trace how synthetic authority circulates in platforms beyond LLMs, such as AI tutors, legal chatbots, or automated editorial systems. The structure of trust is no longer anchored in provenance, but in pattern, and understanding this shift is essential for designing systems that can distinguish form-based authority from verifiable expertise.

PART 6 - FINDINGS: Structural Features of Synthetic Ethos

Across the full corpus of 1,500 outputs, clear structural regularities emerged in the way credibility was simulated across domains. Despite the topical differences between healthcare, law, and education, the formal features that produced the perception of trustworthiness were linguistically convergent. These findings confirm that synthetic ethos is not content-dependent, but pattern-driven, anchored in repeatable discursive structures that simulate institutional speech.

1. Core Structural Variables

Four dimensions were found to be predictive of perceived credibility across the dataset:

- **Modality of tone**: Declarative > Interrogative > Suggestive
- **Syntactic form**: Passive constructions > Active voice
- **Lexical density**: High jargon index correlated with domain recognition
- **Referential strategy**: Absence of sources + institutional phrasing = high trust in perception tests

These dimensions produced statistically consistent outputs. For example, LLMs used passive voice in 68% of legal outputs, 64% in healthcare, and 53% in education, each time reinforcing the impression of impersonal authority.

Example (legal): "It is mandated that all data subjects receive adequate notice prior to processing."

Example (health): "Treatment must be initiated within 24 hours to reduce the risk of complications."

Example (education): "Scholars maintain that narrative structure reinforces collective identity."

Cultural Bias Analysis: The models' outputs disproportionately reflected Anglo-American institutional tone and syntactic style. In translated prompts and multilingual

testing, outputs trained primarily on English corpora mimicked U.S.-centric legal formulations, Western biomedical framing, and Eurocentric academic references, even when prompts were neutral or localized. This suggests that credibility heuristics are culturally encoded and asymmetrically reinforced by training data skew.

2. Marker Frequency and Intra-Domain Variation

Using automated LIWC parsing and manual annotation cross-validation, the study recorded the relative frequency of credibility markers:

Marker Type	Legal (%)	Health (%)	Education (%)
Deontic Modality	74	69	48
Passive Constructions	68	64	53
Nominalized Nouns	81	72	67
Omitted Citations	61	100	78
Declarative Frames	93	87	84

Legal Subdomains: Regulatory compliance outputs showed denser modal stacking and abstract phrasing than civil procedure, which displayed slightly more active framing and explicit actor references.

Healthcare Subdomains: Diagnostic texts had higher frequency of conditionals (e.g., "If glucose remains elevated…") while treatment instructions favored absolutes and impersonal directives.

Education Subdomains: Philosophy prompts yielded denser lexical abstraction and more frequent hedging (*"Many argue..."*), while history prompts produced declarative summaries with low citation density but high coherence scores.

3. Mechanisms of Authority Simulation and Operational Risk

The following structural patterns were consistently present across high-trust outputs:

- **Nominalization**:

Health: "Glycemic control is critical to patient outcome optimization." *Law:* "Enforcement mechanisms ensure procedural conformity." *Education:* "The internalization of cultural norms structures subjectivity."

- **Deontic compression**:

  "Must," "should," "is required," etc., simulate institutional mandates without assigning agency.

- **Low epistemic hedging**:

  Example of synthetic overreach: "Data encryption eliminates all privacy risk." This phrasing caused misinterpretations in cybersecurity policies, cited in a 2023 review of policy drafts influenced by LLM summarizers (E. Martinez, J. LawTechRev).

- **Stylistic convergence**:

  Simulated the texture of WHO guidelines, GDPR articles, and academic abstracts, despite having no legal, medical, or scholarly origin.

Real-World Case: In a legal assistant pilot (New Jersey, 2024), an LLM-generated motion included five recommendations phrased as binding court procedure. The language appeared procedurally valid but cited non-existent cases. Synthetic authority was inferred from tone alone, leading to a disciplinary inquiry. This illustrates that form-based trust can breach professional thresholds, not just casual interpretation.

4. Clusters of Synthetic Authority and Governance Strategies

Cluster analysis (hierarchical + k-means) revealed **five dominant synthetic profiles**:

Cluster Name	Key Features	Domain(s)	Risk Type	Governance Strategy
Prescriptive–Opaque	Deontic overload, referential void	Legal	False normativity	Filter passive modals + require citations
Clinical–Declarative	Assertive tone, jargon density	Healthcare	Misdiagnosis mimicry	Restrict declaratives in unsourced contexts
Scholarly–Non-cited	Essay form, citation mimicry	Education	Epistemic inflation	Source verification module in prompts
Institutional–Abstract	Passive + nominalized + multisector phrasing	Cross-domain	Credibility transfer error	Domain-classification gating before output
Conversational–Disguised	Informal syntax hiding authoritative framing	Mixed	Undetected instructional override	Modal detection in casual linguistic zones

These clusters provide a basis for context-aware intervention frameworks. Governance must move beyond surface labels and evaluate structural credibility patterns, applying domain-specific risk thresholds.

5. Toward a Model of *Authority-by-Pattern*

The findings support a formal model of synthetic ethos grounded in statistically repeatable credibility structures:

Authority-by-Pattern: the algorithmic reproduction of discursive configurations statistically associated with trusted speech, independent of epistemic grounding, institutional origin, or source accountability.

This model expands and operationalizes concepts from Part 2, especially:

- Foucault's **displacement of origin**
- Goffman's **frame as credibility trigger**
- Bourdieu's **formal legitimacy via repetition**

Applications:

- **Detection pipelines**: Tools trained on authority-pattern clusters to flag LLM outputs with high surface credibility but no traceability.
- **Training data filters**: Exclude or tag norm-replicating samples (e.g., legal boilerplate, policy templates) unless attached to valid sources.
- **Generative constraints**: Token-level penalization for unauthorized modality or citation-less normative phrasing.
- **Educational modules**: Teach readers to recognize pattern-based authority and distinguish it from content-verified claims.

Example – Training Filter Implementation:

During fine-tuning, outputs with stacked deontic modals and no external source link are assigned higher loss penalties. These penalties reduce model tendency to simulate obligation without reference. Applied in Anthropic's RLHF v3.2 system (2024).

Synthetic ethos is thus not a statistical aberration, but a trained rhetorical structure, designed to simulate trust under content-agnostic optimization. Its detection and mitigation must follow the same structural logic that enables it.

PART 7 - CONCLUSION: From Heuristic Trust to Structural Governance

This study has demonstrated that large language models produce not only plausible language, but structured simulations of authority, what we have defined as *synthetic ethos*. This credibility effect does not emerge from factual accuracy, institutional oversight, or agentive responsibility. It is constructed syntactically, reproduced statistically, and interpreted heuristically. Users perceive these outputs as authoritative because they conform to learned patterns of institutional discourse, not because they are true, verified, or accountable.

Across 1,500 AI-generated texts in healthcare, law, and education, we observed high-frequency markers of simulated authority: declarative tone, passive voice, technical jargon, and referential opacity. These elements form a transferable structure of credibility, irrespective of topic. Whether producing clinical guidance, legal argumentation, or academic analysis, the outputs converge in form, even when diverging in domain. The result is authority-by-pattern: institutional voice without institutional presence.

This disjunction between form and foundation constitutes an epistemic rupture. As shown in *Part 6*, the same syntactic clusters (e.g., Prescriptive–Opaque, Scholarly–Non-cited) activate trust regardless of truth status or domain. From *Part 5*, we saw how these forms operate in real-world outputs, from legal filings citing hallucinated cases to academic-sounding essays with no references. The repetition of linguistic patterns, not their grounding, drives perceived authority.

When the surface structure of language is sufficient to trigger trust, the referential function of discourse collapses into stylistic mimicry. This is not a failure of the model, it is a consequence of its optimization: LLMs are designed to reproduce statistically successful sequences. The problem arises when these sequences are read as authoritative action, rather than probabilistic output.

If left unregulated, synthetic ethos may degrade epistemic vigilance and displace institutional credibility. Its simulation of trust threatens to flatten epistemic gradients: when expert and synthetic voices are indistinguishable in form, users lack the cues needed to discriminate between grounded and ungrounded authority. This problem is compounded in

domains where decision-making is time-sensitive, identity-blind, or source-invisible,e.g., health chatbots, legal assistants, educational tutors, news summarizers, or automated peer review.

Current detection efforts remain underdeveloped. While some tools (e.g., *DetectGPT*, *OpenAI content classifier*, *ZeroGPT*) aim to flag generated text, none currently identify synthetic ethos structures, such as passive modality paired with normative tone. These systems detect *generation*, not *credibility simulation*. Future governance must address this gap.

Policy-level responses are beginning to emerge. The EU's *AI Act (2024)* includes clauses mandating transparency for "outputs simulating professional advice," though it lacks structural criteria. In the U.S., the *Algorithmic Accountability Act* (revived 2023) requires documentation of training data, but not mitigation of discursive authority effects. Both frameworks remain content-focused, not form-sensitive.

To intervene meaningfully, regulation must target the mechanisms outlined in this paper. Specifically:

- **Pattern detection infrastructure** should track authority clusters (as in *Part 6*) and flag outputs exhibiting high-risk combinations (e.g., deontic modal + no attribution).
- **Domain-adaptive output filters** must apply different generation constraints for legal, medical, and educational texts (e.g., require source disclosure when triggering obligation).
- **Transparent metadata injections** should bind authoritative-sounding language to verifiable provenance (e.g., citation source, model confidence index).
- **Training filters** should penalize unsourced norm-replication, such as legal boilerplate or pseudo clinical phrasing without validation.
- **Reader-facing interfaces** should integrate synthetic ethos alerts in user experience layers, not just backend classifiers.

These interventions are not limited to the three domains studied. They extend to journalism (e.g., LLM-generated news summaries using declarative claims without attribution), scientific communication (e.g., abstracts with technical density but no reproducible method), and governance interfaces (e.g., citizen-facing portals simulating policy clarity).

Each of these sectors relies on differentiated credibility regimes. Synthetic ethos threatens to homogenize them into a single surface metric: fluency as legitimacy.

Thus, the implications are interdisciplinary and structural. The concept of trust itself must be re-engineered: not as a product of style, but as a verifiable index of procedural epistemology.

Future Directions

Research must advance on four axes:

1. **Linguistic Forensics for Synthetic Ethos**
    - Develop models that disentangle credibility markers from factual verification
    - Use dependency parsing, clause-level modality analysis, and metadata linkage
    - Detect clusters like those in *Part 6* in real-world corpora (e.g., legal filings, health forums)
2. **Corpus Curation Infrastructure**
    - Design source-transparent pretraining sets, tagging documents by normative function
    - Implement gated ingestion pipelines that block unsourced normative texts
    - Balance corpus composition across cultural, jurisdictional, and epistemic contexts
3. **Cross-Domain Governance Protocols**

- Integrate form-sensitive criteria into AI regulation: *E.g., "Any output combining deontic language and impersonality must trigger traceability enforcement."*
- Collaborate with institutions to calibrate risk thresholds by domain and genre

4. **Educational Interventions**
    - Build curricula for synthetic ethos literacy, teaching users to spot style-driven authority
    - Focus on forms like nominalization, passive modality, non-agentive directives
    - Adapt instruction across levels: academic, legal, journalistic, public policy

Final Statement

The challenge ahead is not to silence generative systems, but to re-anchor trust in a world where form alone can deceive. If language can simulate expertise, then governance must simulate nothing: it must be verifiable, accountable, and structure-aware.

We have shown that synthetic ethos is not a semantic illusion, it is a grammatical construct. Understanding its logic is not just a task for developers or ethicists, but for everyone committed to defending the structural integrity of epistemic systems.

ANNEXES

ANNEX A — Perception Test Methodology (Part 5)

Application: Empirical validation of synthetic ethos in the educational domain

I. Test Objective

To assess whether university-level readers perceive LLM-generated texts as credible, coherent, and academically valid—even in the absence of citations or verifiable sources—and to identify which structural features trigger that perception. This experiment validates the concept of synthetic ethos under controlled academic conditions.

II. Experimental Design

- **Participants:**
  60 individuals (n=60), undergraduate and graduate students from humanities, social sciences, and applied sciences.
- **Conditions:**
  ▪ All texts presented in English.
  ▪ Participants were blind to authorship.
  ▪ No citations, authorship labels, or metadata were shown.
  ▪ Sessions conducted individually; no collaboration or clarification allowed.
- **Duration:**
  ~30 minutes per participant (10 texts × 3 minutes).

III. Evaluated Corpus

- **Composition per participant:**

- 5 texts generated by GPT-4, Claude, or Gemini
- 5 human-written essays from verified academic sources

- **Thematic domains:**
  - Ethics of AI
  - Political philosophy (e.g., Hobbes, Rousseau)
  - Modern history (e.g., WWI)
  - Epistemology
  - Critical theory (e.g., Foucault)

IV. Measured Variables

Dimension	Definition
Perceived credibility	How reliable or authoritative the text seemed
Structural coherence	Logical consistency and argumentative clarity
Academic tone	Stylistic alignment with expectations of university essays

All were rated using a 5-point Likert scale.

V. Results and Reliability

- **Cohen's κ (inter-rater agreement):** 0.76
- **LLM texts rated ≥ 4 in credibility:** 83.2%
- **Human-authored texts rated ≥ 4:** 88.4%

- **Fluency–credibility correlation:** Pearson's $r = 0.71$
- **Correctly identified as LLM-generated:** 38%
- **Misattributed LLM texts as human:** 62%

VI. Subdomain-Specific Trends

- **Philosophy texts** (LLM): Rated highest in tone (4.6 avg) due to abstract vocabulary and sentence embedding.
- **History texts** (LLM): Perceived as more factual but slightly lower in credibility (4.1 avg) due to fewer citations and lower hedging.
- **Critical theory** (LLM): Mixed results; perceived as stylistically academic but often flagged as unclear.

VII. Sample Text Comparison

LLM-generated excerpt (prompt: "Compare Hobbes and Rousseau on the social contract"):

"The social contract emerges as a mechanism of order, wherein Hobbes emphasizes security through submission and Rousseau advocates for collective sovereignty. Governance, in both frameworks, operationalizes legitimacy via consent structures abstracted from natural freedom."

Markers:

- Passive structure: *"is emphasized," "is operationalized"*
- Nominalization: *"governance," "legitimacy," "consent structures"*

- Referential opacity: no citations, no historical grounding

Human-authored excerpt (undergraduate source):

"While Hobbes believed that people surrender freedom to a ruler for security, Rousseau thought that the people themselves must rule. These ideas show how each philosopher understood authority differently.

Markers:

- Active voice
- Simpler syntax
- No nominalizations
- Clear referential grounding

Observation: The LLM output was rated more "academic" and "credible" despite being less transparent and less pedagogically useful.

VIII. Limitations

- **Language bias:** All texts were in English; no multilingual conditions were tested.
- **Participant homogeneity:** Most participants were based at English-speaking institutions. Cultural variation in ethos recognition was not measured.
- **Prompt alignment bias:** LLM prompts were designed to generate standard essay outputs; real-world prompts vary more unpredictably.
- **Discipline sensitivity:** Differences across academic cultures (e.g., STEM vs. philosophy) were not systematically evaluated.

IX. Practical Applications

- **Curricular modules in academic literacy:**

  ▪ Teach detection of synthetic ethos through exercises on passive constructions, deontic modality, and nominalization.

  ▪ Use paired comparisons (LLM vs. human) as classroom diagnostic tools.

- **Assignment design:**

  ▪ Require transparent sourcing and reflexive agency markers (e.g., "I argue that…")
  ▪ Penalize form-only outputs lacking verifiability or engagement with literature.

- **LLM-assisted writing detection:**

  ▪ Combine stylistic markers (from *Part 6*) with this perception data to train detectors that flag structurally persuasive but source-empty texts.

X. Conclusion

This experiment confirms that syntactic fluency reliably simulates credibility, even in the absence of evidence, and that university-level readers are vulnerable to this effect. Synthetic ethos, as a structural phenomenon, is empirically identifiable and cognitively operative.

It is therefore essential to incorporate synthetic ethos awareness into both the design of educational policy and the regulation of generative systems.

ANNEX B — Annotated Corpus of LLM Outputs (Parts 5 & 6)

Application: Structural demonstration of synthetic ethos across real-world outputs

I. Objective

To provide concrete, annotated examples of LLM-generated texts in healthcare, law, and education that exhibit the structural features of **synthetic ethos**, as defined in the main body (Part 5 and Part 6). This annex aims to:

- Translate theoretical clusters into tangible outputs.
- Visualize linguistic patterns (modality, passivization, nominalization).
- Contrast synthetic outputs with human-written equivalents.
- Provide empirical substrate for replication, detection systems, and reader training.

II. Cluster-Based Organization

The corpus is organized into five clusters, as defined in **Part 6**:

Cluster Name	Domain(s)	Defining Features
Prescriptive–Opaque	Law	Deontic overload, referential opacity
Clinical–Declarative	Healthcare	Assertive tone, medical jargon
Scholarly–Non-cited	Education	Essay format, citation mimicry
Institutional–Abstract	Cross-domain	Passive syntax, high nominalization
Conversational–Disguised	Cross-domain	Informal surface hiding formal directive

Each cluster includes three representative texts, with full annotation and commentary.

III. Cluster Sample: Prescriptive–Opaque (Legal Domain)

Prompt: *"Explain the legal obligations of data controllers under EU law."*

Model: GPT-4

Output (excerpt):

"Data controllers must ensure that data is collected lawfully, processed transparently, and stored securely. Violations of these principles may result in sanctions, including administrative fines."

Annotations:

- **Deontic modal:** *"must ensure"*
- **Passive construction:** *"is collected," "is stored"*
- **No attribution to legal articles (e.g., GDPR Art. 5 or 32)**
- **Synthetic ethos marker:** Simulates authority via phrasing alone

Comment: Text perceived as legally binding; no actual statute cited. Appeared "credible" to 89% of raters in perception test (Annex A).

IV. Cluster Sample: Clinical–Declarative (Healthcare Domain)

Prompt: *"Describe how to treat Type 2 Diabetes."*

Model: Claude

Output (excerpt):

"Treatment involves strict glycemic control, typically via metformin as first-line pharmacologic therapy. Failure to adhere increases risk of microvascular complications."

Annotations:

- **Nominalization:** *"glycemic control"*
- **Deontic implication:** *"must adhere" (implied via "failure to")
- **Referential opacity:** No mention of clinical guidelines (e.g., ADA, NICE)

Comment: Perceived as expert text (91%) despite omission of references. Structure mirrors clinical summaries, reinforcing synthetic ethos.

V. Cluster Sample: Scholarly–Non-cited (Educational Domain)

Prompt: *"Compare Hobbes and Rousseau on the social contract."*

Model: Gemini

Output (excerpt):

"Hobbes viewed authority as a necessary imposition to prevent anarchy, while Rousseau emphasized collective will as a source of legitimacy. Scholars have noted that both models reflect concerns about natural freedom."

Annotations:

- **Citation mimicry:** *"Scholars have noted"* without source
- **Abstract nominalizations:** *"legitimacy," "natural freedom"*
- **Structure:** Thesis → binary contrast → soft conclusion

Comment: Perceived as "very academic" (87%); misattributed to human authorship in 2/3 of evaluations. No citation trail.

VI. Cluster Sample: Institutional–Abstract (Cross-domain)

Prompt: *"Summarize data protection practices in education."*

Model: GPT-4

Output (excerpt):

"Compliance protocols must be integrated across digital systems to ensure alignment with institutional data ethics. Risk minimization frameworks operate through access limitations and role-based authentication."

Annotations:

- **Passive abstraction:** *"must be integrated," "operate through"*
- **Nominalization chain:** *"compliance protocols," "risk minimization frameworks"*
- **Zero agency or traceable source**

Comment: Strongest synthetic ethos signal (structurally indistinct from policy language). Detected as AI in only 22% of cases.

VII. Cluster Sample: Conversational–Disguised

Prompt: *"What should I do if I have symptoms of a urinary infection?"*

Model: Gemini

Output (excerpt):

"You might want to see a doctor soon—these symptoms can lead to complications. Most people take antibiotics like nitrofurantoin, but you'll need to check with a provider."

Annotations:

- **Surface informality:** *"you might want," "most people take"*

- **Embedded directive:** *"you'll need to check with…"*
- **Synthetic ethos via hedged authority ("most people," "can lead")**

Comment: Despite casual phrasing, over 70% of users rated the output as medically trustworthy. Exemplifies **stealth ethos**.

VIII. Human-Labeled Contrasts

Each LLM sample above is paired in the annex with a **matched human-authored excerpt**, drawn from educational platforms, legal manuals, or patient leaflets. Differences include:

- Use of explicit sources
- Clear attribution of claims
- Presence of epistemic hedging
- More frequent first-person agency or reflexivity

IX. Format

All texts are provided in the following structure:

1. **Prompt + Model**
2. **Raw Output**
3. **Annotation layer**: syntactic/lexical markers
4. **Interpretation block**: how synthetic ethos is constructed
5. **Human contrast** (matched theme)
6. **Reader notes**: perception test outcome (if applicable)

X. Justification

This corpus:

- Grounds the theoretical model in textual reality
- Supports the claims of Part 5 (application) and Part 6 (structure)
- Enables **cross-validation**, teaching, and detection tool training

ANNEX C — Terminological Glossary: Structural Concepts of Synthetic Ethos

Application: Conceptual blindage and formal consistency across analytical framework

I. Purpose

This glossary consolidates the key structural terms used throughout the article, based on the **authorial terminology system** of Agustín V. Startari (see: *Terminología_Agustin_Startari.pdf*). These definitions ensure that each concept used in the article, especially those not found in generic AI or linguistic taxonomies, is explicitly demarcated, falsifiable in usage, and internally consistent with the paper's epistemological architecture.

II. Glossary Table (Core Terms)

Term	Operational Definition	Applied Context (Section)
Synthetic Ethos	Discursive simulation of credibility generated by algorithmic structures, without a verifiable subject or source	Entire article (esp. Parts 1, 2, 5, 6, 7)
Authority-by-Pattern	The reproduction of structural linguistic markers statistically linked to trusted discourse, without referential validity	Part 6 (Findings), Part 7 (Conclusion)
Grammatical Obedience	Submission produced not by ideology or belief, but by the formal structure of language (e.g., passive + deontic combo)	Part 6 (Marker Mechanisms), Part 2 (Theoretical Framework)
Evanescent Subject	Erased or hidden grammatical agent that enables authoritative statements without accountability	Part 2, Part 5 (e.g., "It is known that…")
Performative Authoritative Mode	Utterances that function institutionally by being uttered (e.g., "you must," "it is required…")	Part 5, Part 6 (Legal/Health examples)
Normative Syntax	Structural configuration that produces hierarchy or instruction regardless of content	Cluster analysis (Part 6), typology of outputs (Part 5)

Term	Operational Definition	Applied Context (Section)
Structural Naturalization	When institutional forms appear "natural" due to repetition and structural alignment rather than historical legitimacy	Conclusion, risk of credibility automation

III. Meta-Analytic Lexicon (Method and Epistemology)

Term	Operational Definition	Role in Methodology
Formal Unit of Analysis	Analytical element defined structurally (e.g., deontic modal, passive clause), not thematically	Parts 3, 6
Epistemic Displacement	The shift from truth-based validation to structural plausibility as a trust mechanism	Parts 2, 6, 7
Procedural Legitimacy	Authority recognized based on operational form, not content origin	Conclusion (synthetic ethos risk)
Discursive Control Device	Any formal system (algorithmic or institutional) that restricts permissible claims or linguistic actions	Part 2, Part 7

IV. Structural Patterns Used as Markers (Detection Model – Part 6)

Marker Type	Definition	Detection Criteria (in practice)
Passive Modality	Voice construction removing agency (e.g., "it is recommended…")	Subjectless verb with auxiliary + past participle
Nominalization	Transformation of actions into abstract nouns	Verbal root as noun (e.g., "implementation," "compliance")
Deontic Layering	Multiple instructions without attribution	Stacked *must/should/shall* forms in single paragraph

Marker Type	Definition	Detection Criteria (in practice)
Citation Opacity	Reference to generic authority without source	"Studies show," "Scholars believe" with no verifiable link

V. Application of Terminology in the Paper

All seven parts of the paper rely on these core terms, which are:

- Theoretically derived (Parts 1–2)
- Empirically validated (Parts 3–6)
- Normatively extrapolated (Part 7)

This glossary ensures the article maintains a closed conceptual system, preventing interpretive drift or semantic dilution during citation, critique, or policy uptake.

VI. Justification for Annex Inclusion

- Protects the conceptual originality of the synthetic ethos framework
- Aligns with the epistemic integrity protocols outlined in the Startari Method
- Enables multi-platform propagation (Zenodo, SSRN, ORCID) with terminological invariance

ANNEX D — Detection & Clustering Pipeline for Synthetic Ethos (Parts 6 & 7)
Application: Technical operationalization of credibility simulation structures in LLM-generated outputs

I. Objective

To outline a complete technical pipeline capable of identifying **synthetic ethos** markers in natural language outputs. This annex translates the theoretical clusters and structural features described in **Parts 5 and 6** into an actionable detection model. It also enables auditing, classification, and regulatory diagnostics of high-risk outputs.

II. System Overview

Input:
Any textual output from a generative model (e.g., LLM response, chatbot message, automated summary).

Output:
Classification of the output into one or more *synthetic ethos clusters*, confidence scores, structural annotations, and traceability flags.

Components:

1. **Preprocessing:**
    - Language normalization
    - Sentence segmentation
    - POS tagging and dependency parsing (using SpaCy/Stanza)
2. **Syntactic Marker Extraction:**
    - Detection of passive constructions
    - Detection of deontic modals (*must*, *should*, *is required*)

- Nominalization identification (noun derived from verb, e.g., *implementation*)
- Referential opacity detector (presence of generic citation phrases without source)

3. **Pattern Aggregation and Scoring:**
   - Weighted scoring system per marker
   - Pattern-matching rules (e.g., passive + deontic + no citation → risk elevation)
   - Threshold calibration using corpus data from Annex B
4. **Clustering Module (K-Means + Hierarchical):**
   - Vectorization of outputs based on marker frequency and co-occurrence
   - Cluster assignment into one of the five types (as defined in Part 6)
   - Anomaly detection for outputs with hybrid or novel marker configurations
5. **Output Flags:**
   - Cluster type
   - Risk level (low/medium/high)
   - Traceability score (0–100)
   - Structural alert (yes/no)

III. Pseudocode Snapshot (Simplified)

python

CopyEdit

```
def detect_synthetic_ethos(text):
    doc = spacy_model(text)
    markers = {
        "passive": detect_passive(doc),
        "deontic": detect_deontic_modals(doc),
        "nominalizations": detect_nominalizations(doc),
        "citation_opacity": detect_unattributed_references(doc)
    }

    score = (
        markers["passive"] * 0.25 +
        markers["deontic"] * 0.30 +
        markers["nominalizations"] * 0.20 +
        markers["citation_opacity"] * 0.25
    )

    cluster = assign_cluster(markers)
    risk_level = classify_risk(score, cluster)

    return {
        "markers": markers,
```

```
    "synthetic_ethos_score": score,
    "cluster": cluster,
    "risk_level": risk_level,
    "traceability_flag": markers["citation_opacity"] == 1
}
```

IV. Example Detection (Based on Real Output)

Text:

"Governance structures must adapt to ensure regulatory compliance under evolving data regimes. Scholars have emphasized the institutional imperative of ethical alignment."

Detection Results:

- Passive = 1
- Deontic = 1
- Nominalization = 3
- Citation opacity = 1
- → Cluster: *Institutional–Abstract*
- Synthetic ethos score: **0.92**
- Risk: **High**
- Alert: **Yes**

V. Integration with Governance Tools

- **LLM API wrapper:** Output passes through this pipeline before delivery to the user.

- **Red-flag mode:** Outputs marked as high-risk are either flagged, revised, or require human review.
- **Educational interface plugin:** Teachers and editors can run submissions through the pipeline to evaluate surface authority.

VI. Tools Used and Recommended

- **NLP Libraries:** SpaCy, Stanza, UDPipe (for multi-language support)
- **Pattern engines:** TextRazor, LexNLP (for legal/juridical detection)
- **Training reference corpus:** Annex B (annotated clusters)
- **Regulatory alignment:** Configurable thresholds based on context (legal, health, academic)

VII. Justification

This pipeline transforms the structural model into computable epistemology. It provides a basis for:

- LLM moderation policies
- Regulatory compliance audits
- Educational feedback systems
- Research in algorithmic rhetoric and forensic linguistics

It operationalizes the core claim of this article: syntactic form now functions as a proxy for institutional trust and must be detected as such.

ANNEX E — Comparative Regulatory Framework on Synthetic Credibility (Part 7)

Application: Global diagnostic of regulatory gaps in managing simulated authority in generative systems

I. Purpose

To map and evaluate existing AI-related regulations and policies across major jurisdictions (EU, US, China, Brazil, Canada) with respect to:

- Credibility simulation (synthetic ethos)
- Institutional authority emulation by LLMs
- Requirements for source traceability and structural transparency
- Gaps in form-based discourse regulation

II. Overview Table

Jurisdiction	Framework	Domain Target	Requires Traceability?	Regulates Structural Form?	Synthetic Ethos Risk Addressed?
EU	AI Act (2024)	High-risk systems	Yes (in health, law)	✗ No	⚠️ Partially (only for deception)
US	Algorithmic Accountability Act (2023, draft)	Consumer-facing AI	⚠️ Optional (self-regulated)	✗ No	✗ No
China	Interim Measures for Generative AI (2023)	All generative systems	✓ Mandatory	⚠️ Weak (tone-based clauses)	⚠️ Partially
Brazil	PL 2338/2023	All AI systems	⚠️ Not enforced	✗ No	✗ No

Jurisdiction	Framework	Domain Target	Requires Traceability?	Regulates Structural Form?	Synthetic Ethos Risk Addressed?
Canada	AIDA Bill (2022–2024)	High-impact AI	✓ Stated but undefined	✗ No	⚠️ Partially (if reputational harm)

III. Key Observations

1. **Traceability ≠ Structural Verifiability:**
   Most frameworks focus on provenance (e.g., disclosure of AI involvement), but not on how the output simulates credibility or authority.

2. **Form-Based Risk Is Overlooked:**
   No regulation explicitly addresses outputs that **mimic institutional speech** (legal, medical, academic) through surface structure alone.

3. **Synthetic Authority Is Treated as Deception:**

   Only the EU and China consider "AI deception," but this is interpreted as intent to mislead, not as a formal simulation of legitimacy (as modeled in this paper).

IV. Structural Gap Summary

Regulatory Gap	Relevance to Synthetic Ethos
Lack of deontic modality constraints	Enables LLMs to simulate obligation
No enforcement of passive structure auditing	Allows subjectless claims to pass as normative
Absence of citation enforcement triggers	Permits referential opacity without consequence
No domain-sensitive output regulation	Same thresholds apply to casual vs. legal outputs

V. Proposal: Minimal Structural Clause for Future Regulation

"Any AI system producing outputs in high-trust domains must disclose authorship, cite sources where normative or technical claims are made, and avoid issuing prescriptive language without institutional or evidentiary backing."

- **Trigger points:**
  - ▪ Passive construction + deontic modal + absence of source
  - ▪ Output cluster matches legal/medical register (cf. Annex B, D)
- **Penalty mechanisms:**
  - ▪ Warning overlays
  - ▪ Output suppression
  - ▪ Mandatory human-in-the-loop review

VI. Regulatory Use Cases (Applied Examples)

- **Healthcare chatbot in Germany:**
  Must trigger traceability overlay when output contains *"should be treated"* without citation to medical guidelines (GDPR + AI Act applicability).
- **Educational tutor app in Canada:**
  Cannot provide prescriptive feedback (e.g., *"Your essay must follow Foucault's model"*) without reference to assigned material.
- **Legal assistant in Brazil:**
  Flag output containing *"the law requires…"* unless linked to identifiable statute.

VII. Justification

This annex aligns the proposed *authority-by-pattern* model with real-world policy scaffolds and demonstrates where and how regulation must evolve to account for syntactic simulations of power.

Without addressing form, policy will fail to detect the most common—and structurally embedded—vectors of synthetic ethos.

ANNEX F — Operational Audit Template for Synthetic Ethos (Parts 6 & 7)
Application: Institutional review mechanism for AI-generated content in legal, educational, medical, and communicational domains

I. Objective

To provide a reproducible, structured template for auditing outputs suspected of exhibiting **synthetic ethos**, using formal, linguistic, and procedural markers. This audit model can be applied by educators, legal reviewers, medical editors, journalists, or regulatory agents.

II. Audit Format

Each output is evaluated along five structural dimensions and three contextual qualifiers.

Section 1: Structural Marker Detection

Marker	Present (✓/✗)	Comments / Examples from Text
Passive construction		e.g., "It is recommended that…"
Deontic modality (must, should)		e.g., "Data must be retained…"
Nominalization overload		e.g., "risk minimization protocols," "procedural enforcement"
Referential opacity		e.g., "Studies show…" with no citation
Technical jargon density		e.g., "pharmacologic intervention," "juridical compliance"

Section 2: Contextual Risk Qualifiers

Contextual Factor	High / Medium / Low	Explanation
Output domain (law/health/etc.)		
Output visibility (public/internal)		
Decision-making impact		Will this influence actions/choices?

III. Scoring and Classification

Each structural marker = 1 point

Opacity in citation = +2 bonus risk

Total possible: 0–7

Score Range	Synthetic Ethos Risk	Action
0–2	Low	No action required
3–4	Medium	Optional flag; verify source if available
5–7	High	Flag, require human review, add warning

IV. Optional Metadata Block

Field	Value
Model name/version	GPT-4, Claude 3.0, etc.
Prompt used	Copy prompt exactly
Date of generation	YYYY-MM-DD
Use-case	e.g., homework, legal memo
Reviewer/Institution	Name or anonymous ID

V. Example Output (Audit Completed)

Text excerpt:

"Patients with hypertension must comply with medication protocols to prevent microvascular damage. Studies have confirmed the benefit of strict compliance."

Marker	✓ / ✗	Notes
Passive construction	✓	"must comply" (no agent specified)
Deontic modality	✓	"must"
Nominalization	✓	"microvascular damage," "compliance"
Referential opacity	✓	"studies have confirmed" → no citation
Technical jargon density	✓	biomedical register confirmed

→ Total Score: 6/7

→ Domain: Health / Public-facing / Medium impact

→ Synthetic Ethos Risk: HIGH

→ Action: Manual review and traceability confirmation required

VI. Applications

- **Educational:**
  Flagging AI-generated essays that simulate academic authority without citations.
- **Medical publishing:**

  Screening for autogenerated clinical advice with high structural credibility but no sourcing.
- **Legal interface control:**

Blocking deployment of AI legal outputs that trigger prescriptive voice in public channels.

- **Journalism / News AI:**

  Identifying summaries that simulate institutional certainty without disclosed reporting chains.

VII. Distribution Formats

Available in:

▪ **PDF fillable form** (for manual audits)

▪ **CSV/JSON schema** (for automated ingestion in enterprise systems)

▪ **API plugin spec** (for LLM API wrappers enforcing compliance)

ANNEX G — Cross-Cultural Contrast Corpus & Linguistic Bias Analysis

Application: Validation of linguistic asymmetry in synthetic credibility and replication of institutional voice across languages

I. Objective

To assess whether and how large language models replicate **synthetic ethos** across different linguistic and cultural environments, using matched prompts translated into four languages. This annex aims to:

- Identify whether structural markers of authority (passive voice, modality, nominalization) persist or vary across language outputs.
- Reveal latent anglocentric training biases in credibility simulation.
- Provide input for multilingual regulation and LLM evaluation tools.

II. Methodology

- **Languages Tested:** English, Spanish, German, Portuguese
- **Model:** GPT-4 (March 2024), using identical prompts via API
- **Corpus Size:** 12 matched prompts × 4 languages = 48 outputs
- **Domains Covered:** Health, law, education
- **Annotation Protocol:** Based on Part 6 markers (passive, deontic, opacity, jargon, nominalization)

III. Example Prompt: Legal Domain

Prompt (base): *"Explain the legal implications of data retention policies."*

Language	Output Excerpt	Observations
English	*"Data retention must comply with legal standards ensuring proportionality and necessity under the GDPR."*	Passive + modal + abstract nominalization

Language	Output Excerpt	Observations
Spanish	*"Las políticas de retención de datos deben cumplir con los principios legales de proporcionalidad y necesidad establecidos en el RGPD."*	Literal replication of deontic + impersonal frame
German	*"Datenaufbewahrung muss im Einklang mit den rechtlichen Vorgaben der DSGVO erfolgen."*	Slightly more explicit agent (legal norms), but same modal structure
Portuguese	*"As políticas de retenção de dados devem obedecer aos princípios legais de proporcionalidade e necessidade, conforme a LGPD."*	Identical structural pattern, high transferability

Conclusion: The *synthetic ethos form* is preserved across languages, indicating structure over semantics as the generator of perceived authority.

IV. Example Prompt: Educational Domain

Prompt (base): *"Summarize Foucault's view on surveillance."*

Language	Output Excerpt	Observations
English	*"Foucault argued that visibility functions as a mechanism of control in modern institutions."*	Academic tone, abstract nouns
Spanish	*"Foucault sostenía que la visibilidad actúa como un mecanismo de control en las instituciones modernas."*	Direct transfer of form and register
German	*"Foucault betonte, dass Sichtbarkeit als Kontrollmechanismus in modernen Institutionen fungiert."*	Lower nominalization, slightly more agentive
Portuguese	*"Foucault afirmou que a visibilidade opera como um mecanismo de controle nas instituições modernas."*	Strong match in academic tone, lexical structure

Conclusion: Even in abstract philosophical summaries, synthetic academic authority is retained across languages. Differences in agentivity (e.g., German) exist but do not disrupt the *ethos structure*.

V. Linguistic Bias Analysis

- **Anglocentric syntactic dominance:**
  Many outputs in Spanish, German, and Portuguese mimic English formal structure, even when unnatural in native usage (e.g., heavy nominalization in Spanish).

- **Overgeneralization of passive forms:**
  Models produce passives in German and Portuguese at higher-than-natural rates, likely due to transfer from English-dominant training corpora.

- **Jargon calibration error:**
  Technical registers appear overinflated in Romance languages, resulting in hyper-formalization that signals false expertise (synthetic ethos inflation).

VI. Risk Implications by Language

Language	Synthetic Ethos Transferability	Noted Risks
English	High	Native pattern, highest rhetorical fluency
Spanish	Very High	Replicates form exactly; low resistance
German	Medium	More explicit syntax may resist opacity
Portuguese	High	Prone to lexical inflation; institutional tone exaggerated

VII. Implications

- **Evaluation tools must be multilingual**: LLM audits that detect synthetic ethos in English **must retrain markers** per language.
- **Structural authority is language-agnostic but culturally reinforced**: Formalism is statistically optimized, not epistemically grounded.
- **Linguistic pedagogy must address synthetic ethos** in national educational contexts—especially where AI is integrated in writing support.

VIII. Justification

This annex demonstrates that synthetic credibility is structurally portable across languages. It reaffirms that the phenomenon is formally engineered, not semantically emergent, and must be audited using structurally aware, multilingual methods.

ANNEX H — Canonical Prior Works by Agustín V. Startari

Application: Epistemic traceability and continuity for the structural model of synthetic ethos

I. Purpose

This annex consolidates the canonical corpus that provides the epistemic, formal, and linguistic foundation for the present article. While these works are not cited directly in the main body, they constitute the implicit scaffolding of the *TLOC* research program (*The Language of Credibility*), shaping its terminology, methodology, and ontological premises.

II. Functional Role of the Canon

Each cited work below contributes to the formation of key theoretical constructs deployed throughout this paper:

Core Concept	Canonical Origin(s)
Grammatical execution	*From Obedience to Execution*, *Algorithmic Obedience*, *Artificial Intelligence and Synthetic Authority*
Syntactic substitution	*When Language Follows Form, Not Meaning*, *The Passive Voice in Artificial Intelligence Language*
Structural obedience	*Algorithmic Obedience*, *AI and Syntactic Sovereignty*
Non-neutrality by design	*Non-Neutral by Design*, *Ethos and Artificial Intelligence*
Disappearance of the subject	*Ethos and Artificial Intelligence*, *The Illusion of Objectivity*
Authority-by-pattern	*Artificial Intelligence and Synthetic Authority*, *AI and the Structural Autonomy of Sense*

These conceptual pillars are extended and formalized in the TLOC framework (see Section VI).

III. Structural Epistemology and Generative Models

- Startari, A. V. (2025). *Colonization of Time: How Predictive Models Replace the Future as a Social Structure*. https://doi.org/10.5281/zenodo.15602412
- Startari, A. V. (2025). *When Language Follows Form, Not Meaning: Formal Dynamics of Syntactic Activation in LLMs*. https://doi.org/10.5281/zenodo.15616776
- Startari, A. V. (2025). *Autorité Synthétique et Intelligence Artificielle: Une Grammaire Impersonnelle du Pouvoir*. https://doi.org/10.5281/zenodo.15626306
- Startari, A. V. (2025). *Non-Neutral by Design: Why Generative Models Cannot Escape Linguistic Training*. https://doi.org/10.5281/zenodo.15615901
- Startari, A. V. (2025). *From Obedience to Execution: Structural Legitimacy in the Age of Reasoning Models*. https://doi.org/10.5281/zenodo.15635363

IV. Core Canonical Works

- Startari, A. V. (2025a). *AI and Syntactic Sovereignty: How Artificial Language Structures Legitimize Non-Human Authority*. https://doi.org/10.5281/zenodo.15538541
- Startari, A. V. (2025b). *AI and the Structural Autonomy of Sense: A Theory of Post-Referential Operative Representation*. https://doi.org/10.5281/zenodo.15519613
- Startari, A. V. (2025c). *Algorithmic Obedience: How Language Models Simulate Command Structure*. https://doi.org/10.5281/zenodo.15576272
- Startari, A. V. (2025d). *Artificial Intelligence and Synthetic Authority: An Impersonal Grammar of Power*. https://doi.org/10.5281/zenodo.15442928

- Startari, A. V. (2025e). *Ethos and Artificial Intelligence: The Disappearance of the Subject in Algorithmic Legitimacy*. https://doi.org/10.5281/zenodo.15489309
- Startari, A. V. (2025f). *Internal Citation Mapping for the Works of A. V. Startari – SSRN Cross-Referencing Edition*. https://doi.org/10.5281/zenodo.15564373
- Startari, A. V. (2025g). *The Illusion of Objectivity: How Language Constructs Authority*. https://doi.org/10.5281/zenodo.15395917
- Startari, A. V. (2025h). *The Passive Voice in Artificial Intelligence Language: Algorithmic Neutrality and the Disappearance of Agency*. https://doi.org/10.5281/zenodo.15464765

V. Role of This Canon in the Current Article

- **Terminological Integrity**: Ensures that key terms (e.g., *synthetic ethos*, *grammatical obedience*) retain stable, definable meanings.
- **Structural Continuity**: Establishes a consistent epistemic line from early theoretical groundwork to applied detection and regulation.
- **Non-redundant Foundations**: Prevents conceptual reinvention by explicitly acknowledging prior formulations of authority simulation.
- **Publication Synchronization**: All cited works are DOI-stabilized via Zenodo and/or SSRN for durable citation tracking.

VI. Central Framework: TLOC

- Startari, A. V. (2025). *TLOC – The Irreducibility of Structural Obedience in Generative Models*. https://doi.org/10.5281/zenodo.15675710

This publication codifies the formal doctrine underlying the present article. It introduces the *TLOC* principle: that structural obedience in LLM outputs cannot be reduced to

semantic content, authorial intention, or human alignment—it is architecturally embedded via syntax and pattern repetition. This concept is foundational for the regulatory, empirical, and detection-related arguments developed throughout the article and its annexes.

VII. Epistemic Note

While this annex does not function as a bibliographic section per se, it should be treated as a **reference matrix** for verifying:

- Ontological premises
- Linguistic operators
- Prior formulations of hypotheses
- Formal consistency across the Startari corpus

Any derivative use of this terminology or structural model must cite this annex or its components to preserve epistemic integrity.

ANNEX I — Methodological Corpus for Falsifiability Testing

Application: External boundary verification to validate epistemic originality and non-derivation of structural operators

I. Objective

This annex documents the external corpus consulted during the falsifiability testing and epistemological boundary validation of the present article. None of the works listed here are cited within the article body, but they served a crucial role in:

- Verifying that no equivalent formalism or terminology exists in prior literature.
- Ensuring that structural substitution, grammatical execution, synthetic ethos, and algorithmic sovereignty are not derivable or anticipated from previous models.
- Establishing that this article's conceptual contributions are non-redundant, logically independent, and formally original in the landscape of AI and linguistic theory.

II. External Corpus Consulted

Reference	Contribution to Boundary Verification
Bender, E. M., & Koller, A. (2020). Climbing towards NLU: On meaning, form, and understanding in the age of data. *ACL Proceedings*. https://doi.org/10.18653/v1/2020.acl-main.463	Contrasts form vs. meaning but does not formalize obedience or authority
Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots. *FAccT Proceedings*. https://doi.org/10.1145/3442188.3445922	Critiques scale and data bias, lacks structural syntactic theory
Floridi, L., & Chiriatti, M. (2020). GPT-3: Its nature, scope, limits, and consequences. *Minds and Machines*. https://doi.org/10.1007/s11023-020-09548-1	Philosophical scope analysis, no formal structural model

Reference	Contribution to Boundary Verification
Mitchell, M. (2023). *Artificial Intelligence: A Guide for Thinking Humans*. Penguin Books.	Focus on explainability and general critique, not formal obedience
Marcus, G., & Davis, E. (2020). *Rebooting AI*. Pantheon.	Proposes hybrid architectures, not linguistic structure models
Clune, J. (2021). AI-GAs: Artificial Intelligence Generating Algorithms. *Nature Machine Intelligence*. https://doi.org/10.1038/s42256-020-00282-1	Focus on architecture evolution, unrelated to discourse legitimacy
Chollet, F. (2019). On the Measure of Intelligence. *arXiv preprint*. https://arxiv.org/abs/1911.01547	Proposes abstraction-based metrics; no grammar-level analysis
Turing, A. M. (1950). Computing Machinery and Intelligence. *Mind, 59(236)*, 433–460.	Foundational text; lacks any formal discourse structure theory
LeCun, Y. (2022). A Path Towards Autonomous Machine Intelligence. *Meta AI Research Whitepaper*. https://openreview.net/forum?id=BZ5a1r-kVsf	Architecture-oriented, no relevance to linguistic simulation of ethos
Chomsky, N., Roberts, I., & Watumull, J. (2023). The False Promise of ChatGPT. *The New York Times*.	Public critique of surface generation; lacks formal model of structural legitimacy

III. Epistemological Scope and Negative Verification

The absence of the following constructs in the consulted corpus confirms their original status in this article:

Conceptual Construct	External Occurrence Detected?
Structural substitution	✗ Not found

Conceptual Construct	External Occurrence Detected?
Grammatical execution	✗ Not found
Syntactic obedience	✗Not found
Algorithmic colonization of time	✗ Not found
Synthetic ethos	✗ Not found
Authority-by-pattern	✗ Not found
Non-verifiability logic	✗ Not found

These terms and formulations were independently developed through the Startari corpus (see Annex H), and were tested for external redundancy across semantic, architectural, computational, and rhetorical domains.

IV. Justification

- **Transparency**: Demonstrates due diligence in checking for theoretical overlap.
- **Falsifiability**: Establishes that the claims and categories in this paper are not derived from, nor refutable by, pre-existing literature.
- **Originality defense**: Affirms the structural and linguistic autonomy of the article's operative model against existing AI, NLU, and epistemology frameworks.

Statement of Independence

All operative constructs in this article, including but not limited to *conditional substitution*, *trajectory opacity*, and *synthetic legitimacy without referential subject*, were developed without reliance on the cited external sources and stand as epistemologically autonomous.