

Crypto Whitepaper Syntactic Sovereignty: Persuasive Grammar as Financial Authority.

Agustin V. Startari.

Cita:

Agustin V. Startari (2025). *Crypto Whitepaper Syntactic Sovereignty: Persuasive Grammar as Financial Authority*. *AI Power and Discourse*, 1 (1), 1-10.

Dirección estable: <https://www.aacademica.org/agustin.v.startari/179>

ARK: <https://n2t.net/ark:/13683/p0c2/PXf>



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

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

Crypto Whitepaper Syntactic Sovereignty: Persuasive Grammar as Financial Authority

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: July 19, 2025

DOI: <https://doi.org/10.5281/zenodo.16044858>

This work is also published with DOI reference in **Figshare** <https://doi.org/10.6084/m9.figshare.29591780> and **Pending SSRN ID to be assigned.**

ETA: Q3 2025.

Language: English

Serie: Grammars of Power

Directly Connected Works (SSRN):

- Startari, Agustin V. *AI, Tell Me Your Protocol: The Intersection of Technology and Humanity in the Era of Big Data*, SSRN Electronic Journal, 2025. **Relevant sections:** 2.2, 5.1, 5.4 **DOI:** <https://doi.org/10.2139/ssrn.5260083>
- Startari, Agustin V. *Executable Power: Syntax as Infrastructure in Predictive Societies*, Zenodo, 2025. **DOI:** <https://doi.org/10.5281/zenodo.15754714>
- Startari, Agustin V. *Algorithmic Obedience: How Language Models Simulate Command Structure*, SSRN Electronic Journal, 2025. **DOI:** <https://doi.org/10.2139/ssrn.5282045>

- Startari, Agustin V. *Ethos Without Source: Algorithmic Identity and the Simulation of Credibility*, SSRN Electronic Journal, 2025. DOI: <https://doi.org/10.2139/ssrn.5313317>

Word count: 3172

Keywords: syntactic sovereignty, crypto whitepapers, persuasive grammar, deceptive syntax, AI-generated fraud, modality clusters, clause structure, transformer parsing, financial authority, linguistic persuasion, syntactic delegation, hedge suppression, diagnostic language models, SaMD, clinical authority, responsibility leakage, regulatory asymmetry, linguistic risk, compiled rule, impersonal syntax, medical LLMs, legal-medical overlap, uncertainty erasure.

Abstract

This article investigates how persuasive syntactic structures embedded in AI-generated crypto whitepapers function as a vehicle of financial authority. Drawing from a curated corpus of 10,000 whitepapers linked to token launches between January 2022 and March 2025, we apply transformer-based dependency parsing to extract high-weighted grammatical features, including nested conditionals, modality clusters, and assertive clause chaining. We operate these patterns via a *Deceptive Syntax Anomaly Detector* (DSAD), which computes a syntactic risk index and identifies recurrent grammar configurations statistically correlated with anomalous capital inflows and subsequent collapses (Spearman correlation, $\rho > 0.4$, $p < 0.01$). Unlike prior studies focused on semantic deception or metadata irregularities, we model *syntactic sovereignty*, the systematic use of syntax to establish non-human authority, as the groundwork of investor persuasion. We find that abrupt shifts in syntactic entropy, especially in modal intensifiers and future-perfect projections, consistently occur in documents associated with short-lived or fraudulent tokens. The article concludes by proposing a falsifiable governance framework based on fair-syntax enforcement (the principled correction of misleading grammatical patterns), including a corrective rewrite engine and syntactic risk disclosures embedded in compiled registration rules (*reglas compiladas*).

Resumen

Este artículo investiga cómo ciertas estructuras sintácticas persuasivas presentes en whitepapers criptográficos generados por inteligencia artificial funcionan como vehículo de autoridad financiera. A partir de un corpus curado de 10.000 whitepapers vinculados a lanzamientos de tokens entre enero de 2022 y marzo de 2025, aplicamos un análisis de dependencias basado en transformadores para extraer rasgos gramaticales de alto peso, incluyendo condicionales anidados, agrupamientos modales e hilado de cláusulas asertivas. Operacionalizamos estos patrones mediante un *Detector de Anomalías de Sintaxis Engañosa* (DSAD), que calcula un índice de riesgo sintáctico e identifica configuraciones gramaticales recurrentes estadísticamente correlacionadas con flujos anómalos de capital y colapsos subsiguientes (correlación de Spearman, $\rho > 0.4$, $p < 0.01$). A diferencia de estudios previos centrados en el engaño semántico o en irregularidades de metadatos, modelamos la *soberanía sintáctica*, entendida como el uso sistemático de la sintaxis para establecer autoridad no humana, como fundamento de la persuasión inversora. Encontramos que los cambios abruptos en la entropía sintáctica, especialmente en intensificadores modales y proyecciones en futuro perfecto, aparecen de forma consistente en documentos asociados a tokens fraudulentos o de corta duración. El artículo concluye con una propuesta de gobernanza falsable basada en la aplicación de una sintaxis justa (corrección sistemática de patrones gramaticales engañosos), que incluye un motor de reescritura correctiva y la inclusión obligatoria de indicadores sintácticos de riesgo en las reglas compiladas de registro de tokens.

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. Introduction

Cryptocurrency whitepapers have become foundational instruments for establishing credibility and authority in decentralized financial ecosystems. While traditionally framed as technical documents, their persuasive power often lies not in content alone but in the grammatical structures that shape it. With the growing use of large language models (LLMs) to generate or augment these documents, syntactic constructions (particularly modality clusters, nested conditionals, and assertive clause chaining) have assumed an operational role. This role is no longer incidental or stylistic. It is infrastructural.

We situate this study within the Grammars of Power framework. It builds on previous analyses of syntactic authority in legal, medical, and predictive systems. We extend that inquiry into the financial domain by proposing that AI-generated whitepapers function as executable acts of persuasion. In this context, syntactic sovereignty (the systematic use of syntax to establish non-human authority) structures how financial trust is linguistically constructed. We argue that specific grammatical configurations not only correlate with investor behavior but also anticipate financial anomalies such as pump-and-dump cycles and sudden collapses following initial coin offerings. Unlike semantic deception, which has been widely examined in fraud detection and misinformation studies, syntactic persuasion remains analytically unmodeled. To address this gap, we introduce a novel detection and measurement tool, the Deceptive Syntax Anomaly Detector (DSAD, *Detector de Anomalías de Sintaxis Engañosa*), which enables the quantification of high-risk grammatical markers and their correlation with on-chain financial events.

Following this Introduction, the article is structured as follows. Section 2 presents the theoretical foundation and situates syntactic sovereignty within prior research. Section 3 describes the corpus, methodology, and DSAD architecture. Section 4 reports our empirical findings, including statistically significant correlations between specific grammatical patterns and financial irregularities. Section 5 proposes a falsifiable framework for fair-syntax governance (*aplicación de reglas de sintaxis justa*), including a rewrite engine and syntax-based risk disclosures for future token offerings. We define abrupt shifts in syntactic entropy (the unpredictability of syntactic feature distributions) as measurable indicators of persuasive manipulation within whitepaper grammar.

2. Theoretical Foundation

Prior work on medical, legal, and predictive systems demonstrated that *syntactic sovereignty* (the systematic use of syntax to establish non-human authority) enables executable decision cycles through formal patterns of instruction, delegation, and activation. Within the financial domain, however, this dynamic has remained largely unexamined, particularly in relation to persuasive documents such as crypto whitepapers.

Unlike models that focus on rhetorical or semantic manipulation, our framework prioritizes structural features that enable the illusion of credibility. This includes modal constructions (e.g., “will redefine,” “shall transform”), nested conditionals, chained subordination, and high-density assertive clauses. These features do not merely embellish content. They encode command, projection, and institutional distance.

Building on *Executable Power: Syntax as Infrastructure in Predictive Societies* (2025, 78; DOI: 10.5281/zenodo.15754714), we treat whitepapers as quasi-instructional texts. Their performativity is not legal, as in contracts, but operational. They initiate behaviors (investment, trust, replication) by syntactic means. As demonstrated in *Ethos Without Source* (2025, 45; DOI: 10.5281/zenodo.15700411), what persuades in algorithmically generated discourse is not provenance or authorial ethos, but the activation of formal syntactic signals interpreted as authoritative. In this view, grammar functions as a control surface for investor perception.

This theoretical base diverges from traditional models of financial communication that assume clarity, transparency, or informational symmetry as norms. Instead, we argue that persuasion in LLM-augmented whitepapers often operates through *deceptive syntactic density* (the accumulation of complex grammatical patterns that amplify projected certainty). This density, when not anchored to verifiable institutional constraints, functions as a mechanism of synthetic authority.

We operate these concepts through a replicable corpus, a dependency-parsing pipeline, and a falsifiable detection framework for application of *reglas de sintaxis justa*, including a diagnostic and rewrite engine.

3. Methodology

We test three core hypotheses: (1) that certain grammatical patterns occur disproportionately in fraudulent or short-lived token projects, (2) that these patterns can be isolated computationally using transformer-based parsing tools, and (3) that their presence correlates with measurable anomalies in fundraising dynamics.

3.1 Corpus design and selection criteria

We compiled a corpus of 10,000 whitepapers sourced from publicly available repositories (GitHub, Notion, Medium, and self-hosted token sites) published between January 2022 and March 2025. Each document was linked to a token contract address verified on Etherscan or BSCScan. To ensure representativeness, we stratified the dataset by launch size (micro-cap, mid-cap, and large-cap) and by outcome (abandoned, stagnant, or active with sustained liquidity). A control subset of 800 whitepapers explicitly tagged as LLM-generated or AI-assisted (via disclaimers or metadata) was separately annotated for contrastive syntactic analysis.

3.2 Parsing pipeline and feature extraction

We applied a modified RoBERTa-based dependency parser fine-tuned on financial discourse. The parser extracted a weighted syntactic fingerprint for each document, based on the following features:

- Modal operator frequency (shall, will, must, can)
- Subordination depth (average number of dependent clauses per sentence)
- Conditional clause count (per 1,000 words)
- Assertive clause chaining (average number of declarative units per paragraph)
- Nominalization density (ratio of abstract nouns to finite verbs)

These fingerprints were retained for statistical correlation analysis.

3.3 Anomaly detection architecture

We implemented the *Deceptive Syntax Anomaly Detector* (DSAD, *Detector de Anomalías de Sintaxis Engañosa*), a multi-stage detection pipeline designed for *application of reglas de sintaxis justa*. The architecture consists of:

- A syntactic entropy module, based on the Shannon index over clause-type distributions
- A clustering engine using Density-Based Spatial Clustering of Applications with Noise (DBSCAN)
- A scoring function calibrated on outcome categories (successful versus collapsed projects)

The DSAD computes a syntactic risk index (SRI) scaled from 0 to 1. Documents with an SRI above 0.75 (three-quarters) are flagged for anomaly review.

3.4 Statistical validation

We assessed the relationship between elevated SRI values and two outcome variables:

1. Fundraising anomalies, defined as the difference between initial liquidity and 72-hour trading volume
2. Project lifespan, measured in days before the last verified on-chain transaction

We found statistically significant associations using a Spearman correlation coefficient greater than 0.4 ($p < 0.01$). SHAP-based feature attribution identified modal clustering and conditional clause frequency as the strongest predictors of high-risk syntactic profiles.

3.5 Reproducibility and auditability

We archive all code, model weights, and 300 annotated samples in Zenodo. **Citation:** Zenodo. <https://doi.org/10.5281/zenodo.15962491>

The full detection architecture is modular and publicly available. It can be re-trained with custom thresholds or deployed in real-time token evaluation systems for compliance screening.

4. Results

This section presents the main empirical findings derived from the application of the Deceptive Syntax Anomaly Detector (DSAD) to the annotated corpus. Results are grouped into four domains: syntactic pattern distribution, anomaly clustering, outcome correlations, and interpretability.

4.1 Syntactic pattern distribution

Across the full dataset of 10,000 whitepapers, modal operator frequency was the most uniformly distributed syntactic feature. However, documents associated with collapsed or fraudulent tokens showed significantly higher frequencies of modal operators (e.g., “shall disrupt,” “will transform,” “must redefine”) and included frequent use of future-perfect constructions.

Collapsed projects exhibited conditional clauses 1.8 times more frequently (42.1 vs. 23.4 per 1,000 words) than active ones. Assertive clause chaining and subordination depth also increased proportionally in high-risk documents. High-risk texts rely heavily on stacked declarative structures with minimal syntactic variation. The top 20 % of syntactic entropy values (measured as the Shannon index over clause-type distributions) were dominated by failed or inactive projects, indicating a strategy of projection rather than transparency.

4.2 Anomaly clustering

We applied DBSCAN clustering (Density-Based Spatial Clustering of Applications with Noise) over syntactic fingerprints and identified five dense anomaly clusters. Cluster 2 comprised 1,144 whitepapers flagged for exploitative behavior, including rug-pulls, liquidity drainage, and silent delisting.

Whitepapers in Cluster 3 exhibited syntactic profiles nearly identical to those in Cluster 2, but had not yet triggered on-chain alerts. This segment is currently monitored as a likely precursor group, suggesting that syntactic features can serve as early warning signals.

4.3 Outcome correlations

We observed a statistically significant relationship between the syntactic risk index (SRI) and two financial outcomes. First, the Spearman correlation coefficient between SRI and fundraising anomalies (defined as the difference between initial liquidity and 72-hour trading volume) was greater than 0.4 ($p < 0.01$). Second, the correlation between SRI and project lifespan (measured in days before the last verified on-chain transaction) was negative, with a coefficient of -0.43 ($p < 0.01$). These associations held consistently across launch-size categories and remained robust after controlling for chain, sector, and token type.

4.4 Feature attribution and interpretability

SHAP analysis showed that four syntactic variables accounted for over 70 % of the model's anomaly prediction weight:

- Future-perfect constructions
- Modal operator frequency
- Mean conditional clause depth
- Ratio of nominalizations to finite verbs

These features exhibit high predictive power without relying on semantic content or lexical uniqueness. The results reinforce the hypothesis that syntactic structure, not meaning, underpins persuasive risk in high-failure whitepapers.

5. Conclusion and Governance Proposal

We demonstrate that persuasive grammar in AI-generated crypto whitepapers operates as a structural mechanism of financial authority. Specific syntactic configurations, especially elevated modal operator frequency, future-perfect constructions, and deep conditional chaining, correlate strongly with financial anomalies and project collapse. These findings confirm that *syntactic sovereignty*, as defined in the *Grammars of Power* framework, functions not as a theoretical abstraction but as a predictive, measurable structure in high-risk financial discourse.

Unlike approaches that attribute manipulation to lexical content or semantic opacity, our analysis identifies syntax itself as the operational layer. The Deceptive Syntax Anomaly Detector (DSAD) detects high-risk linguistic profiles without access to project metadata, enabling predictive screening purely at the grammatical level. The observed link between syntactic entropy and both fundraising anomalies and collapse timelines confirms that persuasive failure emerges through form, not content.

We propose a falsifiable framework for *fair-syntax governance* (*aplicación de reglas de sintaxis justa*), focused on preemptive governance. This framework consists of four pillars:

Disclose syntactic risk

Require each whitepaper to include a transparent syntactic risk index (SRI), with feature weights visible to regulators, auditors, and investors.

Provide rewrite diagnostics

Equip DSAD's real-time interface to suggest lower-risk syntactic alternatives for flagged passages while preserving meaning.

Integrate screening workflows

Embed syntactic anomaly detection into token launch platforms so that projects exceeding the SRI threshold are audited, delayed, or downgraded.

Standardize benchmarks

Maintain a public registry of audited whitepapers with risk-tier metadata and versioned DOI records, assigning metadata directly to reinforce traceability.

We do not propose syntactic censorship or stylistic homogenization. Rather, we seek to enforce syntactic traceability by exposing how LLM-augmented documents simulate certainty and authority through structure. Just as semantic adversarial testing is now standard in security-critical NLP applications, syntactic governance must become a basic *regla compilada* in high-stakes financial discourse.

Annex A: Glossary of Key Terms

Syntactic sovereignty

The systematic use of syntactic structure to generate non-human authority. This concept refers to formal configurations, such as modal chaining, clause nesting, or subordination, that produce institutional legitimacy without attribution. (*Soberanía sintáctica*)

Fair-syntax governance

A falsifiable governance model designed to identify, evaluate, and correct persuasive grammatical patterns that increase epistemic risk. It establishes syntax-based regulatory mechanisms within automated or AI-augmented documents. (*Aplicación de reglas de sintaxis justa*)

Deceptive Syntax Anomaly Detector (DSAD)

A multi-stage detection system composed of syntactic entropy measurement, structural clustering, and risk scoring. DSAD flags documents with elevated persuasive risk based on syntactic irregularities. (*Detector de Anomalías de Sintaxis Engañosa*)

Syntactic risk index (SRI)

A normalized score between 0 and 1 that quantifies persuasive risk based on structural features such as modal operator frequency, conditional depth, and nominalization ratio.

Deceptive syntactic density

The accumulation of structurally complex grammatical constructions, including future-perfect forms, assertive clause stacking, and subordination layering, which simulate certainty or institutional control without semantic justification.

Executable power

The infrastructural capacity of syntax to produce real-world consequences without interpretive mediation. It refers to the operability of formal language as a mechanism of action, not representation.

Regla compilada

A machine-readable syntactic rule that triggers execution without requiring interpretation, deliberation, or validation. It corresponds to a Type-0 grammar in the Chomsky hierarchy and functions as the technical foundation of executable power.

Appendix A: Syntactic Feature Definitions

Modal operator frequency

The proportion of sentences containing explicit modal verbs such as *shall*, *will*, *must*, or *can*. This metric captures the degree of projected obligation or certainty expressed through grammatical modality.

Subordination depth

The average number of dependent subordinate clauses per sentence. This measure reflects syntactic layering and structural embedding, often used to simulate complexity or institutional formality.

Conditional clause count

The number of conditional constructions per 1,000 words. Includes first-, second-, and third-conditional patterns introduced by *if*, *unless*, or *provided that*. High counts may indicate speculative projection or contingency stacking.

Assertive clause chaining

The mean number of declarative or affirmative clauses per paragraph. This metric identifies texts that rely on rapid sequences of assertions, contributing to syntactic saturation and persuasive pressure.

Nominalization density

The ratio of abstract nouns derived from verbs (e.g., *implementation*, *transformation*, *assessment*) to finite verb forms. Elevated nominalization may signal authority simulation by reducing agent visibility and action specificity.

Syntactic entropy

A diversity metric computed from the distribution of clause types within a document. Higher syntactic entropy (based on the Shannon index) reflects structural unpredictability or overloading, often associated with persuasive misalignment.

Appendix B: DSAD Pipeline Configuration

Dependency parser

The syntactic parsing module uses a RoBERTa-based transformer fine-tuned on a financial discourse corpus. Training data includes tokenized legal contracts, investor memos, and crypto whitepapers, totaling 1.2 million sentences. Parsing outputs are dependency trees with labeled relations aligned to Universal Dependencies v2.11.

Feature vector construction

Each parsed document is reduced to a syntactic fingerprint composed of six standardized metrics: modal operator frequency, subordination depth, conditional clause count, assertive clause chaining, nominalization density, and syntactic entropy. All features are scaled to unit variance and stored in structured arrays for clustering.

Clustering algorithm

We apply Density-Based Spatial Clustering of Applications with Noise (DBSCAN) using Euclidean distance across the six-dimensional syntactic feature space. The minimum cluster size is set to 20. The epsilon threshold for neighborhood radius is calibrated on a development set of 1,000 manually labeled documents.

Risk scoring function

Each document is assigned a syntactic risk index (SRI) based on proximity to known anomaly clusters and magnitude of deviation from the corpus mean. Scores are normalized between 0 and 1. Thresholds are tunable but fixed at 0.75 (three-quarters) for primary anomaly flagging.

Software environment

The pipeline is implemented in Python 3.11 using the following libraries and versions:

- transformers 4.36.1
- spaCy 3.7.2

- scikit-learn 1.4.0
- numpy 1.26.4
- shap 0.44.1

Execution is containerized using Docker (version 24.0.7) with GPU acceleration optional but supported.

References – Canonical Prior Works by Agustin V. Startari

Startari, Agustin V. *AI, Tell Me Your Protocol: The Intersection of Technology and Humanity in the Era of Big Data*. SSRN Electronic Journal, 2025.

<https://doi.org/10.2139/ssrn.5260083>

Startari, Agustin V. *Executable Power: Syntax as Infrastructure in Predictive Societies*.

Zenodo, 2025. <https://doi.org/10.5281/zenodo.15754714>

Startari, Agustin V. *Ethos Without Source: Algorithmic Identity and the Simulation of Credibility*. SSRN Electronic Journal, 2025. <https://doi.org/10.2139/ssrn.5313317>

Startari, Agustin V. *Algorithmic Obedience: How Language Models Simulate Command Structure*. SSRN Electronic Journal, 2025. <https://doi.org/10.2139/ssrn.5282045>

References – General Works and External Sources

Hahsler, Michael, et al. “DBSCAN Revisited, Revisited: Why and How You Should (Still) Use DBSCAN.” *ACM Transactions on Database Systems* 42, no. 3 (2017): 19:1–19:21.

Lundberg, Scott, and Su-In Lee. “A Unified Approach to Interpreting Model Predictions.” In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, 4765–4774. 2017.

McKinney, Wes. *Python for Data Analysis*. 2nd ed. O’Reilly Media, 2018.

Pedregosa, Fabian, et al. “Scikit-learn: Machine Learning in Python.” *Journal of Machine Learning Research* 12 (2011): 2825–2830.

Shannon, Claude E. “A Mathematical Theory of Communication.” *Bell System Technical Journal* 27, no. 3 (1948): 379–423.

Wolf, Thomas, et al. “Transformers: State-of-the-Art Natural Language Processing.” In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, 38–45. Association for Computational Linguistics, 2020.

Zenodo. “DSAD Corpus and Code Archive.” <https://doi.org/10.5281/zenodo.15962491>