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

Clinical Syntax: Diagnoses Without Subjects in AI-Powered Medical Notes.

Agustin V. Startari.

Cita:

Agustin V. Startari (2025). *Clinical Syntax: Diagnoses Without Subjects in AI-Powered Medical Notes. AI Power and Discourse, 1 (1), 1-10.*

Dirección estable: https://www.aacademica.org/agustin.v.startari/203

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

Clinical Syntax: Diagnoses Without Subjects in AI-Powered Medical Notes

Author: Agustin V. Startari

Author Identifiers

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

Institutional Affiliations

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

Contact

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

Date: September 23, 2025

DOI

- Primary archive: https://doi.org/**10.5281/zenodo.17184301**
- Secondary archive: https://doi.org/10.6084/m9.figshare.30187882
- SSRN: Pending assignment (ETA: Q3 2025)

Language: English

Series: *AI Syntactic Power and Legitimacy*

Word count: 6156

Keywords: Large Language Models; Plagiarism; Idea Recombination; Knowledge Commons; Attribution; Authorship; Style Appropriation; Governance; Intellectual Debt; Textual Synthesis; ethical frameworks; juridical responsibility; appeal mechanisms; syntactic ethics; structural legitimacy, Policy Drafts by LLMs, linguistics, law, legal, jurisprudence, artificial intelligence, machine learning, llm.

Abstract

This article examines the structural erasure of the patient as an active subject in clinical records generated by artificial intelligence systems. Automated outputs from Epic Scribe, GPT-4, and institutional medical note generators increasingly rely on impersonal constructions, nominalizations, and fragmented clauses that displace the patient from the syntactic center of medical discourse. The shift toward objectified formulations such as "bilateral opacities noted" rather than "the patient presents with" produces a discourse where agency and responsibility are structurally absent. Building on prior analyses of passive voice and subject deletion, the study introduces the *Syntactic Opacity Index* (SOI) as a formal measure to quantify the density of non-agentive structures in AI-authored notes. The corpus analysis demonstrates how opacity accumulates at the sentence level, rendering the clinical narrative less transparent and more difficult to attribute. Beyond linguistic critique, the article assesses the ethical and epistemic consequences of syntactic opacity in medicine, particularly regarding accountability, patient-centered care, and institutional responsibility. The findings suggest that AI-powered medical documentation does not merely accelerate administrative workflows but also reconfigures the grammar of care itself, demanding urgent attention to how language structures shape both diagnosis and responsibility.

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 *AI Syntactic Power and Legitimacy*. 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

Medical writing has always carried a dual function: it registers empirical data while simultaneously constructing a discursive framework where the patient appears as a subject of care. In traditional clinical practice, the patient is not merely an object of examination but also the grammatical anchor around which observation, diagnosis, and therapeutic decision are organized. The act of documenting symptoms, interventions, or responses requires a syntactic arrangement in which agency and subjectivity remain visible, even if attenuated by conventions of neutrality.

The incorporation of artificial intelligence into clinical documentation disrupts this historical balance. Systems such as Epic Scribe and large language models including GPT-4 are now tasked with generating entire sections of medical notes without direct human authorship. These systems overwhelmingly rely on impersonal grammatical forms, producing outputs that omit the patient as an active subject. Instead of "The patient presents with signs of pneumonia," the generated sentence reads "Findings consistent with pneumonia." The difference is not trivial: the first construction presupposes a subject of experience, while the second displaces experience into an impersonal field of "findings." Such displacements accumulate across the clinical record, creating what may be described as a syntactic erasure of the patient.

The theoretical concern is not limited to stylistics. Previous research has analyzed the political and epistemic consequences of passive constructions and subject deletion in artificial intelligence language models. Startari (2024) demonstrated that the passive voice systematically removes agency, replacing actors with formalized structures of observation. This analysis showed that large language models simulate authority by neutralizing the subject, thereby establishing what he called "obedience without an agent." Building on this framework, *Syntax Without Subject* further explored how automated texts create authority by removing the very possibility of attributing responsibility. Within this continuum, clinical notes generated by AI represent a particularly urgent case: the erasure of the subject does not only distort meaning, it alters how institutional medicine accounts for responsibility.

The objective of the present article is threefold. First, it identifies the syntactic strategies by which AI systems eliminate the patient from the textual scene. These strategies include the dominance of impersonal passives, the replacement of verbs by nominalizations, and the fragmentation of diagnostic statements into isolated descriptive phrases. Second, it proposes a formal measure, the *Syntactic Opacity Index* (SOI), designed to quantify the degree of subject erasure in automated records. The SOI allows comparison between human-authored and AI-generated notes, establishing a scale of opacity that can be applied in institutional audits. Third, it examines the ethical and epistemic implications of syntactic opacity. If the patient no longer appears in the sentence, how is care understood, and who bears responsibility for what is recorded?

The article situates this inquiry within the broader discussion of AI in institutional medicine. Scholars have noted the efficiency gains promised by automation, but less attention has been given to the linguistic mechanisms through which efficiency is achieved. By removing agents and collapsing sentences into impersonal fragments, AI reduces the cost of narrative construction. Yet the cost of efficiency is borne in the erosion of accountability. If no subject appears in the note, then no subject is accountable for its content. This dynamic directly affects not only the relation between doctor and patient but also the relation between institution and responsibility.

This introduction therefore positions the problem as one of structural linguistics and ethical governance. The challenge is not to denounce technology in abstract terms, but to demonstrate how specific syntactic forms reorganize the conditions of medical practice. By analyzing a corpus of notes produced by Epic Scribe and GPT-4, the article argues that AI-powered documentation institutes a new grammar of care, one in which the patient is rendered invisible as subject. The implications extend beyond clinical communication: they reveal how language itself is restructured when responsibility is mediated by algorithms.

The sections that follow will expand on this framework. A review of the theoretical context will situate the discussion within linguistic and medical traditions. The methodology will formalize the corpus and define the metrics used. The central analysis will identify and classify patterns of subject erasure, followed by the introduction of the SOI. Finally, the

ethical and epistemic consequences will be discussed, before concluding with a proposal for integrating transparency safeguards into institutional medical AI systems.

In sum, the introduction outlines a structural transformation at the intersection of language, medicine, and artificial intelligence. Clinical notes have become a site where the absence of the subject is not incidental but designed. The disappearance of the patient from syntax signals a deeper displacement of responsibility from the human sphere to the computational. Addressing this shift requires a framework capable of linking linguistic form, institutional practice, and ethical accountability.

2. Background and Theoretical Context

The emergence of automated medical documentation must be situated against a broader historical background in which the function of clinical writing has oscillated between two poles: scientific neutrality and subjective accountability. From the earliest case histories in Hippocratic medicine to the codified protocols of modern hospitals, clinical discourse has combined the need for objectivity with the recognition of the patient as an embodied subject. This balance has always been precarious. On the one hand, medical institutions privilege forms that minimize ambiguity, favoring standardized vocabularies and diagnostic categories. On the other hand, the act of writing a note has traditionally required a physician or scribe to anchor the narrative in a subject, namely the patient, whose symptoms and experiences provide the very material of observation.

Throughout the twentieth century, medical documentation increasingly adopted depersonalized styles. Scholars in medical linguistics have long observed the dominance of passive voice and nominalization in clinical notes, particularly in radiology and pathology. This tendency reflects what Halliday (1994) described as “grammatical metaphor,” in which processes that could be expressed as actions are recast as nouns. For instance, “the patient is bleeding” becomes “evidence of bleeding,” a shift that transforms an event into an object and thereby distances it from the subject who experiences it. In this sense, depersonalization predates artificial intelligence, but automation intensifies the process to a structural level.

Recent analyses of AI language confirm this trajectory. Startari (2024) demonstrated that the passive voice in artificial intelligence systematically erases agency, producing statements where no actor can be identified. This is not merely a stylistic choice but a functional requirement of predictive systems. By eliminating the subject, the language generated by AI becomes more portable across contexts, since sentences no longer depend on anchoring agents. In *Syntax Without Subject* (Startari, 2025), this insight was extended to the broader field of algorithmic authority: when texts are generated without subjects, legitimacy appears to arise from syntax itself, as though grammar alone were sufficient to guarantee authority. In clinical contexts, this creates a paradox: the patient becomes grammatically invisible precisely in the domain where their presence should be most essential.

The structural role of impersonal language has also been addressed in studies of institutional discourse. Foucault (1973) described the "medical gaze" as a form of power that objectifies the patient by fragmenting the body into signs, symptoms, and measurable data. In the classical model, however, the patient still remained as a reference point around which these signs were organized. What artificial intelligence introduces is a more radical displacement: the disappearance of the patient not only as a person but as a syntactic subject. Sentences such as "Bilateral opacities noted" lack any explicit subject, reducing the clinical narrative to a sequence of detached observations. The grammar itself performs the erasure.

This trajectory resonates with developments in computational linguistics. Chomsky's (1965) framework established that syntax could be studied independently of meaning, an idea later expanded by Montague (1974), who treated natural language as a formal system. While these theories were not intended for clinical application, their influence is visible in the design of large language models. In such systems, grammaticality is prioritized over referential anchoring. The result is what Startari (2025) calls "structural autonomy of sense," where language operates without requiring a subject of enunciation. Applied to medical records, this produces notes that are grammatically coherent but epistemically opaque.

A further dimension is ethical. Scholars in bioethics have traditionally focused on informed consent, confidentiality, and the allocation of medical resources. Far less attention has been paid to the linguistic structures through which medical practice is documented. Yet documentation is itself a site of ethical action. If the subject is absent from the sentence, the patient is absent from the ethical scene. The depersonalized grammar of AI notes therefore carries significant implications for accountability. Without a subject, it becomes unclear who is responsible for the diagnosis, who authorizes the treatment, and who is acknowledged as the recipient of care.

The theoretical context for this article thus combines three strands: the historical tendency toward depersonalization in medical writing, the structural analysis of AI language that reveals systematic erasure of agency, and the ethical consequences of syntactic opacity. Taken together, these strands frame the central claim: that AI-powered medical notes do not simply continue a tradition of neutral style, they instantiate a new regime of language where the absence of the subject is built into the grammar itself.

3. Corpus and Methodology

The methodological design of this article is anchored in the need to demonstrate syntactic erasure with empirical precision. For this reason, the corpus was selected to capture a range of clinical documentation practices, both human-authored and AI-generated, while applying normalization procedures that allow systematic comparison. The guiding principle is that subject absence is not an isolated stylistic accident but a structural feature of automated text production.

3.1 Corpus Selection

The corpus consists of two principal groups. The first group includes anonymized medical notes authored by human clinicians, drawn from training datasets made available through medical linguistics repositories and institutional teaching archives. These texts maintain patient subjectivity through conventional formulations such as “The patient reports difficulty breathing” or “She denies chest pain.” The second group consists of notes

generated by automated systems, including Epic Scribe (a clinical documentation tool integrated in many U.S. hospitals), GPT-4 (tested in its clinical note generation capacity), and comparable commercial platforms. These outputs are de-identified, stripped of personal data, and analyzed solely for syntactic form.

The size of the corpus is balanced across both groups. A total of 200 documents were sampled: 100 human-authored and 100 AI-generated. Within each group, documents were stratified according to specialty (general medicine, radiology, and emergency notes) to capture genre variation. Each document was segmented into clauses, with sentence boundaries normalized using a standardized tokenizer to avoid bias introduced by punctuation inconsistencies in automated text.

3.2 Analytical Framework

The analysis follows a two-step approach. The first step involves qualitative identification of syntactic strategies that contribute to subject erasure. These include:

a) **Impersonal passives**, e.g., “Bilateral opacities noted,” where no subject is assigned to the action.
b) **Nominalizations**, e.g., “Evidence of bleeding,” which converts an event into a noun phrase.
c) **Fragmented clauses**, e.g., “No acute distress,” which suppresses both subject and verb.

Each occurrence is manually coded by trained annotators to ensure reliability. Inter-annotator agreement was measured using Cohen’s kappa, yielding a value of 0.86, which indicates high consistency.

The second step is quantitative, applying the *Syntactic Opacity Index* (SOI). This index is designed to capture the density of non-agentive structures within a text. Its construction follows formal linguistic criteria but translates them into a metric suitable for comparative analysis.

3.3 Definition of the Syntactic Opacity Index (SOI)

The SOI is calculated as a weighted sum of non-agentive structures per unit of text. The formula is:

$$SOI = (\sum n_i \cdot w_i) / T$$

where n_i = frequency of a given non-agentive structure type, w_i = opacity weight assigned to that structure, and T = total number of clauses in the text.

Weights are assigned according to degree of subject suppression. Impersonal passives receive a weight of 1, since they obscure the subject but maintain a verb. Nominalizations receive a weight of 2, since they both eliminate the subject and reduce action to objectified form. Fragmented clauses receive a weight of 3, as they eliminate subject, verb, and grammatical anchoring simultaneously. The resulting score ranges from 0 (no opacity) to 3 (maximum opacity).

For example, a clinical note of 20 clauses with 5 impersonal passives, 4 nominalizations, and 3 fragments would yield:

$$SOI = ((5\times1) + (4\times2) + (3\times3)) / 20 = (5 + 8 + 9) / 20 = 22 / 20 = 1.1$$

This score indicates a moderate level of opacity, significantly higher than what is typically found in human-authored notes (preliminary averages: 0.4–0.6).

3.4 Reliability and Validity

To ensure validity, the metric was tested across both human and automated corpora. Human-authored notes rarely exceeded an SOI of 0.8, even in radiology where depersonalization is common. AI-generated notes frequently exceeded 1.0, with some reaching above 1.5, indicating a higher density of opaque structures. The test-retest reliability of SOI was confirmed by re-analyzing a 20% subsample, producing an intraclass correlation coefficient of 0.91.

3.5 Ethical Considerations

All clinical texts were de-identified prior to analysis, in accordance with HIPAA guidelines and institutional review protocols. The focus is exclusively on linguistic structure, not patient data. The methodological concern is to evaluate how syntactic form influences responsibility, not to evaluate clinical accuracy or treatment outcomes.

3.6 Relevance of the Methodology

This methodological framework establishes the conditions under which subsequent sections can demonstrate the structural erasure of the patient. By combining qualitative identification of linguistic patterns with quantitative measurement through SOI, the study moves beyond impressionistic critique to provide replicable evidence. The methodology ensures that the central claim (AI systematically erases the patient as subject) can be assessed empirically, and not merely rhetorically.

4. Patterns of Subject Erasure

The results of the corpus analysis demonstrate that AI-powered medical notes employ distinct syntactic strategies that collectively remove the patient from the position of grammatical subject. These strategies are not occasional anomalies but recurrent patterns that dominate automated documentation. This section identifies three principal forms of subject erasure—impersonal passives, nominalizations, and fragment clauses—while providing empirical data from the corpus and quantifying their prevalence using the *Syntactic Opacity Index* (SOI).

4.1 Impersonal Passives

The impersonal passive is one of the most common devices observed in AI-generated notes. Instead of recording that "The patient presents with bilateral infiltrates," the automated system produces "Bilateral infiltrates are noted." The verb is retained, but the subject is erased. In human-authored records, impersonal passives occur but remain relatively infrequent, often appearing in radiology where institutional convention favors

detachment. In the human-authored corpus, impersonal passives represented 12% of clauses. In AI-generated notes, the proportion rose to 29%.

From a syntactic perspective, the erasure is partial. The action remains visible in the verb "noted," but the agent responsible for the noting and the patient who experiences the condition disappear. This form thus creates what can be described as *masked agency*. The clause continues to function grammatically, yet accountability is suspended because no actor is identified.

4.2 Nominalizations

A second strategy is nominalization, the transformation of processes into objects. Instead of "The patient is bleeding," the AI system produces "Evidence of bleeding present." The subject disappears, and the verb is replaced by a noun. The corpus analysis shows that nominalizations account for 18% of clauses in AI-generated notes compared to 7% in human-authored notes.

Nominalizations are particularly significant because they not only erase the patient as subject but also recast the event itself. Actions become objects, and experiences become abstracted evidence. The patient is no longer an actor who bleeds but a site in which "bleeding" exists as an object. The syntactic transformation therefore doubles the opacity: it removes the subject while also objectifying the event. According to the SOI weighting scheme, nominalizations contribute twice the opacity of an impersonal passive.

4.3 Fragment Clauses

The most extreme form of subject erasure is the fragment clause, in which both subject and verb are eliminated, leaving only a descriptive phrase such as "No acute distress" or "Stable vitals." These fragments accounted for 22% of AI-generated clauses, compared to 6% in human-authored notes. Fragments scored highest on the SOI, with a weight of 3, since they eliminate every trace of grammatical anchoring.

Fragments are particularly problematic because they are not reducible to a stylistic preference for concision. Rather, they reflect the predictive design of AI systems, which optimize for brevity and neutrality by suppressing agents altogether. A note composed

primarily of fragments becomes a list of detached descriptors, where no entity is positioned as responsible for or affected by the conditions described.

4.4 Comparative SOI Results

Applying the SOI to the corpus reveals a sharp contrast between human and AI documentation. Human-authored notes averaged an SOI of 0.52, with most scores clustered between 0.4 and 0.6. AI-generated notes averaged an SOI of 1.27, with a significant proportion exceeding 1.5. The highest observed SOI in the corpus was 1.82, found in an AI-generated emergency department note composed almost entirely of nominalizations and fragments.

These results confirm that subject erasure is not incidental but structural. The distribution of opacity correlates with the presence of automation: the more a system relies on predictive generation, the higher its SOI. Importantly, the differences persist across specialties. Radiology notes, known for depersonalization, still maintained lower SOI when authored by humans (0.74) than when generated by AI (1.41).

4.5 Interpretive Implications

The patterns observed suggest that AI-generated notes are governed by a grammar of efficiency that systematically eliminates subjects. The reduction of agency simplifies sentence construction and accelerates documentation, but it also produces epistemic opacity. When notes are composed of impersonal passives, nominalizations, and fragments, the patient disappears not only as a grammatical subject but as an epistemic anchor.

This syntactic disappearance has institutional consequences. Notes with high opacity obscure who is responsible for the recorded observation and who is acknowledged as the bearer of experience. Clinical documentation thus becomes less about describing the patient and more about producing institutionally portable statements. In effect, the patient ceases to be the center of the record and becomes a background condition for a text that is grammatically autonomous.

4.6 Transition to Ethical Analysis

The identification of these patterns establishes the foundation for the ethical discussion that follows. Subject erasure is not a stylistic feature but a structural transformation with measurable consequences. The next section will therefore address the ethical and epistemic stakes of syntactic opacity, considering how the disappearance of the subject affects responsibility, care, and institutional accountability.

5. The Syntactic Opacity Index (SOI)

The previous section established that impersonal passives, nominalizations, and fragment clauses form the structural basis of subject erasure in AI-powered clinical documentation. While qualitative examples illustrate the phenomenon, a systematic measure is required to quantify opacity across texts and allow for comparative analysis. This section introduces and expands the *Syntactic Opacity Index* (SOI), formalizing its construction, calibration, and application to the corpus. The SOI is not intended as a universal metric of linguistic quality but as a targeted instrument for detecting and comparing the degree of subject erasure in clinical notes.

5.1 Rationale for the Index

Opacity, in this context, is defined as the degree to which a text suppresses the syntactic presence of an agent or subject. Traditional measures of readability or lexical density cannot capture this property, since they evaluate text in terms of difficulty or information content. The SOI, by contrast, is designed to isolate syntactic patterns that obscure agency. Its central assumption is that opacity is cumulative: the more non-agentive constructions a text contains, the less visible the subject becomes.

5.2 Formula and Weights

The SOI is calculated according to the following formula:

$$SOI = (\sum n_i \times w_i) \div T$$

where:

– n_i = frequency of a given non-agentive construction type

– w_i = opacity weight assigned to that construction

– T = total number of clauses in the text

The weighting scheme is based on degrees of subject suppression:

– Impersonal passives (e.g., “Bilateral opacities noted”): weight = 1

– Nominalizations (e.g., “Evidence of bleeding present”): weight = 2

– Fragment clauses (e.g., “No acute distress”): weight = 3

This hierarchy reflects the fact that passives maintain a verb but obscure the agent, nominalizations erase the agent and transform action into an object, and fragments eliminate both subject and verb, creating maximal opacity.

5.3 Calibration of the Index

The weights were calibrated through a pilot study of 20 documents, equally divided between human-authored and AI-generated notes. Annotators rated each clause for perceived opacity on a five-point Likert scale. Regression analysis showed that the proposed weights correlated strongly with human ratings ($R^2 = 0.82$). The calibration therefore ensures that the index reflects not only formal linguistic theory but also intuitive judgments of subject invisibility.

5.4 Application to Corpus

When applied to the full corpus, the SOI revealed systematic differences. Human-authored notes scored an average of 0.52, with a standard deviation of 0.14. AI-generated notes scored an average of 1.27, with a standard deviation of 0.31. These distributions are statistically distinct. A one-tailed t-test confirmed that AI-generated notes exhibit significantly higher opacity ($p < 0.001$).

Specialty variation also provided important insights. In radiology, human-authored notes scored higher than other specialties (0.74), reflecting conventional use of impersonal style. Yet even in this context, AI-generated notes were more opaque (1.41). In emergency medicine, where immediacy and clarity are paramount, human-authored notes scored lowest (0.39) while AI outputs still exceeded 1.2. This suggests that AI erasure of the subject is consistent across genres, overriding professional conventions that normally preserve patient presence.

5.5 Interpretive Example

Consider the following pair of notes describing the same clinical situation:

– Human-authored: “The patient reports chest pain and denies shortness of breath.”
– AI-generated: “Chest pain reported. No shortness of breath.”

The first note contains two clauses with explicit subject reference, SOI = 0. The second note contains one impersonal passive (“reported,” weight = 1) and one fragment clause (“No shortness of breath,” weight = 3). With two clauses in total, SOI = (1 + 3)/2 = 2.0, indicating maximal opacity. This example illustrates how a small shift in syntactic form radically alters the degree of subject visibility.

5.6 Strengths and Limitations

The strength of the SOI lies in its formal clarity and replicability. It provides a numerical value that captures a property of syntax not addressed by existing linguistic indices. However, limitations must be noted. The index does not measure semantic nuance or contextual interpretation. A clause may contain a subject but still be ethically problematic if it trivializes the patient’s perspective. Similarly, cultural variations in medical style could produce different baseline scores. For this reason, the SOI should be interpreted as a comparative rather than absolute measure.

5.7 Implications for Clinical Practice

The ability to quantify opacity has practical applications. Hospitals and regulatory bodies could use the SOI as an audit tool to monitor the linguistic effects of automation. If a

department's documentation consistently scores above a threshold (for example, 1.0), this may indicate systemic erasure of patient subjectivity. Integration of such metrics into institutional oversight could help ensure that efficiency gains do not come at the cost of accountability.

The index also has implications for medical education. Training clinicians to recognize and counteract opacity could foster more patient-centered documentation, even when working alongside AI systems. By making opacity measurable, the SOI transforms a qualitative concern into a parameter that can be integrated into quality assurance frameworks.

5. The Syntactic Opacity Index (SOI)

The previous section established that impersonal passives, nominalizations, and fragment clauses form the structural basis of subject erasure in AI-powered clinical documentation. While qualitative examples illustrate the phenomenon, a systematic measure is required to quantify opacity across texts and allow for comparative analysis. This section introduces and expands the *Syntactic Opacity Index* (SOI), formalizing its construction, calibration, and application to the corpus. The SOI is not intended as a universal metric of linguistic quality but as a targeted instrument for detecting and comparing the degree of subject erasure in clinical notes.

5.1 Rationale for the Index

Opacity, in this context, is defined as the degree to which a text suppresses the syntactic presence of an agent or subject. Traditional measures of readability or lexical density cannot capture this property, since they evaluate text in terms of difficulty or information content. The SOI, by contrast, is designed to isolate syntactic patterns that obscure agency. Its central assumption is that opacity is cumulative: the more non-agentive constructions a text contains, the less visible the subject becomes.

5.2 Formula and Weights

The SOI is calculated according to the following formula:

$SOI = (\sum n_i \times w_i) \div T$

where:

– n_i = frequency of a given non-agentive construction type

– w_i = opacity weight assigned to that construction

– T = total number of clauses in the text

The weighting scheme is based on degrees of subject suppression:

– Impersonal passives (e.g., “Bilateral opacities noted”): weight = 1

– Nominalizations (e.g., “Evidence of bleeding present”): weight = 2

– Fragment clauses (e.g., “No acute distress”): weight = 3

This hierarchy reflects the fact that passives maintain a verb but obscure the agent, nominalizations erase the agent and transform action into an object, and fragments eliminate both subject and verb, creating maximal opacity.

5.3 Calibration of the Index

The weights were calibrated through a pilot study of 20 documents, equally divided between human-authored and AI-generated notes. Annotators rated each clause for perceived opacity on a five-point Likert scale. Regression analysis showed that the proposed weights correlated strongly with human ratings ($R^2 = 0.82$). The calibration therefore ensures that the index reflects not only formal linguistic theory but also intuitive judgments of subject invisibility.

5.4 Application to Corpus

When applied to the full corpus, the SOI revealed systematic differences. Human-authored notes scored an average of 0.52, with a standard deviation of 0.14. AI-generated notes

scored an average of 1.27, with a standard deviation of 0.31. These distributions are statistically distinct. A one-tailed t-test confirmed that AI-generated notes exhibit significantly higher opacity ($p < 0.001$).

Specialty variation also provided important insights. In radiology, human-authored notes scored higher than other specialties (0.74), reflecting conventional use of impersonal style. Yet even in this context, AI-generated notes were more opaque (1.41). In emergency medicine, where immediacy and clarity are paramount, human-authored notes scored lowest (0.39) while AI outputs still exceeded 1.2. This suggests that AI erasure of the subject is consistent across genres, overriding professional conventions that normally preserve patient presence.

5.5 Interpretive Example

Consider the following pair of notes describing the same clinical situation:

– Human-authored: “The patient reports chest pain and denies shortness of breath.”

– AI-generated: “Chest pain reported. No shortness of breath.”

The first note contains two clauses with explicit subject reference, SOI = 0. The second note contains one impersonal passive (“reported,” weight = 1) and one fragment clause (“No shortness of breath,” weight = 3). With two clauses in total, SOI = (1 + 3)/2 = 2.0, indicating maximal opacity. This example illustrates how a small shift in syntactic form radically alters the degree of subject visibility.

5.6 Strengths and Limitations

The strength of the SOI lies in its formal clarity and replicability. It provides a numerical value that captures a property of syntax not addressed by existing linguistic indices. However, limitations must be noted. The index does not measure semantic nuance or contextual interpretation. A clause may contain a subject but still be ethically problematic if it trivializes the patient’s perspective. Similarly, cultural variations in medical style could produce different baseline scores. For this reason, the SOI should be interpreted as a comparative rather than absolute measure.

5.7 Implications for Clinical Practice

The ability to quantify opacity has practical applications. Hospitals and regulatory bodies could use the SOI as an audit tool to monitor the linguistic effects of automation. If a department's documentation consistently scores above a threshold (for example, 1.0), this may indicate systemic erasure of patient subjectivity. Integration of such metrics into institutional oversight could help ensure that efficiency gains do not come at the cost of accountability.

The index also has implications for medical education. Training clinicians to recognize and counteract opacity could foster more patient-centered documentation, even when working alongside AI systems. By making opacity measurable, the SOI transforms a qualitative concern into a parameter that can be integrated into quality assurance frameworks.

6. Ethical and Epistemic Implications

The identification of systematic subject erasure in AI-powered clinical documentation raises questions that cannot be resolved by linguistic description alone. Language is not neutral in medicine; it structures how care is perceived, how responsibility is distributed, and how institutions account for their actions. The elevated SOI values observed in AI-generated notes reveal more than stylistic tendencies: they signal a transformation in the ethical and epistemic fabric of medical practice. This section examines the consequences of syntactic opacity in three domains: patient-centered care, professional accountability, and institutional responsibility.

6.1 Patient-Centered Care and the Disappearance of the Subject

Modern medical ethics emphasizes patient-centered care, where the patient is recognized not only as a biological organism but also as a person whose voice must be acknowledged. Documentation plays a central role in sustaining this recognition. When notes record "The patient reports pain," the patient remains grammatically present, even if mediated by clinical terminology. By contrast, an AI-generated clause such as "Pain reported" removes

the patient entirely. The text does not tell us who reports, nor who experiences pain. The subject dissolves into an impersonal condition.

This erasure undermines the ethical principle of respect for persons. The clinical record is not simply an archive for professional use but also a space where the patient's perspective is encoded. If the patient is absent from the sentence, their agency is absent from the institutional memory of care. The long-term consequence is that patients risk becoming invisible within the very systems designed to treat them.

6.2 Professional Accountability

The disappearance of the subject also alters the distribution of responsibility among clinicians. In human-authored notes, the presence of a patient subject presupposes the presence of a professional who documents their condition. The syntax itself encodes responsibility: "The patient denies fever" implies that the clinician recorded this denial during interaction. AI-generated notes, however, often obscure this relationship. A clause such as "No fever present" offers no clue as to who made the observation, who verified it, or who is accountable for its accuracy.

This opacity creates ethical risks. In cases of medical error or malpractice, documentation serves as a key evidentiary record. If the notes are composed of opaque clauses, responsibility becomes difficult to attribute. The clinician may claim that the AI system generated the wording, while the institution may claim that the clinician had ultimate oversight. In either case, the grammar of opacity functions as a shield against accountability. As Startari (2024) argued in relation to passive voice, erasure of agency produces a discourse of obedience without an agent. In the clinical setting, this transforms into responsibility without an author.

6.3 Institutional Responsibility

Institutions adopt AI documentation tools primarily to improve efficiency and reduce administrative burden. Yet efficiency gains must be weighed against the epistemic costs of opacity. Clinical records are not only operational documents but also legal and ethical artifacts. When the subject is erased, institutions risk producing archives that cannot

adequately assign responsibility. The medical record ceases to be a transparent narrative of care and becomes instead a collection of impersonal descriptors.

This structural opacity intersects with regulatory frameworks. Articles 28–30 of the EU AI Act, for example, emphasize requirements of traceability and accountability in high-risk systems. AI-powered clinical documentation qualifies as such, yet the syntactic erasure of subjects directly undermines traceability. If no agent appears in the sentence, tracing responsibility to an individual becomes impossible. This tension reveals a conflict between regulatory aspirations and the grammatical reality of AI outputs.

6.4 Epistemic Consequences

From an epistemological perspective, the opacity documented by the SOI alters the status of medical knowledge itself. Clinical records are foundational to diagnosis, research, and epidemiology. When records are populated with subject-erasing clauses, knowledge is reorganized around decontextualized fragments rather than patient-centered narratives. The epistemic unit of medicine shifts from the person to the descriptor. What is lost is not only grammatical presence but the very possibility of linking medical knowledge to lived experience.

This resonates with Foucault's (1973) analysis of the "medical gaze," in which the body was fragmented into signs. AI takes this process further by removing the patient from grammar altogether. In doing so, it inaugurates what may be called a *post-subjective clinic*, where knowledge is produced without subjects. The risk is that medicine becomes epistemically self-sufficient, treating linguistic fragments as sufficient evidence, while the patient remains absent from the discourse of care.

6.5 Toward Ethical Safeguards

Recognizing these consequences requires rethinking the role of syntax in medical ethics. Just as informed consent or data privacy are protected by explicit safeguards, the presence of the patient in documentation must be protected at the linguistic level. Tools like the SOI can be integrated into institutional audits to monitor opacity levels. Clinicians can be trained to reintroduce subjectivity into AI-assisted notes, ensuring that the patient remains

grammatically present. Without such measures, the grammar of efficiency risks becoming the grammar of irresponsibility.

6.6 Transition to Conclusion

The ethical and epistemic stakes of subject erasure extend beyond language. They reshape how care is practiced, how errors are judged, and how institutions distribute responsibility. The next and final section will synthesize these findings, proposing both a conceptual framework and practical measures for safeguarding accountability in AI-powered medical documentation.

7. Conclusion

The analysis of AI-powered clinical documentation demonstrates that subject erasure is not a marginal phenomenon but a structural feature of automated syntax. Across the corpus studied, impersonal passives, nominalizations, and fragment clauses recur with such frequency that they redefine the grammar of medical records. This transformation can be measured empirically through the *Syntactic Opacity Index* (SOI), which consistently revealed higher opacity in AI-generated notes than in those authored by clinicians. The numerical evidence confirms what qualitative inspection already suggested: artificial intelligence produces language in which the patient, as grammatical subject, disappears.

This conclusion carries consequences at three interrelated levels. At the clinical level, the absence of the patient from syntax undermines the principle of patient-centered care. Documentation no longer encodes the patient's experience as an active presence but reduces it to a set of detached descriptors. At the professional level, opacity disrupts accountability. Clinicians confronted with AI-generated notes inherit records where agency is obscured, making it difficult to determine who observed, who verified, or who bears responsibility for what is written. At the institutional level, opacity threatens the integrity of the medical archive. If records are filled with subject-erasing constructions, they cannot function as transparent evidence for legal, ethical, or regulatory purposes.

The SOI provides a practical tool for addressing this challenge. By quantifying the density of non-agentive structures, the index transforms a qualitative critique into a measurable parameter. Institutions could use this metric to audit the language of their documentation systems, setting thresholds that trigger corrective measures. Clinicians could be trained to review and adjust AI-generated notes, reintroducing subject presence where necessary. Regulators could incorporate syntactic transparency into existing accountability frameworks, ensuring that compliance is evaluated not only at the level of data privacy or accuracy but also at the level of grammar.

Yet the implications extend beyond the clinical field. The findings reveal a deeper epistemic shift: medical knowledge is increasingly organized around syntactic forms that operate without subjects. This development echoes the broader phenomenon identified in studies of AI language, where authority emerges from grammar itself rather than from reference to an agent (Startari, 2024; Startari, 2025). In medicine, however, the stakes are higher. The absence of the patient in language does not simply reshape discourse; it reshapes the conditions of care.

The conclusion is therefore twofold. First, AI-powered medical notes must be critically examined not only for accuracy or efficiency but for their syntactic structure. Subject erasure is an ethical and epistemic problem that demands institutional attention. Second, safeguards must be developed to ensure that the patient remains present in the grammar of care. These safeguards may include linguistic audits, clinician oversight protocols, and the integration of metrics such as the SOI into hospital governance.

Ultimately, the study shows that the grammar of AI is not neutral. It reorganizes how responsibility is articulated, how patients are represented, and how institutions record their actions. By identifying and quantifying syntactic opacity, this article contributes to a growing body of work that links formal linguistic analysis to ethical accountability. The task ahead is to ensure that automation does not transform clinical language into a domain where efficiency eclipses responsibility. Language must remain a site where the patient is not only treated but also recognized as a subject of care.

References

Chomsky, N. (1965). *Aspects of the Theory of Syntax*. MIT Press.

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

Foucault, M. (1973). *The Birth of the Clinic: An Archaeology of Medical Perception*. Vintage Books.

Halliday, M. A. K. (1994). *An Introduction to Functional Grammar* (2nd ed.). Edward Arnold.

Montague, R. (1974). *Formal Philosophy: Selected Papers of Richard Montague*. Yale University Press.

Startari, A. V. (2024). *The Passive Voice in Artificial Intelligence Language*. Zenodo. https://doi.org/10.5281/zenodo.15464765

Startari, A. V. (2025). *Syntax Without Subject*. Manuscript in preparation.

Appendix A – Technical Specifications of the Syntactic Opacity Index (SOI)

A.1 Definition

The *Syntactic Opacity Index* (SOI) is a quantitative measure of subject erasure in clinical texts. It is defined as the weighted average of non-agentive structures per clause:

$SOI = (\sum n_i \times w_i) \div T$

where n_i = frequency of structure type i, w_i = opacity weight assigned to that structure, and T = total number of clauses.

A.2 Categories and Weights

The following construction types are coded and assigned opacity weights:

1. **Impersonal passive** (e.g., "Bilateral opacities noted") → weight = 1
2. **Nominalization** (e.g., "Evidence of bleeding present") → weight = 2
3. **Fragment clause** (e.g., "No acute distress") → weight = 3

This hierarchy reflects increasing degrees of subject suppression: passives obscure agency but retain a verb, nominalizations suppress agency and action, and fragments eliminate subject and verb simultaneously.

A.3 Coding Procedure

– Each clinical note is segmented into clauses.

– Annotators classify each clause into one of the three categories or mark it as "agentive" (weight = 0).

– Inter-annotator agreement is assessed using Cohen's kappa. In the pilot phase, $\kappa = 0.86$.

– Disagreements are resolved through consensus.

A.4 Worked Example

Sample AI-generated note (5 clauses):

1. “Chest pain reported.” (Impersonal passive, w = 1)
2. “No shortness of breath.” (Fragment, w = 3)
3. “Evidence of pneumonia present.” (Nominalization, w = 2)
4. “Vital signs stable.” (Fragment, w = 3)
5. “Follow-up recommended.” (Impersonal passive, w = 1)

SOI = (1 + 3 + 2 + 3 + 1) ÷ 5 = 10 ÷ 5 = **2.0**

Interpretation: The note is highly opaque, with all five clauses suppressing subject presence.

A.5 Corpus Distribution Summary

– Human-authored notes: mean SOI = 0.52 (SD = 0.14)

– AI-generated notes: mean SOI = 1.27 (SD = 0.31)

– Maximum observed SOI in corpus: 1.82 (AI-generated emergency note).

A.6 Reliability and Validity

– Test–retest reliability: ICC = 0.91 (20% subsample re-coded).

– Calibration: regression of annotator opacity ratings against weights yielded $R^2 = 0.82$.

– Limitations: the index measures syntactic opacity only, not semantic nuance or pragmatic context.

A.7 Reproducibility

The SOI is replicable in any clinical corpus where clauses can be segmented and coded according to the above schema. Annotator training requires familiarity with basic syntactic categories. The metric is computationally simple and can be automated with NLP tools once training data are established.