AI Power and Discourse, vol. 2, núm. 1, 2026, pp. 1-10.

Plagiarism Ex Machina: Structural Appropriation in Large Language Models.

Agustin V. Startari.

Cita:

Agustin V. Startari (2026). *Plagiarism Ex Machina: Structural Appropriation in Large Language Models. AI Power and Discourse, 2 (1), 1-10.*

Dirección estable: https://www.aacademica.org/agustin.v.startari/230

ARK: https://n2t.net/ark:/13683/p0c2/0su

Plagiarism Ex Machina: Structural Appropriation in Large Language Models

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: May 7, 2026

DOI

- Primary archive: https://doi.org/**10.5281/zenodo.20070859**
- Secondary archive: https://doi.org/10.6084/m9.figshare.32204832
- SSRN: Pending assignment (ETA: Q3 2026)

Language: English

Series: *AI Syntactic Power and Legitimacy*

Word count: 22375

Keywords: Large Language Models; Plagiarism; Structural Appropriation; Recombinative Plagiarism; Synthetic Originality; Predictive Authorship; Referential Opacity; Attribution Collapse; Latent Intellectual Debt; Corpus Extraction; Corpus Parasitism; Invisible Intellectual Labor; Generative Provenance; AI Authorship; Source Transparency; Epistemic Extraction; Knowledge Commons; Algorithmic Attribution; Predictive Text Generation; Academic Integrity; AI Governance; Intellectual Property; Copyright; Citation Ethics; Authorship Theory; Machine-Generated Text; Textual Recombination; Provenance Auditing; Synthetic Knowledge; Structural Dependency; Authority Without Source; Indexical Collapse; Ethos Without Source; AI-Assisted Writing; Academic Credit; Knowledge Governance; Generative AI; Training Data; Dataset Opacity; Post-Hoc Citation; Source Lineage.

Abstract

Large language models have destabilized the classical definition of plagiarism. Traditional plagiarism frameworks presuppose identifiable duplication, traceable authorship, and recoverable chains of intellectual attribution between an original source and a reproducing subject. Generative systems disrupt this architecture. Rather than copying isolated passages, they absorb massive corpora of human production, compress conceptual and stylistic patterns into statistical representations, and generate outputs whose epistemic lineage becomes structurally opaque. This paper defines this phenomenon as *structural appropriation*: the extraction and recombination of intellectual labor without transparent mechanisms of provenance reconstruction. The article argues that contemporary debates centered on copyright infringement, memorization, and fair use fail to capture the deeper transformation introduced by predictive language systems. The central problem is not limited to literal duplication. It is the emergence of *recombinative plagiarism*, a regime in which originality is synthetically produced through probabilistic recombination of absorbed human knowledge. Under these conditions, attribution collapses not because sources disappear from training data, but because the generated output dissolves reconstructible links between intellectual origin and textual manifestation. Through a linguistic, epistemological, and infrastructural analysis, the paper introduces several operational concepts: *referential opacity*, *synthetic originality*, *latent intellectual debt*, and *corpus parasitism*. It demonstrates how generative architectures convert distributed human authorship into predictive synthesis while preserving the economic and symbolic value extracted from collective intellectual production. The model therefore functions not merely as a textual generator, but as an infrastructural intermediary that redistributes legitimacy while obscuring the underlying debt relations embedded in its outputs. The paper further examines the consequences of structural appropriation for academia, journalism, law, and knowledge governance. It argues that existing attribution systems are structurally incapable of auditing probabilistic generation because they were designed for document-level similarity rather than latent recombination across high-dimensional semantic spaces. As a corrective framework, the article proposes the foundations of a *generative provenance model* capable of integrating probabilistic source visibility, attribution mapping, and transparency layers into future language systems.

Ultimately, the paper contends that the defining characteristic of plagiarism in predictive systems is no longer duplication, but the disappearance of recoverable epistemic provenance under conditions of synthetic generation. In large language models, originality increasingly emerges as a statistical effect of recombination depth rather than as evidence of autonomous intellectual creation.

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]

Part I. The End of Classical Plagiarism

Classical plagiarism presupposes a stable relation between source, copy, and author. It assumes that an identifiable subject reproduces, paraphrases, or appropriates a prior textual object without sufficient acknowledgment. This model depends on three structural conditions: the existence of a recoverable original, the possibility of comparing that original with a later text, and the attribution of responsibility to a human agent who either concealed or failed to disclose the borrowed material. Under this framework, plagiarism appears as a violation of textual integrity. The wrong consists in presenting another author's words, ideas, or distinctive formulations as one's own. The detection of that wrong therefore depends on comparison: one text is placed against another, overlap is measured, and responsibility is assigned through the relation between the borrowed material and the author who failed to cite it.

Large language models destabilize this architecture. They do not operate primarily by selecting one source and reproducing it as a discrete act of copying. They absorb enormous corpora of human writing, transform those materials into statistical parameters, and produce outputs through probabilistic generation. The result may contain no verbatim duplication, no directly identifiable passage, and no single recoverable source. Yet the output still depends on prior intellectual labor. The absence of literal copying does not eliminate appropriation. It changes its form. This distinction is central because conventional plagiarism frameworks remain oriented toward visible textual similarity, while generative systems operate through latent statistical dependency.

The central problem is therefore conceptual. Existing definitions of plagiarism were built for document-level comparison, not for model-level absorption. They detect overlap between texts, not structural dependency within predictive systems. Similarity software can identify copied sentences, paraphrased passages, or suspicious lexical proximity. It cannot reconstruct the distributed intellectual debt embedded in a model's parameters. Once source materials are compressed into a high-dimensional generative architecture, attribution no longer operates as a visible relation between texts. It becomes a buried relation between corpus, model, and output. This is the point at which plagiarism ceases to be only a textual violation and becomes an infrastructural problem.

This paper defines that condition as structural appropriation. Structural appropriation occurs when a generative system extracts patterns from prior human production and reactivates them in new outputs without preserving transparent provenance. The appropriated object is not only a sentence or paragraph. It may be an argumentative structure, a conceptual sequence, a stylistic rhythm, a taxonomy, a rhetorical frame, or a recognizable academic posture. The generated text can appear original because it does not duplicate any one source exactly. Its originality, however, is produced through recombination of absorbed intellectual materials whose origins cannot be reconstructed by the reader, the user, or often even the system provider.

This shift requires separating plagiarism from duplication. Duplication is one possible evidence of plagiarism, but it is not the entire phenomenon. A system may avoid verbatim reproduction while still converting uncredited human knowledge into commercial, academic, or symbolic value. In classical settings, the plagiarist hides the source. In generative settings, the system dissolves the source. The difference is decisive. Hidden sources may be uncovered through comparison. Dissolved sources are structurally unrecoverable because the output is not a copy of one document but a synthesis produced from statistical relations across many documents.

The insufficiency of duplication-based analysis becomes clearer when placed against the broader literature on large-scale language modeling. Bender, Gebru, McMillan-Major, and Shmitchell (2021) argue that language models reproduce statistical regularities from large datasets while obscuring the social and documentary conditions embedded in those datasets. Their critique does not reduce the problem to isolated copying. It identifies a deeper relation between dataset scale, opacity, and the reproduction of human language as machine output. In the context of plagiarism, this means that the relevant object of analysis cannot be limited to the generated sentence. It must include the training infrastructure that makes the sentence possible.

The same problem appears in discussions of automation bias and human reliance on machine-generated outputs. Parasuraman and Riley (1997) show that automated systems can be misused or over-trusted when users treat outputs as reliable without sufficiently inspecting the underlying process. In predictive writing, this produces a specific authorship

problem. The user receives fluent language without knowing the dependency structure that produced it. The model's surface coherence becomes a substitute for provenance. Under these conditions, plagiarism becomes difficult to identify because the user may not possess either the knowledge or the technical access required to reconstruct source relations.

This is why the phrase "AI plagiarism" is often inadequate when used only to describe students submitting generated essays or models reproducing copyrighted passages. Those are visible cases, but they remain secondary. The deeper transformation is infrastructural. Large language models convert collective authorship into a generative substrate. They ingest texts written by scholars, journalists, programmers, translators, bloggers, fiction writers, legal professionals, technical writers, and public users. Those materials are not merely stored. They are reorganized into predictive capacity. The system then returns fluent outputs that carry the value of that prior labor while suppressing its trace.

The issue is not reducible to intention. Classical plagiarism frequently depends on intentional concealment or negligent failure to cite. Structural appropriation can occur without a conscious plagiarist in the ordinary sense. The user may not know which sources informed the answer. The model does not cite unless instructed or architecturally connected to retrieval. The provider may disclose general training practices without exposing the specific lineage of any given output. Responsibility becomes distributed across dataset construction, model training, interface design, user prompting, and institutional adoption. The appropriation is therefore not merely moral or individual. It is procedural.

This procedural character distinguishes structural appropriation from ordinary textual theft. In ordinary plagiarism, an author takes from a source and removes attribution. In structural appropriation, the system is built upon prior sources in such a way that attribution becomes technically and institutionally displaced. The output arrives as a self-contained answer. It may be polished, coherent, and apparently novel. Yet its fluency depends on patterns acquired from uncounted acts of prior writing. The model's authority is parasitic on a corpus it cannot transparently acknowledge at the point of generation.

This condition extends the problem developed in *Citation by Completion*, where predictive writing systems were shown to redistribute academic credit by shaping which citations are

suggested, accepted, and normalized during composition (Startari, 2025a). In that framework, the central mechanism was not only content prediction, but the syntactic framing of legitimacy: phrases such as “as established by,” “following the seminal work of,” or “canonical framework” alter the perceived authority of a source before the writer has independently evaluated its relevance. The present paper moves one layer deeper. Before citation can be redistributed, source dependency itself has already been obscured. Plagiarism is no longer only a failure to cite an identifiable work. It becomes the systemic production of text from absorbed intellectual labor under conditions where citation cannot be reconstructed. The uploaded draft of *Citation by Completion* confirms this continuity, since it defines the problem of predictive citation systems as a redistribution of academic credit through autocomplete, authority-bearing syntax, and citation concentration.

This shift also follows from the theory of shared intellectual debt developed in *Borrowed Voices, Shared Debt*, where plagiarism and idea recombination were treated not as isolated acts of textual theft, but as structural events within the knowledge commons (Startari, 2025b). Large language models radicalize that condition by scaling recombination beyond the threshold of ordinary attribution. The model does not borrow from a single author in a recoverable way. It absorbs distributed textual production, converts it into predictive capacity, and returns outputs that may preserve conceptual, stylistic, or argumentative value while erasing visible dependency. The result is not simply unattributed borrowing. It is borrowing without a recoverable borrower-source relation.

The same logic intersects with the problem of source disappearance developed in *Ethos Without Source* and *Indexical Collapse* (Startari, 2025c, 2025d). In both cases, authority survives after the disappearance of its referential anchor. Generated text can sound legitimate, complete, and institutionally usable even when its source conditions remain unavailable. In the context of structural appropriation, this means that plagiarism no longer appears as an identifiable relation between one text and another. It appears as a referential failure built into the generative architecture itself. The system can produce an authoritative surface while withholding the lineage that would allow readers to evaluate intellectual debt.

This referential failure also connects to the theory of post-referential operative representation developed in *AI and the Structural Autonomy of Sense* (Startari, 2025e). If

generated language can operate without stable reference to an originating subject or source, then attribution becomes structurally weakened before any legal or academic evaluation begins. The output functions, persuades, explains, summarizes, or instructs, but the source relation remains displaced. In that sense, structural appropriation is not an accidental defect of large language models. It is a consequence of the same post-referential condition that allows generated text to circulate as meaningful without exposing the conditions of its formation.

The collapse of classical plagiarism is therefore a collapse of scale, traceability, and agency. Scale changes the object of appropriation: not one book, article, or paragraph, but entire textual ecosystems. Traceability changes the evidence: not visible similarity, but latent dependency. Agency changes the responsible structure: not only the individual writer, but a chain of technical and institutional actors who enable the transformation of uncredited language into generative output. These three changes make inherited plagiarism frameworks insufficient. They were designed for disputes between texts. They were not designed for systems that convert many texts into a generative capacity whose outputs may not match any one source while still depending on the extracted value of all of them.

A legal or academic system focused only on textual similarity will miss the central event. It will ask whether the generated passage matches a protected source. That question remains relevant, but it is incomplete. The stronger question is whether the model's capacity to generate authoritative language depends on unacknowledged extraction from intellectual communities whose work has been converted into predictive infrastructure. The former question concerns copying. The latter concerns appropriation. Copying can be detected through overlap. Appropriation requires analysis of dependency, opacity, and value transfer.

The distinction matters because synthetic originality can function as a laundering mechanism. A generated text may pass similarity checks precisely because the model has recombined enough sources to avoid direct overlap. The deeper the recombination, the more original the surface appears. This produces an inversion of classical plagiarism detection: the absence of textual match becomes evidence of originality, even when the output remains dependent on uncredited prior production. The system does not need to

copy in order to appropriate. It only needs to transform. In predictive systems, transformation can conceal dependency more effectively than concealment itself.

This inversion also alters the meaning of authorship. In the classical model, authorship is compromised when a writer claims another person's words or ideas as their own. In the generative model, authorship is compromised because the output already emerges from a distributed field of prior writings whose contribution cannot be separated from the model's statistical operation. The user may function as prompt initiator, editor, or selector, but not as sole originator of the intellectual structures mobilized by the output. The model may function as generator, but not as autonomous creator. The prior corpus remains active as the hidden substrate of production. Authorship therefore becomes layered, but the interface presents it as singular.

This produces a serious problem for academic legitimacy. Academic writing depends on the ability to distinguish contribution from inheritance. Citation systems, peer review, bibliographies, and intellectual histories all presuppose that claims can be connected to sources. Large language models weaken that presupposition by generating claims whose relation to prior sources cannot be reconstructed through ordinary scholarly methods. A bibliography can be added after generation, but this does not solve the problem if the generated argument was shaped by sources that remain unknown. Post-hoc citation may create the appearance of accountability while failing to disclose the real chain of dependency.

The same problem applies outside academia. In journalism, generated text can reproduce frames, explanatory sequences, or investigative patterns without visible debt to prior reporting. In law, generated memoranda can synthesize doctrinal language while obscuring which legal analyses shaped the output. In technical writing, generated documentation can reproduce patterns learned from open-source communities while removing the connection to those communities. Across these domains, structural appropriation operates by converting prior communicative labor into new textual products without a stable mechanism for recognizing the labor that made them possible.

Classical plagiarism theory was designed for a world in which authorship left visible traces. Large language models operate in a world where traces are statistically absorbed, recombined, and erased from the surface of the text. The result is not the end of plagiarism, but the end of its classical detectability. What emerges is a new regime of intellectual extraction in which the copied object is no longer the sentence, but the generative condition of the sentence. Structural appropriation names that regime.

This paper begins from that premise: plagiarism in large language models must be redefined at the level of infrastructure. The relevant question is not only whether a model repeats protected expression. It is whether predictive generation converts human intellectual labor into unattributed output through mechanisms that prevent provenance from being recovered. Under those conditions, plagiarism becomes less visible, more scalable, and more difficult to contest. It no longer appears as a breach inside a text. It appears as the condition under which the text becomes possible.

Part II. Corpus Extraction and Invisible Intellectual Labor

Training data is not a neutral technical input. In large language models, the corpus functions as the primary infrastructure through which human intellectual labor is converted into predictive capacity. Every generated sentence depends on prior acts of writing, classification, editing, translation, annotation, publication, discussion, and circulation. These acts are distributed across academic articles, books, journalism, code repositories, legal documents, forum posts, technical manuals, creative writing, institutional records, and user-generated content. Once aggregated into training data, they lose their original status as authored contributions and become statistical material for generation. The transformation is not merely computational. It is epistemic and economic.

The classical view of training data treats texts as examples from which a system learns linguistic regularities. This description is technically partial but analytically insufficient. It conceals the fact that the model's generative ability depends on accumulated human production. The corpus does not simply provide language. It provides conceptual distinctions, explanatory patterns, argumentative sequences, domain-specific vocabularies,

genre conventions, citation habits, stylistic markers, and institutional forms of legitimacy. The model does not create these structures ex nihilo. It extracts them from prior communicative labor and converts them into latent predictive relations. This is why corpus extraction must be understood as the first stage of structural appropriation.

The extraction process produces a double displacement. First, it displaces authorship by severing texts from the persons, institutions, and communities that produced them. Second, it displaces compensation by converting those texts into model capacity without a proportional mechanism for recognizing or rewarding the labor absorbed. A book, article, dataset, comment, or code file may contribute to the model's ability to answer, summarize, classify, imitate, explain, or generate. Yet the generated output usually does not identify which prior works shaped its structure. The labor remains active, but its attribution is erased from the visible surface of production.

This condition differs from ordinary quotation or citation. In quotation, borrowed language remains attached to a source. In citation, intellectual dependency is made explicit through a referential marker. In training, dependency is absorbed into model parameters. The source does not appear as source. It appears as fluency, competence, stylistic control, or domain familiarity. The user experiences the output as immediate generation, while the system's capacity depends on a long history of prior textual production. The corpus therefore becomes an invisible labor archive. It is not merely stored knowledge. It is transformed work.

This point extends the problem developed in *Citation by Completion*, where predictive systems were shown to redistribute academic credit by influencing which sources are suggested and how authority is syntactically framed during composition (Startari, 2025a). That paper examined redistribution at the level of citation behavior. The present argument identifies a prior layer: before credit can be redistributed among suggested citations, the underlying corpus has already been transformed into an unattributed generative substrate. The uploaded draft of *Citation by Completion* confirms this continuity, since its abstract defines predictive writing systems as mechanisms that shape academic citations, authority syntax, concentration, novelty, and legitimacy phrasing.

The difference between citation redistribution and corpus extraction is structural. Citation redistribution concerns visible references inside generated or assisted texts. Corpus extraction concerns the invisible absorption that makes generation possible before any citation appears. A model can cite a source in an output while still depending on countless uncited sources in its internal formation. This means that citation after generation cannot fully repair attribution collapse. It can only attach visible references to the final text. It cannot reconstruct the complete intellectual debt embedded in the model's capacity. The problem is not only that the model may fail to cite. It is that the model often cannot expose the real provenance of its own linguistic competence.

This is the core of invisible intellectual labor. Human authors produce the linguistic and conceptual materials from which the model learns. The model provider converts those materials into a commercial or institutional tool. The user receives generated output as if it were produced by the system in the present moment. The prior labor is hidden behind the interface. This structure creates an asymmetry between extraction and recognition. The model accumulates value from the corpus, but the corpus contributors do not remain visible as contributors. Their work becomes infrastructural, and infrastructural work is precisely the kind most easily erased.

The term corpus parasitism names this dependency. Corpus parasitism occurs when a generative system relies on large-scale human textual production while providing no transparent mechanism for proportional attribution, compensation, or provenance recovery. The term does not require the claim that every act of training is legally unlawful. It identifies a structural relation: one system derives productive capacity from another body of labor while masking the dependency relation at the point of output. The issue is not only ownership. It is the conversion of public, academic, creative, and professional language into private predictive infrastructure.

This parasitic relation becomes especially significant because large language models operate at scale. A human plagiarist may appropriate one article, one paragraph, or one idea. A model can absorb entire fields. The scale changes the moral and epistemological structure of appropriation. When millions or billions of textual units are ingested, the resulting dependency cannot be managed through ordinary citation logic. No bibliography

can list the totality of sources that shaped a generated answer. No footnote can reconstruct the statistical weight of every document involved in the production of a sentence. The scale of extraction exceeds the architecture of attribution.

This excess produces a new form of opacity. Bender, Gebru, McMillan-Major, and Shmitchell (2021) describe large language models as systems built from massive datasets whose composition and social consequences are often difficult to inspect. Their critique is central to the present argument because plagiarism in LLMs cannot be separated from dataset opacity. If the corpus cannot be fully audited, then the intellectual debt embedded in outputs cannot be fully traced. Opacity at the dataset level becomes opacity at the authorship level. The system's inability or refusal to expose training provenance directly affects whether generated text can be evaluated as original, derivative, or appropriative.

The extraction of intellectual labor also changes the status of style. Classical plagiarism usually focuses on words and ideas. Generative models absorb more than both. They also absorb stylistic habits, disciplinary tone, rhetorical pacing, genre conventions, and forms of explanation. A model trained on academic writing learns how academic authority sounds. It learns the distribution of hedges, transitions, nominalizations, passive constructions, citations, section structures, and evaluative formulas. These elements may not belong to one author in isolation, but they emerge from accumulated disciplinary labor. When reproduced without provenance, they create a style of synthetic expertise that draws from many communities while acknowledging none.

This is why structural appropriation cannot be limited to the question of exact textual reproduction. Exact reproduction is only the most visible case. The more consequential case is the extraction of form. A model can reproduce the architecture of a legal memorandum without copying a specific memorandum. It can reproduce the argumentative rhythm of academic prose without copying one article. It can reproduce the explanatory style of technical documentation without duplicating a particular manual. It can reproduce the authority posture of institutional language without citing the institutions from which that posture was learned. In each case, the appropriated object is structural rather than literal.

The same logic applies to concepts. A concept may circulate across a field and become part of its technical vocabulary. When a model absorbs that vocabulary, it also absorbs the conceptual labor that stabilized it. The model may later use the concept fluently, but without preserving the history of debate, refinement, disagreement, and authorship that produced it. This creates a flattened epistemic surface. Concepts appear available, detached from origin, and ready for recombination. The generated text then presents knowledge as if it were a free-floating resource rather than the outcome of situated intellectual work.

This flattening corresponds to the problem of post-referential operation developed in *AI and the Structural Autonomy of Sense* (Startari, 2025e). If generated language can function without maintaining stable reference to originating sources, then intellectual labor can remain operational while becoming referentially invisible. The model can use patterns extracted from human writing without preserving the source relations that would normally make scholarly accountability possible. Structural appropriation therefore emerges from the same post-referential condition: language continues to operate after its origin has been displaced.

Invisible labor also produces a political economy of textual extraction. Large language models transform human writing into productive capital. The output may be sold through subscriptions, APIs, enterprise tools, academic platforms, search interfaces, writing assistants, coding assistants, or institutional systems. In each case, previously authored material contributes to the commercial value of the system. Yet the authors whose work trained or shaped the model usually do not participate in that value chain. Their contribution is neither individually measurable nor institutionally recognized. The model converts distributed intellectual work into concentrated platform value.

This economic concentration parallels the symbolic concentration described in *Citation by Completion* (Startari, 2025a). In citation systems, predictive suggestions can concentrate academic credit around already visible authors. In corpus extraction, training systems can concentrate economic and epistemic value around platform owners. Both processes depend on the same structural move: distributed human contribution is absorbed into an automated system, then returned as output under conditions that obscure the original contributors. The

first redistributes recognition. The second redistributes productive capacity. Together, they describe a broader regime of synthetic authorship.

The problem is intensified by the fact that many contributors never consented to this transformation in any meaningful sense. Public availability of a text does not automatically imply consent to its absorption into a generative model. A text may be public for reading, citation, indexing, preservation, or scholarly debate. These uses are not identical to converting the text into training material for a system that can generate competing outputs. The distinction matters because structural appropriation concerns not only access to texts but conversion of texts into generative infrastructure. Reading a text and training on a text are not the same operation. Citation and extraction are not equivalent acts.

The academic implications are direct. Scholars write in order to contribute to fields, establish arguments, refine concepts, and enter citation networks. When their work becomes part of model training without traceable attribution, the ordinary mechanisms of academic recognition are bypassed. Their ideas may influence generated outputs, but their names may not appear. Their conceptual structures may be reused, but their authorship may remain invisible. Their phrasing may shape the model's style, but no reader can reconstruct that influence. This produces latent intellectual debt: a dependency relation that exists in the model's generative capacity but cannot be made visible through ordinary citation practice.

Latent intellectual debt is not identical to plagiarism in the classical sense. It does not always involve a directly copied passage or a single identifiable victim. It names the debt relation produced when prior intellectual labor remains active inside generation while being detached from attribution. This concept is necessary because plagiarism theory requires a category for non-local, non-verbatim, non-reconstructible appropriation. Without such a category, legal and academic systems will continue to treat generated originality as genuine whenever similarity checks fail. That would mistake surface difference for intellectual independence.

The connection with *Borrowed Voices, Shared Debt* is therefore central. That work treated plagiarism and idea recombination as problems of knowledge commons, not only as

violations between isolated individuals (Startari, 2025b). Large language models intensify the commons problem by automating recombination at scale. The model borrows voices without preserving the debt structure that borrowing creates. It generates text that appears singular while depending on a plurality of absorbed sources. The debt is shared, but the output is presented as unified. This is the structural contradiction at the center of generative authorship.

The same contradiction appears in relation to authority. Generated outputs often sound authoritative because they reproduce forms learned from authoritative corpora. Academic prose, legal analysis, policy language, medical explanation, and technical documentation all contain recognizable signals of expertise. The model learns those signals and redeploys them. This produces authority without transparent source, a problem previously developed in *Ethos Without Source* (Startari, 2025c). In the present context, the issue is not only that the generated speaker lacks a stable identity. It is that the generated authority depends on prior human ethos that has been absorbed and anonymized.

This absorbed ethos gives the model a rhetorical advantage. It can speak in the style of expertise without disclosing the experts whose writing helped produce that style. It can generate legal, academic, journalistic, or technical language without exposing the corpus-based debt behind the performance. The output becomes credible because it sounds like the institutions it has learned from. Yet the institutions, authors, and communities that created those forms of credibility remain invisible. This is not ordinary imitation. It is infrastructural style extraction.

The legal implications remain unresolved because existing frameworks often separate copyright, plagiarism, and authorship into distinct categories. Copyright asks whether protected expression has been copied or whether use falls within recognized exceptions. Plagiarism asks whether attribution has been withheld. Authorship asks who created the work. Structural appropriation cuts across all three. It may not always reproduce protected expression. It may not always permit identification of a missing citation. It may not allow a clear assignment of authorship. Yet it still converts prior work into new value. The inadequacy of existing categories does not eliminate the appropriation. It reveals the need for a broader framework.

Such a framework must begin from the corpus. If the corpus is the hidden infrastructure of generation, then transparency cannot be limited to output-level disclosure. A system that states “this text was generated by AI” discloses the immediate production tool, but not the source dependencies embedded in that tool. Output disclosure identifies the generator. It does not identify the extracted labor. Similarly, a list of general data categories does not provide provenance for a specific output. It may describe the system’s training environment, but it does not reveal which intellectual structures shaped the generated text. Transparency must therefore be designed as provenance, not merely as labeling.

A provenance-centered model would require at least three levels of disclosure. First, dataset-level disclosure: what kinds of texts were used, under what licenses, from which domains, and with what exclusions. Second, model-level disclosure: how training transforms textual material into generative capacity, including whether memorization, style transfer, or domain-specific concentration risks were audited. Third, output-level disclosure: whether a generated response can provide probabilistic source mappings, citation candidates, or confidence indicators about likely dependency clusters. Without these levels, users cannot distinguish between generated explanation and structurally appropriated knowledge.

The technical difficulty of full provenance does not invalidate the ethical requirement. Many forms of accountability begin as imperfect approximations. Citation itself is not a complete map of influence. It is a normative mechanism for making intellectual debt visible enough for evaluation. Generative provenance would serve the same function under predictive conditions. It would not need to reconstruct every parameter relation. It would need to prevent the total disappearance of source dependency. The current problem is not that provenance is incomplete. It is that the dominant interface presents generated text as if provenance were irrelevant.

This is why corpus extraction must be placed at the center of any serious theory of AI plagiarism. If analysis begins only with the final output, it will remain trapped in similarity detection. It will ask whether the answer matches something already written. But structural appropriation begins before the output. It begins when texts are ingested, stripped of their

original contexts, transformed into training material, and converted into predictive capacity. The generated sentence is only the final surface of a deeper extraction process.

The category of invisible intellectual labor makes that process visible. It names the human work hidden inside model fluency. It identifies the authors, editors, translators, coders, reviewers, teachers, journalists, researchers, and users whose textual production becomes part of the model's operative competence. It also clarifies why generative systems can appear intelligent: they condense the traces of distributed human labor into a single interface. The interface appears singular because the labor has been made invisible.

Part II therefore establishes the infrastructural basis for the rest of the paper. Structural appropriation is not an accidental by-product of occasional memorization. It begins with corpus extraction. The model's outputs are possible because prior human production has been absorbed, compressed, and operationalized without transparent attribution. The next part develops the mechanism through which that absorbed labor reappears as apparent novelty: recombinative plagiarism.

Part III. Recombinative Plagiarism

Recombinative plagiarism names the mechanism through which structural appropriation becomes visible as apparent originality. If corpus extraction is the first stage of generative appropriation, recombination is the second. Large language models do not usually reproduce prior texts as direct copies. They generate by reorganizing learned statistical relations into new surface forms. This technical fact is often presented as evidence that the output is not plagiarized. The argument is insufficient. A text may be non-identical to any prior source and still depend on the uncredited extraction of conceptual, stylistic, and argumentative structures. The absence of verbatim duplication does not prove intellectual independence. It may only prove that appropriation has passed through a deeper layer of transformation.

Classical plagiarism detection assumes that the copied object remains textually recoverable. Recombinative plagiarism breaks that assumption. The appropriated object

may no longer appear as a passage, sentence, or phrase. It may reappear as an argumentative architecture, a sequence of distinctions, a conceptual hierarchy, a recognizable mode of explanation, or a domain-specific rhetorical posture. In such cases, similarity is displaced from the lexical surface to the structural level. The generated text may pass conventional similarity checks because it does not reproduce the same words in the same order. Yet it may still reactivate intellectual labor extracted from prior texts. This is why recombinative plagiarism is harder to detect than ordinary plagiarism: its evidence is not local overlap, but patterned dependency.

The mechanism depends on statistical abstraction. During training, language models encode associations among words, phrases, genres, concepts, formats, and discourse patterns. These associations do not preserve texts as bibliographic objects. They convert them into distributed predictive relations. When the model generates new output, it does not retrieve authorship in the scholarly sense. It activates probable continuations based on learned regularities. The generated result can therefore combine elements derived from many prior sources without identifying any of them. The output is new at the surface and derivative at the infrastructural level. This combination is the central form of synthetic textual production.

This point matters because debates on AI plagiarism often remain trapped between two inadequate poles. One pole argues that plagiarism exists only when generated text reproduces a protected or identifiable source. The other pole argues that any model trained on prior texts is inherently plagiaristic. Both positions are too blunt. Recombinative plagiarism requires a more precise category. It does not claim that every generated sentence is plagiarism in the classical sense. It claims that generative systems create a new appropriation regime in which extracted intellectual structures can be reused without traceable attribution even when the final expression is formally distinct. The issue is not mere use of prior language. It is the systematic conversion of prior intellectual labor into outputs that conceal their dependency relations.

The concept of recombination has already been central to the analysis of plagiarism, knowledge commons, and large language models in *Borrowed Voices, Shared Debt* (Startari, 2025b). That framework treats idea recombination not as an innocent circulation

of influence, but as a debt-bearing process whose legitimacy depends on whether attribution, context, and intellectual lineage remain visible. The present argument extends that model. In large language models, recombination is no longer an occasional act performed by an author. It is the default mode of production. The system generates by recombining latent patterns at scale. As a result, the problem of intellectual debt is no longer exceptional. It is built into the architecture of predictive generation.

This differs from ordinary influence. Human authors are always shaped by prior reading, education, genre conventions, and disciplinary traditions. No writing begins from an empty field. But academic and literary cultures developed mechanisms to regulate influence: citation, quotation, bibliography, acknowledgment, intellectual history, peer review, and explicit positioning within a field. These mechanisms do not eliminate borrowing, but they make it accountable. Large language models weaken that accountability because they generate without preserving the sources from which their patterns were learned. The model may produce text influenced by thousands of prior works, but the output cannot distinguish influence from invention. The result is recombination without memory.

Recombinative plagiarism therefore has three core features. First, it is non-local. It does not depend on one identifiable source, but on distributed relations across a corpus. Second, it is non-verbatim. Its appropriated object is often structural rather than lexical. Third, it is non-reconstructible. The output usually does not allow readers to recover the sources whose labor shaped it. These features separate it from paraphrase plagiarism, patchwriting, and conventional unattributed borrowing. In paraphrase plagiarism, a source can often be located. In patchwriting, fragments can be compared. In recombinative plagiarism, the source relation is dispersed across the model's latent space. The debt persists, but the creditor cannot be named with ordinary tools.

The most important object of recombinative plagiarism is not the sentence. It is the pattern. A language model can appropriate the structure of an argument without reproducing the argument's original wording. It can reproduce the order in which a field explains a problem: definition, historical background, critique, case example, normative conclusion. It can reproduce the balance between caution and assertion that gives academic prose its authority. It can reproduce the genre logic of legal reasoning: issue, rule, application,

conclusion. It can reproduce the explanatory style of technical manuals, the narrative pacing of journalism, or the institutional voice of policy documents. These are not accidental ornaments. They are forms of intellectual organization.

A generated paragraph may therefore appear original because its words are new, while its intellectual architecture is inherited. This is the failure point of similarity-centered systems. They treat language as a surface. Recombinative plagiarism requires treating language as form, sequence, and function. The plagiarized element may be the arrangement of claims, the movement from premise to conclusion, the recurrent pairing of concepts, or the rhetorical distribution of certainty and doubt. These structures are often more valuable than individual sentences because they organize how knowledge becomes persuasive. To appropriate them without trace is to extract not only language, but the labor of conceptual formation.

The problem becomes sharper when models generate domain-specific outputs. In legal writing, recombination may reproduce doctrinal templates, argumentative pathways, or interpretive patterns learned from prior legal analysis. In academic writing, it may reproduce literature-review structures, theoretical distinctions, or methodological scaffolds. In journalism, it may reproduce narrative framing and investigative synthesis. In software development, it may reproduce coding idioms, documentation patterns, or solution architectures derived from open-source communities. In each case, generated output may differ from any one source, but still benefit from the extracted labor of a field. The model's originality is therefore not autonomous. It is assembled.

This assembled character is what distinguishes synthetic originality from genuine originality. Genuine originality does not require absolute independence from prior work. It requires accountable transformation. A scholar may build on sources, challenge them, cite them, and produce a new argument. The novelty is legitimate because the relation to prior work is made visible enough for evaluation. Synthetic originality, by contrast, emerges when the system recombines prior patterns while concealing the lineage of transformation. The output appears new because the path of dependency is unavailable. The surface is original, but the infrastructure is appropriative.

This dynamic connects directly to the problem of source disappearance developed in *Ethos Without Source* (Startari, 2025c). There, authority was shown to survive even when its originating subject becomes unavailable. In recombinative plagiarism, originality survives even when its originating sources become unavailable. The generated text speaks with the tone of intellectual competence while lacking recoverable authorship behind its structures. It therefore produces ethos through recombination: a credible surface assembled from prior forms of credibility. The system does not merely borrow content. It borrows the conditions under which content appears authoritative.

The connection with *Indexical Collapse* is also direct. If reference can disappear while authority remains, then generated text can function without exposing where its claims, structures, or stylistic authority come from (Startari, 2025d). Recombinative plagiarism is an application of that collapse to authorship. The output may refer to concepts, fields, or arguments, but its own dependency structure remains indexically unstable. It cannot say, in any precise sense, "this came from here." It can only generate a plausible continuation. The model's inability to restore stable origin is not peripheral. It is the condition that allows appropriation to remain hidden.

A further problem is that recombination can be mistaken for creativity. Because the output is not identical to any source, users may infer that the model has produced an original synthesis. This inference is weak. Novel arrangement is not necessarily original authorship. A system can combine extracted materials in statistically novel ways without generating independent intellectual responsibility. The distinction is critical. Creativity involves more than variation. It involves situated judgment, accountable selection, and a traceable relation to prior discourse. Predictive recombination produces variation without necessarily preserving judgment, accountability, or lineage.

This does not mean that all machine-assisted recombination is illegitimate. The question is not whether recombination occurs, but under what conditions it is disclosed, governed, and attributed. Human knowledge itself develops through recombination. Scientific fields advance by connecting existing theories, modifying inherited models, and reusing methods in new domains. The difference is that legitimate recombination remains embedded in accountable scholarly practice. It cites, contests, locates, and contextualizes. Large

language models can bypass these practices by producing recombined output as if it were self-originating. The problem is therefore not recombination itself. It is recombination without provenance.

This condition also explains why post-hoc citation is insufficient. A user may ask a model to provide sources after generating an answer. The model may then list plausible references. But such references do not necessarily represent the true lineage of the output. They may be relevant to the topic, but not causally connected to the generation. This distinction is essential. A citation added after generation can support a claim, but it cannot prove that the claim's structure originated from the cited source. Post-hoc citation may create bibliographic respectability while leaving recombinative debt unresolved. It can legitimize the surface without exposing the infrastructure.

The problem was partially anticipated in *Citation by Completion*, where predictive writing systems were shown to influence academic credit by shaping citation suggestions and authority-bearing syntax (Startari, 2025a). That paper focused on the redistribution of visible citation credit during writing. Recombinative plagiarism identifies a prior and more opaque stage. The model may generate the conceptual structure before any citation is requested. Once the argument exists, citations can be attached as decoration or support, but the real source-dependency remains hidden. The uploaded draft confirms that *Citation by Completion* centers on citation concentration, novelty reduction, and authority syntax in predictive writing systems, which makes it a direct precursor to the present analysis.

Recombinative plagiarism also reveals why plagiarism cannot be reduced to copyright. Copyright generally protects expression, not abstract ideas, methods, or styles in the broadest sense. But structural appropriation often operates precisely at those levels: argumentative architecture, disciplinary form, conceptual pacing, and rhetorical organization. The law may not treat these as protected expression in many cases. That does not make the appropriation epistemically irrelevant. Academic legitimacy depends on more than legal ownership. It depends on visible intellectual lineage. A generated text can be legally non-infringing and still epistemically appropriative if it converts prior intellectual structures into unattributed output.

The distinction between legal infringement and epistemic appropriation must remain central. A plagiarism theory for LLMs cannot simply wait for courts to determine whether training or output generation violates copyright. Legal categories move through specific doctrines of protectable expression, substantial similarity, fair use, licensing, and market harm. Structural appropriation concerns a different but overlapping question: whether a system extracts and redeploys intellectual labor while preventing the recovery of attribution. The answer may be epistemically significant even when legal liability remains uncertain. Plagiarism is not exhausted by copyright.

The deepest form of recombinative plagiarism occurs when a model absorbs conceptual innovations and later returns them as generic knowledge. A distinctive framework, once repeated enough across accessible texts, may become part of the model's predictive repertoire. The model can then reproduce adjacent formulations without naming the originator. Over time, the framework may be detached from its author and reappear as anonymous explanatory language. This is not merely loss of citation. It is the de-authoring of conceptual labor. The system converts authored innovation into unattributed linguistic availability.

This de-authoring process is especially damaging for emerging scholars, independent researchers, small-language communities, and non-dominant intellectual traditions. Their work may be absorbed into training data without achieving the citation density necessary to remain visible in outputs. Dominant authors may be named because they are statistically frequent. Less visible authors may contribute to the model's conceptual capacity while disappearing from generated attribution. The result is a double inequality: their labor is available to the system, but their recognition is not. Recombinative plagiarism therefore reinforces existing asymmetries in the knowledge economy.

This dynamic parallels the Matthew effect in science, where already recognized scholars accumulate disproportionate visibility and credit (Merton, 1968). In predictive systems, the effect is intensified because statistical frequency can govern both generation and attribution. High-frequency names are more likely to be reproduced as citations. Low-frequency contributions are more likely to be absorbed without name recovery. The model thus separates contribution from recognition. It can use the intellectual labor of the less

visible while attributing authority to the already dominant. This is not merely a bibliometric distortion. It is a structural redistribution of authorship.

Recombinative plagiarism must therefore be understood as a regime of asymmetrical transformation. The system transforms texts into latent capacity, transforms latent capacity into new outputs, and transforms source dependency into surface originality. At each stage, attribution weakens. The original author loses visibility. The model gains productive competence. The platform gains commercial or institutional value. The user gains fluent output. The only element that disappears is the chain of intellectual debt. This disappearance is not incidental. It is what allows the output to appear autonomous.

The category also clarifies why transparency cannot be limited to dataset lists. Even if a provider disclosed that certain domains or corpora were included in training, that would not identify which sources shaped a particular output. Recombinative plagiarism occurs at the relation between source patterns and generated structure. A dataset inventory may expose the possibility of extraction, but not the actual lineage of recombination. What is required is a more granular model of generative provenance: a system capable of indicating likely dependency clusters, domain influences, stylistic origins, and conceptual proximity to prior works. Without such mechanisms, recombination remains opaque.

This does not require perfect reconstruction of every influence. Academic citation has never provided perfect reconstruction either. It provides a socially enforceable minimum of intellectual traceability. A generative provenance system would serve an analogous function. It would not claim omniscience over every training relation. It would provide enough visibility to prevent synthetic originality from operating as a laundering mechanism. The goal is not exhaustive genealogy. It is accountable approximation.

Recombinative plagiarism also changes how originality should be evaluated in AI-assisted writing. A text should not be treated as original merely because it is statistically unique. Originality must be assessed through the relation between surface novelty, source transparency, and intellectual responsibility. A generated output may be unique in wording but derivative in structure. It may be novel in arrangement but opaque in debt. It may be

fluent in style but dependent on unacknowledged disciplinary labor. Without evaluating these layers, institutions will confuse lexical novelty with epistemic originality.

The educational implications are direct. Students and researchers using generated text may believe they are avoiding plagiarism because the output is "new." This belief reproduces the classical error that plagiarism equals copying. Under predictive generation, the risk is different. The user may submit a text whose structure, argument, and style derive from untraceable recombination. Even when the user edits the output, the underlying architecture may remain borrowed. The ethical question is not only whether the student copied from the model. It is whether the model produced a text whose intellectual debt cannot be disclosed. Academic integrity policies that ignore this distinction will remain obsolete.

The same applies to professional settings. A lawyer using generated analysis may unknowingly rely on patterns derived from prior legal memoranda. A journalist may publish a generated synthesis shaped by earlier reporting without credit. A consultant may deliver a strategy document whose conceptual structure reflects absorbed industry reports. A researcher may use generated literature framing that reproduces field hierarchies without visible attribution. In all cases, recombinative plagiarism operates by hiding the labor that made the output possible. The fact that the final wording is different does not resolve the problem.

Part III therefore establishes recombination as the operational mechanism of structural appropriation. Corpus extraction creates the hidden archive of absorbed labor. Recombinative generation reactivates that archive as surface novelty. The system avoids classical detectability because it does not need to copy directly. It can transform, rearrange, and synthesize. But transformation without provenance is not innocence. It is a more advanced form of appropriation.

The next part turns to the condition that makes recombinative plagiarism so difficult to contest: referential opacity. Once generated output loses recoverable links to its sources, attribution collapse becomes not an accidental failure, but a structural feature of predictive authorship.

Part IV. Referential Opacity and Attribution Collapse

Referential opacity is the condition under which generated text cannot disclose the source relations that made it possible. In classical authorship, attribution depends on the recoverability of reference. A claim, phrase, method, quotation, or conceptual frame can be linked back to a prior author, document, school, discipline, or archive. Even when citation is incomplete, the system of scholarly accountability presupposes that such links can in principle be reconstructed. Large language models disrupt that presupposition. Their outputs may be coherent, useful, and apparently original, but the path from source material to generated sentence is not available as a stable referential chain. The output appears without its genealogy.

This opacity is not merely a failure of user interface design. It is deeper than the absence of footnotes or bibliography. Referential opacity emerges from the way predictive systems transform source texts into distributed parameters. Once texts are absorbed into training, their role in later generation is no longer preserved as a document-level relation. The model does not usually generate by retrieving a specific passage from a specific source. It generates by activating learned relations across a high-dimensional statistical space. The result is a text whose dependency may be real but non-local, diffuse, and technically difficult to reconstruct. Attribution collapse begins at precisely this point: when the source remains operative but no longer referentially available.

The problem is therefore not that LLMs “forget” to cite. That formulation remains too weak. A missing citation presupposes that a citation could be supplied if the system or user behaved correctly. Referential opacity describes a stronger condition: the system may be unable to identify the actual source lineage of a generated output because the lineage has been dissolved into model weights. A citation added after generation may identify relevant literature, but relevance is not provenance. A source can support a claim without being the source from which the generated structure emerged. This distinction is central. Attribution requires more than plausible bibliographic support. It requires a recoverable relation between intellectual origin and textual production.

This is where structural appropriation becomes difficult to contest. If a generated text copies a protected passage verbatim, the source relation can be demonstrated through comparison. If the model reproduces a near-paraphrase, the relation may still be inferable. But when the output recombines patterns learned from many sources, no single comparison can establish the full debt structure. The system can produce language that depends on prior writing while leaving no direct textual fingerprint. The absence of a fingerprint is then mistaken for absence of appropriation. Referential opacity converts evidentiary difficulty into apparent innocence.

The collapse of attribution therefore has two dimensions. The first is technical: model architecture does not preserve transparent source-output mappings. The second is institutional: current academic, legal, and editorial systems are organized around artifacts that can be inspected, compared, and cited. They are not designed to evaluate latent dependency inside generative systems. Similarity detectors, citation audits, copyright analysis, and plagiarism policies all presuppose that the relevant relation can be located between texts. Referential opacity breaks that relation by moving appropriation into the infrastructure of generation.

This problem extends the argument developed in *Indexical Collapse*, where reference disappears while authority remains operative (Startari, 2025d). In that framework, predictive systems were analyzed as producing language that can function institutionally even when its referential anchors are weakened or absent. The present paper applies that logic to plagiarism and authorship. In structural appropriation, the generated output can function as an academic explanation, legal summary, journalistic synthesis, or technical answer while withholding the source relations that would allow intellectual debt to be evaluated. The text remains operational. Its origin becomes opaque.

The same condition also connects to *Ethos Without Source*, where credibility was shown to emerge without stable origin or accountable subject (Startari, 2025c). Referential opacity produces an equivalent authorship problem. The generated text may display the rhetorical signs of expertise: measured tone, formal structure, domain vocabulary, balanced qualification, methodological vocabulary, and citation-like syntax. These signs produce credibility. Yet the source of that credibility remains displaced. The system performs an

ethos assembled from prior human discourse without identifying whose intellectual labor contributed to that performance. The result is credibility without reconstructible debt.

Attribution collapse is intensified by the fact that outputs often arrive as complete textual objects. A generated response appears as if it were produced in a single present act. The interface hides the historical accumulation that made the output possible. This produces a false temporality of authorship. The user sees generation as immediate, but the output is possible only because of prior corpus extraction. The present sentence is built from past linguistic labor. Referential opacity conceals that temporal depth. It makes historical dependency appear as instantaneous production.

This false temporality matters for academic integrity. In scholarly writing, the time of intellectual formation matters. A concept has a history. A method has a lineage. A theoretical distinction emerges through debate, correction, refinement, and reuse. Citation does not simply decorate academic writing. It marks the temporal and social path by which knowledge becomes available. LLM output can erase that path. It produces statements that appear detached from the history of their formation. Attribution collapse is therefore not only a failure to name sources. It is a failure to preserve the historical structure of knowledge.

The problem becomes clearer when distinguishing three kinds of source relation: direct source, contributing source, and structural source. A direct source is a document from which a passage or idea is visibly borrowed. A contributing source is one among many texts that influence the model's representation of a topic. A structural source is a prior work or corpus segment whose form, method, argument sequence, or conceptual architecture shapes the generated output without being lexically reproduced. Classical plagiarism systems can sometimes detect direct sources. They struggle with contributing sources. They are largely blind to structural sources. Referential opacity is the condition in which structural sources remain active but invisible.

This invisibility is not ethically neutral. If a model generates a framework that resembles an existing author's conceptual architecture without reproducing exact language, the absence of exact overlap does not settle the matter. The relevant debt may lie in the

sequence of distinctions, the hierarchy of concepts, the formulation of a problem, or the mode of argument. These elements can carry substantial intellectual value. Their appropriation may be more consequential than copying a sentence because they shape how an entire field or problem is understood. Yet existing attribution mechanisms are poorly equipped to detect this form of borrowing.

This is why attribution collapse should not be reduced to citation failure. Citation failure is visible. Attribution collapse is structural. A user can fail to cite a source even when the source is known. A model can generate an output for which the source relation is not known by anyone involved in the immediate act of production. The user cannot cite what the system does not expose. The reader cannot verify what the output does not disclose. The institution cannot audit what the provider cannot or will not make available. The result is a distributed epistemic gap.

The literature on large language models and dataset opacity supports this concern. Bender, Gebru, McMillan-Major, and Shmitchell (2021) argue that large-scale language models are built on datasets whose composition, documentation, and social consequences are often insufficiently transparent. Under conditions of insufficient data transparency, attribution collapse becomes predictable. If the training corpus is opaque, then the generated output's source relation is also opaque. Dataset opacity is not merely a technical inconvenience. It becomes an authorship problem because the model's competence depends on materials whose contribution cannot be adequately traced.

Automation bias further compounds the issue. Parasuraman and Riley (1997) show that users can over-rely on automated systems when they treat outputs as reliable without inspecting the underlying process. In generated writing, this over-reliance takes a specific form: users accept fluent outputs as if fluency were evidence of legitimate origin. The output's coherence suppresses the question of provenance. Once the generated text sounds complete, the missing source chain becomes less visible as a problem. This dynamic allows attribution collapse to persist because the surface quality of the answer distracts from the opacity of its formation.

A related risk appears in post-hoc rationalization. Users often request sources after generating content. The system then provides citations that appear to ground the output. This creates a procedural illusion. The presence of references gives the text a scholarly surface, but those references may not correspond to the actual generative dependencies of the argument. They may be topic-relevant, authoritative, or plausible, but not genealogically accurate. Post-hoc citation therefore risks becoming a laundering device. It converts opaque generation into apparently accountable scholarship without solving the underlying provenance problem.

This risk was already visible in *Citation by Completion*, where predictive systems were shown to shape citation behavior through suggestion syntax, authority phrasing, and concentration effects (Startari, 2025a). That article examined how visible citation choices can be redistributed by predictive assistance. The present paper identifies a prior failure: before the system suggests whom to cite, it may already have generated content from source relations that are unavailable. The uploaded draft confirms that *Citation by Completion* treats predictive citation as a structural redistribution of academic credit through autocomplete and authority-bearing syntax, making it a necessary precursor to the present argument.

Referential opacity also creates a problem for legal responsibility. Copyright law and related doctrines often depend on questions of access, similarity, protectable expression, licensing, and market effect. These categories may address specific cases of memorization or direct reproduction. They do not fully address structural source dependence. If the output does not reproduce protected expression but still derives its value from absorbed intellectual labor, the law may fail to register the appropriation. The absence of infringement does not prove absence of epistemic debt. Legal non-liability and intellectual legitimacy are not identical categories.

This distinction is decisive for academic governance. Universities and journals often define plagiarism as presenting another's work or ideas as one's own without appropriate acknowledgment. Under LLM conditions, this definition encounters a practical barrier: the "another" may not be recoverable. The work may be distributed across thousands of sources. The idea may be embedded in a conceptual field rather than a single document.

The phrasing may be statistically assembled from genre conventions rather than copied from one author. If attribution requires a named source, then structurally appropriated outputs can escape accountability precisely because their debt is dispersed.

The dispersion of debt does not eliminate debt. It changes its form. Latent intellectual debt describes this condition: the output depends on prior intellectual labor, but the relation cannot be made visible through ordinary citation. The debt remains embedded in the generative capacity of the model. It appears as fluency, conceptual competence, rhetorical authority, or domain familiarity. But because it is latent, it cannot be assigned, measured, or repaid through conventional scholarly mechanisms. This is the central crisis of attribution collapse.

This crisis is not only theoretical. It affects evaluation. A teacher evaluating AI-assisted work may check for textual similarity and find none. A journal editor may ask whether references support the claims and receive an adequate bibliography. A legal reviewer may ask whether a passage copies protected expression and find insufficient overlap. In each case, the text passes a classical test while the deeper appropriation remains unresolved. Referential opacity creates a gap between compliance and legitimacy. A text can comply with existing detection systems while still being structurally dependent on unacknowledged sources.

The gap is widened by the commercial form of generative systems. Model providers often present outputs as services, not as derivative textual products. The interface frames generation as assistance, productivity, creativity, or search enhancement. This framing shifts attention away from the corpus. Users interact with the immediate tool, not with the archive of extracted labor behind it. The platform becomes the visible producer, while the corpus becomes invisible infrastructure. Attribution collapse is therefore reinforced by product design. The interface makes the system look originative.

The same product design also affects user responsibility. If the interface does not expose provenance, the user must either trust the system or conduct independent verification. But independent verification can only check the claims that appear in the output. It cannot reconstruct all structural sources that shaped the text. This means that responsibility is

asymmetrical. Users are held accountable for outputs, but they are not given full access to the dependencies necessary for responsible attribution. The ethical burden is shifted downward while provenance remains controlled upward by the system provider.

This asymmetry requires a new accountability model. It is insufficient to say that users must cite their sources if the system does not disclose the sources it has structurally absorbed. It is also insufficient to say that providers can avoid responsibility by treating generation as user-directed. Structural appropriation occurs across the full chain: dataset collection, training, model design, interface defaults, prompting, output selection, editing, and publication. Attribution collapse is therefore a system-level event. Accountability must be assigned across the infrastructure, not only at the final point of use.

A provenance-oriented model would begin by distinguishing support citation from generative provenance. Support citation identifies sources that justify or corroborate a claim. Generative provenance identifies sources or source clusters that materially shaped the production of the output. Current scholarly practice is built around support citation. LLM accountability requires movement toward generative provenance. The two can overlap, but they are not identical. A generated paragraph about a legal doctrine may cite a leading case as support while having been structurally shaped by secondary analysis, treatises, briefs, or prior generated outputs. Without generative provenance, the real dependency chain remains hidden.

The technical challenge is considerable, but not conceptually impossible. A system could provide probabilistic source mapping, domain influence indicators, similarity-to-training-cluster estimates, memorization-risk signals, or exposure summaries. These tools would not produce perfect attribution. They would produce accountable approximation. That is already how many scholarly systems work. Citations do not capture every influence. Peer review does not detect every dependency. But they create minimum conditions of traceability. Generative systems need analogous mechanisms because the absence of perfect provenance cannot justify total opacity.

This point connects to *AI and the Structural Autonomy of Sense*, where generated language was theorized as operating beyond stable referential anchoring (Startari, 2025e).

Referential opacity in plagiarism is one consequence of that autonomy. The text functions even when its source relation is unavailable. It can answer, persuade, summarize, recommend, and instruct without showing where its intellectual structures came from. This operational autonomy gives generated text its power. It also creates the conditions for appropriation without recoverable attribution.

Attribution collapse also produces a secondary epistemic effect: it weakens the distinction between knowledge and linguistic availability. If the model can generate a plausible answer, users may treat the answer as knowledge. But knowledge in scholarly contexts requires more than plausible language. It requires grounds, sources, methods, and accountability. Referential opacity allows linguistic availability to replace epistemic grounding. The system can produce language about a topic without exposing the documentary basis of that production. This is not only a citation problem. It is a transformation in the conditions under which claims become acceptable.

The consequences for originality are severe. Originality traditionally depends on a visible relation to prior work. A paper is original because it departs from, extends, revises, or challenges identifiable sources. Without source visibility, originality becomes a surface property. The text appears original if it is not identical to previous text. But under LLM conditions, surface novelty is insufficient. A generated text can be lexically unique and structurally derivative. It can be formally new and intellectually indebted. It can be undetectable as plagiarism while still being built from unattributed conceptual labor. Referential opacity makes this ambiguity permanent unless provenance systems are introduced.

This ambiguity also affects authors whose work enters training data. Their contribution may be absorbed but not recognized. Their concepts may be reused but not cited. Their style may shape outputs but not remain attached to their name. In academic fields, this can erode the relation between contribution and credit. In creative fields, it can erode the relation between style and authorship. In technical communities, it can erode the relation between open contribution and recognition. Attribution collapse therefore redistributes value away from originators and toward systems capable of recombination.

The redistribution is not evenly distributed. Highly visible authors may still be cited because their names occur frequently in training data and public discourse. Less visible authors may be absorbed without being named. This produces an inequality of recoverability. The already dominant remain referentially available. The marginal become latent. The model can learn from both, but attribution returns mainly to the visible. This dynamic reinforces the Matthew effect described by Merton (1968): accumulated recognition attracts further recognition. Under predictive generation, this effect becomes infrastructural. Frequency does not only shape citation. It shapes recoverability itself.

Referential opacity therefore has both epistemic and political consequences. Epistemically, it prevents accurate reconstruction of intellectual lineage. Politically, it determines whose work remains visible after extraction. The collapse of attribution is not a neutral technical limitation. It affects the distribution of credit, legitimacy, and economic value. The inability to trace sources becomes a mechanism by which extracted labor can be used without recognition.

The concept of attribution collapse allows these consequences to be named. Attribution collapse occurs when a generated output depends on prior intellectual labor but the system cannot or does not provide a recoverable chain of provenance sufficient for scholarly, legal, or ethical evaluation. It is not identical to absence of citation. It is the breakdown of the conditions that make citation meaningful. A citation can be added to a text after the fact. Attribution collapse concerns whether the text's actual dependency structure can be disclosed at all.

This distinction prepares the central normative claim of the paper. If LLMs produce referentially opaque outputs from structurally absorbed corpora, then plagiarism governance cannot remain focused on surface similarity. It must evaluate provenance, dependency, and traceability. The primary question becomes: what mechanisms must generative systems provide so that intellectual debt does not disappear into statistical transformation? Without such mechanisms, plagiarism becomes less detectable as models become more capable. Improvement in fluency may therefore worsen accountability.

Part IV establishes that referential opacity is the condition that allows recombinative plagiarism to persist. Corpus extraction supplies the hidden labor. Recombination transforms that labor into surface novelty. Referential opacity prevents the lineage from being reconstructed. Attribution collapse is the result. The next part turns to the authorship effect produced by this collapse: synthetic originality, the appearance of autonomous creation generated by systems built from absorbed human intellectual labor.

Part V. Synthetic Originality and Predictive Authorship

Synthetic originality is the appearance of autonomous creation produced by predictive recombination. It occurs when generated text appears new because it does not reproduce any prior source verbatim, while its conceptual, stylistic, or argumentative structure remains dependent on absorbed human intellectual labor. Under this condition, originality becomes a surface effect. The text differs from its sources, but the difference is produced by statistical transformation rather than by accountable intellectual authorship. The result is not simple imitation. It is originality without recoverable origin.

Classical originality presupposes a relation between inheritance and contribution. An author may borrow from prior works, but the legitimacy of the new work depends on how that borrowing is transformed, cited, challenged, extended, or recontextualized. Originality does not mean absolute independence. It means accountable difference. A scholarly article, legal argument, literary work, or technical framework becomes original when its relation to prior material can be evaluated. The reader can ask what is inherited, what is modified, what is contested, and what is newly contributed. This evaluative structure depends on source visibility.

Large language models weaken that structure because they generate difference without preserving lineage. Their outputs may be lexically distinct, formally coherent, and contextually useful. Yet the process that generates this difference does not provide an adequate account of the intellectual materials transformed into the final text. The model produces novelty at the level of surface arrangement. It does not produce authorship in the classical sense because it does not assume responsibility for the relation between

inheritance and contribution. Predictive generation therefore creates a new authorship problem: the text appears original, but the conditions of originality remain unavailable.

This is the core of synthetic originality. It is not false originality in the simple sense that the output is always worthless or wholly derivative. It is synthetic because it emerges from recombination across absorbed patterns. It is original because the surface configuration may be statistically unique. It is problematic because the uniqueness cannot be separated from untraceable dependency. The generated text may be new in wording while old in structure. It may present a novel arrangement while relying on conceptual labor already performed elsewhere. It may appear as a fresh synthesis while hiding the corpus-based conditions that made the synthesis possible.

The distinction between surface novelty and epistemic originality is therefore essential. Surface novelty refers to the absence of direct textual identity with prior sources. Epistemic originality refers to a contribution whose relation to prior knowledge is visible enough to be judged. Current plagiarism detection systems are much better at measuring surface novelty than epistemic originality. If generated output does not match existing documents, it may pass as original. But this is a weak standard. In predictive systems, the very mechanism of generation can produce enough variation to evade similarity detection while still depending on unacknowledged intellectual extraction.

This condition follows directly from the theory of recombinative plagiarism developed in the previous section. Corpus extraction supplies the hidden archive. Recombination reorganizes that archive into new surface forms. Referential opacity prevents the lineage from being reconstructed. Synthetic originality is the authorship effect produced by this chain. It is the moment when structural appropriation becomes aesthetically and institutionally acceptable because the output looks new. The problem is not that the model copies too visibly. The deeper problem is that it appropriates invisibly enough to appear original.

The same logic extends the argument of *Citation by Completion*, where predictive systems were shown to influence academic recognition by shaping citation suggestions, authority-bearing syntax, and source concentration during writing (Startari, 2025a). That earlier

framework examined how predictive systems redistribute visible credit. Synthetic originality identifies the prior condition under which visible credit can be bypassed altogether. If generated text already appears original before attribution is considered, then citation becomes a secondary repair mechanism. It can decorate the final text, but it cannot reconstruct the full dependency structure of the generative process. The uploaded draft of *Citation by Completion* confirms this continuity by defining predictive writing as a system that alters citation concentration, novelty, and legitimacy phrasing through autocomplete and authority-bearing syntax.

Predictive authorship is the operational form of synthetic originality. It describes a mode of textual production in which the model generates output by selecting probable continuations from learned distributions, while the human user supplies prompts, selection, editing, and final authorization. The resulting text is neither purely human nor autonomously machine-authored. It is produced through a layered relation between corpus, model, interface, prompt, and user. Yet most publication systems still require a singular attribution structure. They ask who wrote the text. Predictive authorship makes that question unstable because the text emerges from distributed, asymmetrical, and partially opaque contributions.

The user may be responsible for initiating the task, refining the output, approving the text, and placing it into circulation. That responsibility is real. But responsibility is not equivalent to origin. The user did not create the model's linguistic capacity. The model did not create the corpus from which that capacity was extracted. The corpus contributors did not authorize each specific recombination. The platform mediates the relation while often withholding provenance. Authorship becomes dispersed across a chain that the final text does not disclose. Predictive authorship is therefore not simply co-authorship. It is authorship under conditions of hidden dependency.

This distinction matters for academic integrity. If a researcher uses an LLM to draft a paragraph, the immediate question is often whether the researcher should disclose AI assistance. Disclosure is necessary but insufficient. Stating that a text was AI-assisted identifies the tool. It does not identify the intellectual debt embedded in the tool's output. The deeper problem is not only whether the researcher used a model. It is whether the

model generated structures, claims, formulations, or argumentative sequences derived from absorbed sources that remain invisible. AI disclosure answers the question of instrument use. It does not answer the question of provenance.

Synthetic originality therefore creates a false sense of compliance. A paper may disclose that AI was used. It may include references. It may pass plagiarism detection. Yet the generated sections may still contain structural dependencies that cannot be traced. The institutional record appears clean because current compliance mechanisms focus on visible authorship markers, declared tools, and textual similarity. Structural appropriation survives because it operates below that level. It is not erased by disclosure alone. It requires source visibility mechanisms capable of distinguishing support citation from generative provenance.

This problem also affects the meaning of creativity. Many descriptions of generative AI emphasize the capacity to produce novel combinations. But novelty is not identical to creativity in the strong intellectual sense. A system can produce variation without situated judgment. It can generate alternatives without responsibility for their epistemic lineage. It can assemble plausible forms without understanding their historical, conceptual, or disciplinary stakes. Synthetic originality should therefore not be equated with creative authorship. It is a formal effect of high-dimensional recombination. It may be useful, but usefulness does not settle the question of attribution.

The same issue appears in the relation between style and ownership. A model may generate text that does not copy a living author's sentences, yet reproduces recognizable features of a genre, field, or authorial posture. In some cases, the system may imitate a specific author if prompted. In other cases, it may absorb diffuse stylistic patterns from many authors and return them as generic fluency. The latter is harder to contest. Style appears as a public resource once it is statistically generalized. But style is also labor. Academic tone, legal precision, journalistic framing, and technical clarity are not natural properties of language. They are accumulated practices produced by communities. Synthetic originality conceals this labor by presenting style as model competence.

This connects directly to *Ethos Without Source*, where credibility was theorized as an effect that can survive after the disappearance of a stable authorial source (Startari, 2025c). Synthetic originality extends that problem from credibility to creation. The generated text appears to possess the ethos of intellectual production without exposing the origin of the structures that sustain that ethos. It speaks as if it had authored itself. Its authority derives from absorbed patterns, but those patterns are returned as a singular, source-neutral output. The system does not merely generate text. It generates the appearance of authorship.

The appearance of authorship is reinforced by interface design. Most LLM interfaces present output as a direct response to user input. The user asks, the system answers. This conversational structure hides the corpus. It makes generation appear immediate and dialogic rather than historical and extractive. The model seems to produce from itself because the interface does not display the chain of absorbed labor behind the answer. This design produces what can be called interface authorship: the illusion that the visible speaker is the origin of the text. In reality, the visible speaker is only the final node in a much longer production chain.

This interface illusion has legal and academic consequences. Legal doctrines of authorship tend to require an identifiable author or rights holder. Academic conventions require named responsibility and citation. Journalistic standards require traceable sourcing. Predictive authorship destabilizes each of these systems because the visible producer is not the full originator. The user may be accountable, but not fully original. The model may be generative, but not authorial in the human sense. The corpus may be constitutive, but not credited. The platform may be infrastructural, but often not treated as a co-producer of specific textual claims. Synthetic originality exists inside this gap.

The gap becomes especially dangerous when generated text is treated as neutral assistance. If the model is framed as a tool equivalent to a spelling checker or grammar assistant, then its contribution appears merely formal. But LLMs do not only correct surface errors. They can supply claims, examples, arguments, categories, transitions, definitions, and entire conceptual structures. Their contribution is substantive. Treating such systems as neutral tools hides the degree to which they shape intellectual production. The more substantive the generation, the more serious the problem of attribution becomes.

Automation bias intensifies this problem. Parasuraman and Riley (1997) show that users can misuse automated systems by over-relying on outputs without adequate verification. In predictive authorship, over-reliance appears as acceptance of generated originality. The user sees a coherent answer and treats coherence as evidence that the text is independently usable. The missing provenance is not examined because the output satisfies the immediate communicative need. This is not merely a user error. It is a predictable consequence of systems designed to produce fluent completion without exposing dependency.

Synthetic originality also affects the economics of writing. If a model can produce apparently original text from absorbed human labor, then markets may substitute generated outputs for commissioned human writing, research assistance, translation, documentation, journalism, or educational support. The economic value extracted from prior human production returns as a competing product. The original contributors are not compensated. The system that absorbed their labor becomes the seller of synthetic originality. This is the economic dimension of structural appropriation. The model does not only produce text. It produces marketable originality from unremunerated textual histories.

This marketable originality creates a structural asymmetry. Human authors must cite, document, and defend their contributions. Generative systems can produce text without comparable source accountability. Human writers can be accused of plagiarism when they fail to disclose sources. LLMs can produce untraceable recombinations at scale, while responsibility is shifted to users or treated as an unresolved policy issue. The asymmetry is not sustainable. If originality is required from human authors, provenance must be required from systems that generate on the basis of human-authored corpora.

The problem becomes sharper in academic publication. A human scholar's originality is judged by relation to literature. Reviewers ask whether the paper contributes something new. That judgment depends on visible engagement with prior work. LLM-assisted writing can simulate such engagement by producing literature-review language, citation frames, methodological summaries, and theoretical positioning. But if these are generated from opaque corpus patterns, the paper may present a false map of its intellectual origin. It may cite some sources while being structurally shaped by others. Synthetic originality can

therefore contaminate peer review by making derivative structures appear as authorial contribution.

This does not mean that AI-assisted scholarship is impossible or illegitimate. It means that legitimacy depends on control, disclosure, and provenance. A scholar may use generative tools for drafting, comparison, language refinement, or exploratory synthesis. But the scholar must remain responsible for verifying claims, selecting sources, marking contribution, and ensuring that the final argument is not merely an opaque recombination of uncredited structures. The problem begins when predictive output is treated as independent intellectual production rather than as a generated surface requiring audit. The standard must be higher than textual uniqueness.

The category of synthetic originality also clarifies why model outputs can feel persuasive. The model has absorbed the forms through which authority is usually communicated. It can produce balanced introductions, precise definitions, structured taxonomies, cautious qualifications, and confident conclusions. These forms make the text appear authored. But authorship is not only form. It is accountable relation to knowledge. When form is detached from provenance, the system produces the signs of originality without its obligations. That detachment is the central risk.

This risk connects to the broader theory of post-referential operative representation in *AI and the Structural Autonomy of Sense* (Startari, 2025e). Generated language can operate effectively even when its referential grounding is weakened. Synthetic originality is one manifestation of that autonomy. The output circulates as original because it functions as original: it answers the prompt, fills the page, supports an argument, or satisfies an institutional requirement. Its functionality masks the absence of recoverable origin. The text works, therefore it is accepted. But operational success is not proof of intellectual independence.

The relation to *Indexical Collapse* is equally direct. In predictive systems, reference can disappear while authority remains (Startari, 2025d). In synthetic originality, source reference disappears while the appearance of creation remains. The text no longer points back to the conditions that generated it. It points forward to its use. It is evaluated by utility,

coherence, and novelty of surface, not by its intellectual lineage. This produces an indexical failure in authorship: the output cannot adequately answer the question, "from where does this structure come?"

The inability to answer that question produces attribution instability. If the user claims authorship, the claim is incomplete because the generated structures derive from opaque model capacity. If the model is treated as author, the claim is conceptually weak because the model lacks situated responsibility and does not disclose its debts. If the corpus contributors are treated as authors, the claim is technically unmanageable because their contributions cannot be individually reconstructed. Predictive authorship therefore produces a triangular instability among user, model, and corpus. Synthetic originality is the textual form taken by that instability.

This instability is not solved by calling LLMs "tools." The tool analogy is inadequate when the tool supplies substantive content. A microscope extends vision but does not write the interpretation. A calculator performs operations selected by the user, but does not construct a theoretical argument. A grammar checker modifies form but does not ordinarily generate the conceptual frame of a paper. LLMs can do all of these things at once: retrieve-like synthesis, stylistic transformation, argument construction, and rhetorical presentation. Their role exceeds mechanical assistance. They are generative intermediaries.

As generative intermediaries, LLMs participate in authorship without satisfying the conditions of authorship. They shape textual outcomes, but they do not assume responsibility. They draw from prior human production, but they do not attribute comprehensively. They produce coherent arguments, but they do not preserve the history of those arguments. This is why predictive authorship must be treated as a distinct category. It is neither ordinary human writing nor simple machine output. It is structurally mediated writing under conditions of partial opacity.

This category also clarifies why generative AI creates institutional pressure. Institutions want productivity, efficiency, and scalable writing. Synthetic originality satisfies those demands because it produces usable text quickly. But the same efficiency weakens the slow mechanisms that preserve intellectual legitimacy: reading, citation, source comparison,

revision, and argumentative accountability. The more efficient generation becomes, the easier it is to bypass the labor that makes originality meaningful. Synthetic originality therefore functions as an institutional temptation. It offers the appearance of contribution without the full cost of intellectual formation.

The educational consequences follow directly. If students learn to treat generated originality as acceptable whenever it is not textually copied, they will internalize a reduced definition of authorship. They will learn that originality means absence of detectable overlap. That definition is false. Originality requires accountable positioning within a field of prior work. AI literacy must therefore include provenance literacy. Students must understand that generated text can be unique, useful, and still structurally indebted. Without that distinction, academic integrity collapses into a technical game of detection avoidance.

Professional consequences are similar. In law, synthetic originality may produce memos that appear analytically independent while relying on opaque doctrinal patterns. In journalism, it may produce articles that appear newly synthesized while drawing from prior reporting frames. In business, it may produce strategy documents that appear customized while recycling absorbed consulting language. In software, it may produce solutions that appear newly written while drawing from uncredited code communities. Across these domains, the core problem is not only whether the output is correct. It is whether the output's apparent originality hides an unacknowledged chain of labor.

This does not imply that all generated outputs must be prohibited. It implies that generated originality must be downgraded as evidence. A text should not be treated as original merely because it is new to the user, new in wording, or undetected by plagiarism software. The burden should shift toward provenance, process documentation, and human verification. The user must show how the final text was produced, what sources were actually consulted, what claims were independently verified, and which parts were generated, edited, or authored. For high-stakes domains, the standard must include traceability, not merely disclosure.

Synthetic originality therefore requires a change in evaluative criteria. The relevant questions are not limited to: Is the text copied? Is the wording unique? Did the user disclose

AI assistance? The stronger questions are: What intellectual labor made this output possible? What source relations can be reconstructed? Which parts of the argument are human-authored, model-generated, or corpus-derived? Are the cited sources actually generative sources or only post-hoc supports? Does the text contain structural dependencies that remain hidden? Without these questions, institutions will mistake generative fluency for legitimate originality.

The concept also exposes why generative systems can become infrastructures of authorship. They do not merely assist writers within existing authorship norms. They redefine the conditions under which text appears as authored. A system that can produce coherent academic, legal, journalistic, or technical prose on demand becomes a standing source of synthetic textual legitimacy. It supplies not only words but the form of contribution. Once integrated into workflows, it can normalize predictive authorship as ordinary authorship. That normalization is the point at which structural appropriation becomes institutionally invisible.

A serious governance framework must therefore distinguish between three layers: textual production, intellectual contribution, and provenance accountability. Textual production concerns who or what generated the words. Intellectual contribution concerns who produced the concepts, structures, methods, or arguments that give the text value. Provenance accountability concerns whether those relations can be disclosed and evaluated. Current AI disclosure policies usually address the first layer. Classical citation practices address part of the second. Neither fully addresses the third under LLM conditions. Synthetic originality emerges in the gap between these layers.

The solution cannot be purely rhetorical. It must be procedural. Authors using generative systems should maintain process logs, source lists, prompt records, revision histories, and verification notes. Platforms should provide source-influence indicators, provenance approximations, memorization risk warnings, and distinction between retrieved support and generated synthesis. Journals and institutions should require disclosure of AI-generated sections and independent source verification. These measures would not eliminate synthetic originality. They would prevent it from masquerading as unmediated authorship.

The theoretical conclusion is direct. Synthetic originality is the authorship form of structural appropriation. It is what happens when corpus extraction, recombinative generation, and referential opacity produce a text that appears independently created. Predictive authorship is the production regime that sustains it. Together, they transform plagiarism from an act of copying into a condition of generated authorship without recoverable lineage. The text is new, but its newness is not enough. Without provenance, novelty becomes a mask.

Part V therefore establishes that originality in LLM-generated text cannot be evaluated at the surface level. The absence of duplication does not prove independence. The presence of fluency does not prove authorship. The disclosure of AI assistance does not prove attribution. A new framework is required, one that treats originality as a relation among surface novelty, intellectual debt, and recoverable provenance. The next part turns to the institutional consequences of failing to adopt that framework: the crisis of academic and legal legitimacy produced by structural appropriation at scale.

Part VI. The Crisis of Academic and Legal Legitimacy

Structural appropriation produces an institutional crisis because academic and legal systems still evaluate originality through frameworks designed for visible authorship. These frameworks assume that intellectual production can be attributed, compared, cited, challenged, and assigned to identifiable agents. Large language models weaken each of those assumptions. They generate texts whose surface may appear original, whose claims may appear grounded, and whose style may appear professional, while the source relations that shaped the output remain opaque. The crisis is therefore not only that LLMs can produce plagiarized passages. The deeper crisis is that they can produce institutionally acceptable texts whose dependency structures cannot be adequately audited.

Academic legitimacy depends on traceable contribution. A scholarly text is not legitimate merely because it is fluent, coherent, or formally original. It must situate itself within prior work, identify its sources, distinguish inheritance from contribution, and expose enough of its method for evaluation. Citation is not ornamental. It is the mechanism through which a

text enters a field of accountability. It allows readers to verify claims, reconstruct intellectual lineage, assess originality, and determine whether the author has properly acknowledged prior work. When predictive systems generate text without recoverable provenance, they undermine the conditions that make citation meaningful.

The problem is not solved by adding citations after generation. A generated paragraph may be followed by plausible references, but those references do not necessarily disclose the actual source relations that produced the paragraph. They may support the topic, but they do not prove generative lineage. This distinction is central to the crisis of academic legitimacy. The bibliography can become a stabilizing fiction: it gives the text an appearance of scholarly accountability while failing to reveal the absorbed intellectual labor that shaped its structure. Under LLM conditions, citation can become post-hoc legitimation rather than genuine attribution.

This issue extends the argument developed in *Citation by Completion*, where predictive writing systems were shown to shape academic credit by influencing citation concentration, novelty, and authority-bearing syntax during composition (Startari, 2025a). That framework demonstrated that LLM writing aids can redistribute visible academic credit by steering which sources appear in a text and how those sources are framed. The present paper identifies a prior and broader institutional problem. Even before citation is suggested, generated language may already depend on absorbed source structures that cannot be identified. The uploaded draft of *Citation by Completion* confirms this continuity, since it defines predictive citation systems as mechanisms that alter source diversity, legitimacy phrasing, and citation concentration through autocomplete and authority syntax.

Academic institutions are poorly equipped for this condition because their integrity systems remain surface-oriented. Plagiarism software detects textual overlap. Citation audits inspect whether references are present and relevant. Peer reviewers evaluate whether the argument appears novel and properly situated. These mechanisms are necessary but insufficient. They assume that the relevant problem appears at the level of document comparison. Structural appropriation appears below that level, in the relation between

corpus extraction, recombinative generation, and referential opacity. A text can pass similarity detection, include relevant citations, and still contain untraceable structural debt.

This creates a gap between compliance and legitimacy. A student, researcher, journalist, or professional writer may comply with existing rules by disclosing AI use, avoiding verbatim copying, and attaching references. Yet the generated text may still be built from unacknowledged conceptual or stylistic structures extracted from prior human production. Compliance becomes easier than accountability. The institution can verify that no obvious copying occurred, but it cannot verify whether the text's apparent originality rests on opaque recombination. The absence of detectable plagiarism becomes a weak substitute for evidence of genuine contribution.

The crisis is especially severe in scholarly publishing. Peer review depends on the reviewer's ability to evaluate contribution against prior literature. If LLM-generated text can simulate literature positioning, methodological framing, and theoretical novelty without exposing its generative dependencies, reviewers face a new evidentiary problem. A paper may look properly situated while its actual intellectual formation remains hidden. The review process can assess the visible argument, but it cannot easily determine whether the argument is an accountable contribution or a synthetic recombination of uncredited structures. This weakens the epistemic function of peer review.

The crisis also affects citation metrics. Citation systems traditionally measure visible recognition. They are already imperfect indicators of influence, as citation behavior reflects social, disciplinary, and institutional factors as well as intellectual contribution. Under predictive generation, the distortion deepens. LLMs can both absorb uncredited work and amplify already visible sources through citation suggestions. Merton's (1968) Matthew effect becomes computationally reinforced: dominant authors remain citeable and recoverable, while less visible contributors may be absorbed into model capacity without appearing as named sources. The result is a separation between contribution and recognition. Some works help produce the model's competence, while other works receive the visible credit.

This produces a structural injustice in academic knowledge economies. Scholars whose texts are frequent, canonical, or institutionally dominant may continue to appear in generated citations. Scholars whose work is less frequent, non-English, independent, regionally situated, or outside dominant indexing systems may contribute to the model's training substrate without being recoverable in outputs. Their intellectual labor becomes usable but not citeable. The model can learn from them without naming them. This is not only a technical failure. It is a redistribution of scholarly visibility.

The legal crisis follows a parallel structure. Copyright law generally asks whether protected expression has been copied, whether the copying is substantial, whether the use is licensed or excused, and whether the market for the original work is harmed. These questions remain relevant for cases involving memorization, direct reproduction, or dataset licensing. But structural appropriation often operates below the threshold of protectable expression. It may reuse conceptual architecture, stylistic formation, argumentative order, or disciplinary convention without reproducing a specific protected passage. Legal analysis may therefore find no infringement while the epistemic problem remains intact.

This distinction between legal infringement and intellectual appropriation is decisive. A generated text can be legally defensible and still epistemically illegitimate. Copyright does not exhaust plagiarism. Plagiarism concerns attribution, authorship, and intellectual honesty. Copyright concerns legally protected expression and authorized use. The two can overlap, but they are not identical. LLMs exploit the gap between them. They can transform prior labor enough to avoid obvious infringement while preserving enough structural dependency to benefit from the appropriated work. In that gap, synthetic originality becomes legally resilient but academically unstable.

The problem becomes even more complex when training data is publicly accessible. Public availability is often treated as if it weakens claims of appropriation. That inference is flawed. A text made public for reading, citation, education, preservation, or debate is not automatically made available for conversion into a commercial generative substrate. Public visibility does not equal unrestricted transformation into predictive infrastructure. The act of reading a text and the act of training a model on it are structurally different. Reading

preserves the text as an object of interpretation. Training converts it into productive capacity. Citation acknowledges dependency. Extraction dissolves it.

Legal systems also face difficulty assigning responsibility. In classical plagiarism, responsibility can often be assigned to the person who copied, paraphrased, or failed to cite. In predictive authorship, responsibility is distributed. The dataset collector may have selected or scraped the material. The model provider may have trained the system. The interface designer may have hidden provenance. The user may have prompted, accepted, edited, and published the output. The institution may have normalized use without adequate standards. No single actor fully contains the act of appropriation, yet the appropriation occurs through their combined structure. This creates accountability diffusion.

Accountability diffusion is one of the central legal and institutional consequences of structural appropriation. When responsibility is distributed, each actor can reduce their own burden by pointing elsewhere. Providers may claim that users control outputs. Users may claim that the system generated the text. Institutions may claim that disclosure policies are sufficient. Developers may claim that source reconstruction is technically difficult. The result is a chain in which no actor fully acknowledges the absorbed intellectual labor behind generated output. Attribution collapse therefore becomes accountability collapse.

The crisis extends into journalism. Journalistic legitimacy depends on sourcing, verification, and editorial responsibility. A generated article can reproduce explanatory frames, story structures, background summaries, and analytic sequences learned from prior reporting without crediting the journalists or outlets whose work shaped those frames. Even when no sentence is copied, reporting labor may be converted into generic generated context. The risk is not limited to factual error. It concerns the transformation of sourced reporting into unsourced synthesis. If generated journalism becomes normalized, the labor of investigation may be absorbed by systems that produce secondary narratives without preserving the debt to original reporting.

Legal writing faces a similar problem. Generated memoranda, case summaries, contract analyses, or regulatory explanations may reproduce doctrinal patterns and interpretive structures learned from prior legal commentary, briefs, treatises, or judicial texts. The

output may not copy any single work, but it can still benefit from accumulated legal analysis. If the sources are not recoverable, the user cannot evaluate whether the argument rests on authoritative law, secondary commentary, outdated doctrine, or hallucinated synthesis. In legal contexts, attribution collapse becomes a risk to professional responsibility because the origin and reliability of legal reasoning matter.

Technical and software communities face another form of the same crisis. Code generation systems may produce solutions structurally shaped by open-source repositories, documentation, forum answers, or prior examples. Even where exact code copying is absent, the generated solution may depend on patterns created by communities that relied on licenses, attribution norms, and collaborative recognition. If generative systems absorb those patterns and return them without provenance, the norms of open contribution are weakened. The issue is not only license compliance. It is whether community labor becomes platform capacity without recognition.

The crisis also reaches education. Academic integrity policies often ask whether students wrote their own work, whether they cited sources, and whether unauthorized assistance was used. LLMs complicate all three questions. A student may submit text that no human source directly wrote. The text may include citations. The assistance may be disclosed. Yet the intellectual structure may be generated from opaque recombination. The policy question “did the student plagiarize?” becomes insufficient. A stronger question is required: did the submitted work preserve accountable relations among sources, student contribution, and generated assistance? Without that relation, integrity becomes procedural rather than epistemic.

The educational problem is not only misconduct. It is formation. If students learn to treat generated text as original whenever it passes similarity checks, they internalize a reduced model of authorship. They learn that originality means lexical uniqueness rather than accountable contribution. This weakens the intellectual habits that citation practices are meant to cultivate: reading, comparison, source evaluation, argumentative positioning, and acknowledgment of debt. Structural appropriation therefore threatens not only the detection of plagiarism but the pedagogy of authorship.

The same institutional weakness appears in AI disclosure policies. Many policies require authors to state whether AI tools were used. This requirement is necessary, but incomplete. Tool disclosure identifies the instrument of production. It does not disclose the sources or source clusters that shaped the generated output. A statement such as "AI was used for drafting" tells the reader that a model participated in the process. It does not tell the reader what intellectual labor the model absorbed, whether generated claims were independently verified, whether citations were post-hoc, or whether the argument structure was human-designed or model-supplied. Disclosure without provenance remains shallow.

A stronger institutional framework would distinguish between four questions. First, was AI used? Second, what parts of the text were AI-generated or AI-assisted? Third, what sources did the human author actually consult and verify? Fourth, what provenance information does the system provide about the generated content? Current policies usually address the first question and sometimes the second. Academic legitimacy under structural appropriation requires the third and fourth. Without them, institutions can document tool use without addressing intellectual debt.

This is why structural appropriation must be treated as a governance problem. It cannot be solved only by punishing individual users. Individual misuse exists, but the deeper issue is systemic. LLMs are designed, trained, distributed, and normalized in ways that make source opacity ordinary. If the infrastructure prevents provenance recovery, then integrity policies aimed only at users will be structurally unfair. They will demand accountability from the final human actor while ignoring the system that made accountability difficult. Governance must therefore operate at the level of model design, dataset documentation, interface disclosure, and institutional review.

The problem also intersects with the theory of authority without source developed in *Ethos Without Source* (Startari, 2025c). Generated text can appear credible because it reproduces the linguistic signs of expertise, even when its source conditions are unavailable. In academic and legal domains, that appearance is dangerous because institutions are trained to recognize formal markers of competence: structured sections, precise terminology, balanced tone, citations, and method-like language. LLMs can produce those markers

without preserving the underlying source chain. The system therefore simulates the form of legitimacy while weakening its evidentiary base.

The same logic follows from *Indexical Collapse*, where reference disappears while authority remains (Startari, 2025d). In the present context, attribution disappears while originality remains institutionally legible. A generated text can be accepted as original because it points to no obvious source. But that failure to point is precisely the problem. The absence of an indexical link to prior labor is not proof of independence. It is evidence of opacity. Academic and legal systems that treat missing source links as absence of debt will misclassify structural appropriation as originality.

The crisis of legitimacy also affects the status of the author. In predictive authorship, the human user may be the named author, but not the sole origin of the text's structure. The model may be the generator, but not an accountable author. The corpus may be the source of competence, but not visibly credited. This triangle destabilizes ordinary attribution. If the named author claims full originality, the claim may be false. If the model is credited as author, the attribution may be conceptually empty because the model does not bear responsibility. If the corpus is credited, the attribution may be impossible without provenance mechanisms. Current systems have no adequate category for this layered authorship.

This layered authorship creates problems for liability. If generated legal advice causes harm, if generated academic claims misrepresent prior work, if generated journalism distorts source reporting, or if generated code violates licensing norms, responsibility cannot be assigned solely by asking who typed the prompt. The prompt initiates output, but it does not explain the training history, interface design, or source opacity that shaped the output. A legitimate accountability framework must account for the entire production chain. Otherwise, institutions will punish visible users while leaving invisible extraction untouched.

The inadequacy of existing frameworks can be summarized in three failures. The first is the similarity failure: institutions over-rely on textual overlap as evidence of plagiarism. The second is the disclosure failure: institutions treat AI-use declarations as sufficient

accountability. The third is the citation failure: institutions accept post-hoc references as if they reveal generative provenance. These failures are connected. They all mistake visible surface markers for source accountability. Structural appropriation operates precisely because the surface can be made compliant while the dependency remains hidden.

The legal concept of market harm also requires expansion in this context. Traditional analysis may ask whether generated output substitutes for a particular copyrighted work. Structural appropriation suggests a broader harm: generative systems may substitute for the labor markets and recognition systems of entire communities whose writing made the model useful. A model trained on journalism may compete with journalists. A model trained on code may compete with programmers. A model trained on academic writing may assist in producing academic-looking text that bypasses scholarly labor. The harm is not always document-to-document substitution. It can be field-level displacement.

This field-level displacement is difficult to litigate because the harmed contributors may be numerous, dispersed, and individually hard to identify. That difficulty should not be mistaken for absence of harm. It is a structural feature of the appropriation. The same scale that makes the model powerful makes the debt hard to assign. Large-scale extraction converts many small contributions into concentrated platform value. The inability to itemize each contribution then becomes a defense against accountability. This is the political economy of attribution collapse.

A credible institutional response must therefore require generative provenance systems. Such systems would not need to produce perfect source reconstruction. They would need to provide enough information to distinguish between unsupported generation, retrieved citation, probable source influence, and verified human sourcing. In academic publishing, this could mean requiring authors to maintain source verification logs for AI-assisted sections. In legal practice, it could mean requiring attorneys to verify all generated authorities and disclose machine-generated drafting where relevant to professional duties. In journalism, it could mean prohibiting generated sourcing unless original reporting or verified references are attached. In software, it could mean license-risk indicators and provenance warnings for generated code.

The aim is not to abolish AI-assisted writing. The aim is to prevent synthetic originality from replacing accountable authorship. LLMs can support drafting, comparison, translation, summarization, and exploratory analysis. But in high-stakes knowledge production, usefulness cannot substitute for provenance. A system that produces credible language without source accountability creates legitimacy without traceability. That is precisely the condition that structural appropriation names.

The crisis of academic and legal legitimacy is therefore not a future risk. It is already implicit in the architecture of predictive generation. Any institution that evaluates writing through surface originality alone is exposed. Any policy that treats disclosure as sufficient is incomplete. Any legal framework that treats non-infringement as full legitimacy misses the epistemic issue. The problem is structural: generated text can be formally acceptable, legally uncertain, academically plausible, and still built from unacknowledged intellectual labor.

Part VII. Toward Generative Provenance Systems

Generative provenance systems are the necessary corrective to structural appropriation. If large language models generate text from absorbed human intellectual labor while obscuring source relations, then governance cannot remain limited to plagiarism detection, AI-use disclosure, or post-hoc citation. These mechanisms address only the surface of the final document. They do not reconstruct the conditions under which the document became possible. A provenance system must therefore operate at the level of source dependency, model mediation, and output traceability. Its purpose is not to eliminate generative writing, but to make the intellectual debt of generative writing auditable.

The central premise is simple: generated text must not be treated as source-neutral. Every output emerges from a chain of prior conditions: corpus selection, dataset preparation, training, model architecture, interface design, prompt formulation, user selection, editing, and publication. Current systems expose only the final stages of that chain. The user sees the prompt and the output. The reader sees the final text. The institution may see a

disclosure statement. But the deeper chain remains hidden. Generative provenance systems would make that chain visible enough for academic, legal, and ethical evaluation.

This requires a shift from output disclosure to dependency disclosure. Output disclosure states that AI was used. Dependency disclosure asks what kinds of prior intellectual labor made the output possible. The difference is decisive. A statement such as "this section was generated with AI assistance" identifies the instrument, but it does not identify the source structures absorbed into the text. It does not distinguish between retrieved information, model-generated synthesis, post-hoc citation, stylistic imitation, or structural recombination. Without dependency disclosure, AI transparency remains shallow. It identifies the tool while concealing the debt.

A generative provenance system should therefore distinguish at least four layers of traceability. The first layer is corpus-level provenance: what categories of texts, licenses, domains, languages, and institutions contributed to the model's training environment. The second layer is model-level provenance: how the system handles memorization risk, style transfer, domain concentration, and source reconstruction. The third layer is output-level provenance: what probable source clusters, retrieved documents, or training-domain influences shaped a given response. The fourth layer is user-level provenance: what prompts, edits, source checks, and verification steps were performed before publication. These layers do not produce perfect attribution. They produce accountable approximation.

Accountable approximation is the correct standard because scholarly attribution itself has never been total. Citations do not map every intellectual influence behind a text. They provide a minimum public structure through which claims can be evaluated, sources can be verified, and intellectual debt can be inspected. Generative provenance would serve the same function under predictive conditions. It would not need to reconstruct every parameter or every training document that contributed to an output. It would need to prevent the complete disappearance of provenance. The objective is not exhaustive genealogy. It is traceability sufficient for accountability.

This distinction follows directly from the problem developed in *Citation by Completion*, where predictive systems were shown to redistribute academic credit through citation

suggestions, concentration effects, and authority-bearing syntax (Startari, 2025a). That earlier analysis demonstrated that visibility can be redistributed at the point of writing. The present framework extends the problem to source dependency itself. If citation suggestions can distort visible credit, then opaque generation can erase deeper forms of intellectual debt before citation appears. The uploaded draft of *Citation by Completion* confirms this continuity, since it defines predictive writing systems as mechanisms that shape academic citations, source diversity, novelty, and legitimacy phrasing through autocomplete and authority syntax.

The first component of a generative provenance system is corpus disclosure. Model providers should disclose the broad composition of training data in a form that is usable for institutional evaluation. This does not require exposing every proprietary detail of model development, but it does require meaningful categories: academic publications, books, journalism, code repositories, legal documents, user content, licensed datasets, public-domain materials, synthetic data, and excluded sources. Without this information, users and institutions cannot determine whether a model's outputs may be structurally dependent on particular domains of intellectual labor. Dataset opacity becomes authorship opacity.

Corpus disclosure must also include licensing status and exclusion logic. A model trained on licensed academic databases raises different provenance questions than a model trained on scraped public web data. A model trained on code under permissive licenses raises different risks than one trained on unknown repositories. A model trained on legal, medical, or institutional documents raises specific responsibility concerns because those domains depend on precision, source validity, and traceability. Generative provenance begins by identifying what kind of intellectual world the model has absorbed. Without that baseline, output-level accountability is impossible.

The second component is model-level auditing. Providers should document whether the system has been tested for memorization, near-reproduction, style imitation, citation hallucination, and domain-specific concentration. Memorization risk matters because direct reproduction remains one form of plagiarism. Style imitation matters because structural appropriation often operates through form rather than exact copying. Citation

hallucination matters because post-hoc references can create false accountability. Domain concentration matters because models may reproduce the authority structures of dominant sources while absorbing less visible sources without recognition. These tests should be reported as part of model documentation.

The third component is output-level provenance. When a model generates text, the system should distinguish between at least three types of source relation: retrieved source, support source, and probable generative influence. A retrieved source is a document actually accessed during generation through retrieval or browsing. A support source is a reference added because it corroborates the claim. A probable generative influence is a source cluster or domain pattern likely to have shaped the output through training, style, or conceptual proximity. Current systems often confuse these categories. A bibliography may contain support sources, but support is not provenance. Generative provenance requires that the distinction be explicit.

This distinction is central to avoiding post-hoc citation laundering. A generated paragraph can be followed by relevant citations without those citations being the true sources of the paragraph's structure. The cited works may support the claim, but they may not explain how the generated formulation emerged. Under structural appropriation, that gap is where intellectual debt disappears. Output-level provenance should therefore indicate whether a citation was retrieved before generation, supplied after generation, suggested by the model, selected by the user, or independently verified. Without such markers, the reader cannot distinguish scholarship from retrospective stabilization.

The fourth component is user-level process documentation. Authors using generative systems in academic, legal, journalistic, or technical settings should maintain records of prompts, generated drafts, human edits, consulted sources, rejected outputs, and verification steps. This requirement does not criminalize AI assistance. It brings it into ordinary standards of accountable authorship. Scholars already keep notes, drafts, bibliographies, and methodological records. Lawyers verify authorities. Journalists retain source materials. Programmers track commits. AI-assisted writing requires an equivalent audit trail because the system can produce substantive intellectual content without exposing its own dependencies.

A minimal user provenance log should include five elements. First, the task given to the model. Second, the sections or claims generated by the model. Third, the sources independently consulted by the human author. Fourth, the revisions made by the author. Fifth, the final verification status of claims, quotations, citations, and domain-specific assertions. This log would not need to be published in full in every case. It should be available for review in high-stakes contexts. The point is to preserve a chain between generated material and human accountability.

The fifth component is provenance scoring. Generated outputs could include a traceability score indicating how much of the response is grounded in retrieved sources, how much is unsupported generation, how much is based on user-provided material, and how much relies on opaque model synthesis. Such a score would not determine truth. It would indicate provenance risk. A text generated entirely from retrieved documents with visible citations has a different risk profile from a text generated from the model's latent knowledge without source exposure. Institutions need that distinction. Without it, all generated text appears equally transparent when it is not.

A provenance score could include four indicators: source visibility, retrieval dependency, citation verification, and structural opacity. Source visibility measures whether sources are available to the user. Retrieval dependency measures whether the model actually used external documents during generation. Citation verification measures whether cited sources were checked against claims. Structural opacity measures how much of the output depends on untraceable model synthesis. These indicators would allow users, editors, and reviewers to evaluate whether a text is suitable for academic or professional use.

The sixth component is distinction between generative and editorial assistance. Not every AI use carries the same attribution burden. A grammar correction tool, a translation assistant, a summarizer, and a full generative drafting system do not occupy the same role. The more a system supplies substantive claims, argument structure, conceptual taxonomy, or source framing, the higher the provenance requirement should be. Institutions should therefore classify AI assistance by contribution level. Surface editing requires disclosure in some contexts, but substantive generation requires provenance. The governance burden should track the intellectual role of the system.

This classification would prevent two errors. The first error is treating all AI assistance as equally problematic. That position is too broad. The second error is treating all AI assistance as ordinary tooling. That position is too weak. A model that corrects punctuation does not create the same authorship problem as a model that drafts a theoretical argument. A model that summarizes user-provided documents does not create the same provenance problem as a model that generates claims from opaque training memory. Generative provenance systems must be calibrated to the function performed.

The seventh component is institutional review. Journals, universities, law firms, newsrooms, and technical organizations should adopt provenance standards proportional to risk. In low-stakes drafting, disclosure and human review may be sufficient. In academic publication, legal analysis, medical communication, public policy, journalism, and software licensing, stronger requirements are necessary. These may include source verification logs, AI-generated section marking, mandatory human review, prohibition of unsupported generated citations, and documentation of retrieval sources. High-stakes language requires high-traceability production.

The academic standard should be especially strict. A submitted paper should not be considered legitimate merely because it discloses AI use and passes plagiarism software. It should show that its claims are source-grounded, its citations are verified, its original contribution is humanly controlled, and its generated sections do not substitute synthetic originality for accountable argument. This does not prohibit AI-assisted scholarship. It requires that scholarship remain scholarship. The author must remain responsible not only for the final wording, but for the intellectual lineage and verification structure of the text.

The legal standard should focus on professional responsibility. Lawyers, judges, clerks, and legal researchers using LLMs must distinguish generated legal language from verified legal authority. A model-generated argument cannot be treated as legally grounded unless the authorities are independently checked. Generated legal prose may reproduce doctrinal style without reliable source lineage. In legal contexts, referential opacity can cause direct harm because legal validity depends on traceable authority. A provenance system should therefore mark whether cases, statutes, regulations, and doctrinal claims were retrieved, verified, or generated without source grounding.

The journalistic standard should focus on sourcing. A generated article or background summary should not convert prior reporting into unsourced synthesis. If the model generates factual or analytical material, the newsroom must verify whether the claims derive from original reporting, cited sources, wire material, public records, or opaque model synthesis. The difference matters because journalism depends on accountability to sources. Synthetic originality can erode that accountability by making accumulated reporting appear as generic knowledge. A provenance requirement would force generated journalism to preserve source visibility.

The software standard should focus on license and community attribution. Code generation systems should provide license-risk indicators, similarity warnings, and source-cluster signals when outputs resemble known repositories or common implementation patterns. Exact copying is not the only concern. Structural reuse of open-source labor can still raise attribution and governance issues. A provenance system cannot solve all licensing questions, but it can prevent the false assumption that generated code is automatically source-neutral. Generated code should be treated as unverified until provenance and license risk are assessed.

The theoretical basis for these standards is already implicit in *Borrowed Voices, Shared Debt*, where plagiarism and idea recombination were framed as problems of knowledge commons rather than isolated acts of copying (Startari, 2025b). Generative provenance systems extend that framework into infrastructure. If knowledge is recombined at scale, then debt must also be tracked at scale. The problem is not only that individual users may fail to cite. The problem is that the system itself lacks mechanisms for making recombinative debt visible. Provenance is therefore the institutional form of shared debt management.

The same basis appears in *Ethos Without Source*, where generated credibility was shown to operate without stable origin (Startari, 2025c). Provenance systems respond to that condition by refusing to let credibility stand alone. A generated text that sounds authoritative must show how its authority is grounded. If it cannot, then its authority should be downgraded. This is a necessary shift in evaluation. Fluency, tone, structure, and

confidence cannot serve as substitutes for source visibility. Provenance becomes the test that separates credible writing from merely credible-sounding writing.

The relation to *Indexical Collapse* is equally direct. If predictive systems allow authority to persist after reference disappears, then generative provenance must restore indexical pressure (Startari, 2025d). A generated claim should be forced to point somewhere: to a retrieved source, a user-provided document, a verified citation, a known dataset category, or a declared zone of unsupported synthesis. The system should not be allowed to speak as if origin were irrelevant. Provenance reintroduces the question that predictive generation suppresses: where does this come from?

This does not mean that every generated sentence must have a single source. That would reproduce the classical error this paper rejects. LLM outputs often arise from distributed patterns rather than identifiable source passages. The correct standard is not one sentence, one source. The correct standard is source relation classification. Some outputs may have direct source support. Some may have domain-level influence. Some may be generic linguistic construction. Some may be unsupported synthesis. Some may carry high structural similarity to known works. A mature provenance system would classify these relations rather than pretending that all attribution must take the form of a conventional citation.

Such classification would also reduce false accusations. Without provenance tools, institutions may over-punish visible AI use while missing deeper structural appropriation. A student who uses a model transparently and verifies sources may be treated more harshly than a professional who uses opaque generated text without disclosure. Provenance systems would shift evaluation from moral suspicion to evidentiary structure. The question would not be “was AI used?” but “was the generated material traceable, verified, and properly integrated?” This is a more precise and fair standard.

A generative provenance framework should include a basic taxonomy of output states. First, verified retrieval output: the model generated from identified and accessible sources. Second, user-grounded output: the model generated from material supplied by the user. Third, latent synthesis output: the model generated from internal statistical capacity

without source disclosure. Fourth, post-hoc supported output: citations were added after generation but do not establish generative lineage. Fifth, high-risk structural output: the text displays conceptual, stylistic, or argumentative proximity to identifiable works without attribution. This taxonomy would allow institutions to evaluate risk more accurately.

The category of latent synthesis output is especially important. Much LLM generation falls into this category. The model produces plausible text from its internal representations, but no source chain is provided. Such output should not be treated as automatically illegitimate. It should be treated as provenance-limited. Its use may be acceptable for brainstorming, drafting, or informal explanation. It becomes problematic when presented as scholarly, legal, journalistic, or technical authority without independent verification. The standard is contextual: the higher the stakes, the lower the tolerance for latent synthesis without provenance.

The category of post-hoc supported output is also critical. Many AI-assisted texts appear properly referenced because sources are added after generation. But a post-hoc source does not necessarily reveal where the argument came from. Institutions should require authors to distinguish between sources that shaped the writing process and sources found afterward to support generated claims. This distinction already exists implicitly in good scholarship. A source used to develop an argument is not the same as a source found later to decorate it. Generative systems make this distinction unavoidable.

The category of high-risk structural output addresses the central thesis of this paper. Structural appropriation often appears not as copied wording, but as borrowed architecture. A provenance system should therefore include tools for detecting structural proximity: similar argument sequence, taxonomy, conceptual framing, rhetorical pattern, or domain-specific arrangement. These tools will be imperfect, but their absence leaves institutions blind to the most important form of generative plagiarism. Similarity detection must expand from lexical overlap to structural dependency.

This expansion would require new metrics. Possible indicators include argument-sequence similarity, concept-cluster overlap, stylistic proximity, citation-pattern convergence, authority-syntax density, and domain-template reuse. These metrics would not prove

plagiarism by themselves. They would identify risk zones for review. The aim is not automatic condemnation. It is structured audit. A text flagged for structural proximity would require human evaluation, source comparison, and process documentation. This is more defensible than relying on opaque AI detectors or surface similarity alone.

The framework also requires separating plagiarism detection from provenance auditing. Plagiarism detection asks whether a text improperly copies or borrows from a source. Provenance auditing asks whether the production chain of the text is sufficiently traceable. A text may fail provenance auditing even if plagiarism is not proven. Conversely, a text may contain properly attributed borrowing and pass provenance review. The two systems should interact, but they should not be collapsed. Structural appropriation demands the broader category.

The role of model providers is unavoidable. Users cannot provide provenance that the system itself withholds. If providers design systems that generate authoritative language without source visibility, then they participate in attribution collapse. A serious governance regime must therefore impose duties on providers: dataset documentation, output provenance tools, citation reliability markers, retrieval logs, memorization-risk disclosures, and user-facing warnings when an answer is generated without source grounding. Responsibility cannot be shifted entirely to the final user.

The role of users is also real. Users must not treat generated text as source-neutral. They must verify claims, inspect citations, document AI involvement, and avoid presenting latent synthesis as independent scholarship. In academic contexts, users should not allow models to supply the core contribution of a paper unless the intellectual lineage is independently controlled and cited. In legal contexts, users should not rely on generated authorities without verification. In journalism, users should not publish generated factual claims without source confirmation. Provenance requires both system design and user discipline.

The role of institutions is to set enforceable thresholds. A university can define what level of AI assistance is permissible in coursework, what documentation is required, and how generated text should be evaluated. A journal can require AI-use declarations, source

verification logs, and exclusion of unsupported generated citations. A court or law firm can require verification of all AI-assisted legal authorities. A newsroom can require that generated background material be linked to verified sources. These rules should not be symbolic. They should specify evidence.

A strong institutional rule would state that AI-generated or AI-assisted text used in scholarly, legal, journalistic, or technical publication must be accompanied by a provenance record sufficient to distinguish human-authored contribution, model-generated synthesis, retrieved source use, and independently verified citation. This rule is stronger than generic AI disclosure. It does not ask merely whether AI was used. It asks how the output was made accountable. That is the necessary standard under structural appropriation.

The framework also requires a change in how originality is evaluated. Originality should no longer be defined negatively as absence of duplication. It should be defined relationally as accountable transformation of prior work. A generated text can be unique and still lack originality if its source relations are opaque. A human-authored text can borrow heavily and still be original if it transforms sources transparently and contributes something identifiable. Generative provenance restores this relational standard. It prevents synthetic originality from being mistaken for intellectual contribution.

This change matters because LLMs can increase surface novelty while decreasing source accountability. As models improve, they may become better at avoiding verbatim reproduction. That improvement may reduce visible plagiarism while intensifying structural appropriation. More fluent models can produce cleaner, more original-looking texts from deeper recombination. Classical detection will therefore become less effective precisely as the systems become more capable. Provenance systems must be built before fluency outpaces accountability.

A further safeguard is provenance-aware citation. In AI-assisted writing, citations should be marked according to function. Some citations may be human-selected sources consulted before writing. Others may be model-suggested sources. Others may be retrieved documents used during generation. Others may be post-hoc supports added during

verification. These categories should not be hidden. A provenance-aware bibliography would allow readers and reviewers to see how sources entered the production process. This would reduce the risk that post-hoc citation disguises opaque generation.

Such a system would also improve scholarly honesty. Authors often use different kinds of sources in different ways: foundational theory, empirical support, methodological reference, background context, and critical contrast. AI-assisted writing adds another dimension: generative influence. A source may influence the generated structure without being consciously selected by the author. If that source cannot be identified, the author should not pretend that the visible bibliography fully captures the intellectual lineage. A statement of provenance limitation may be required. This would be more honest than false precision.

The concept of provenance limitation is important. There will be cases where source lineage cannot be fully reconstructed. Instead of ignoring that fact, authors and systems should disclose it. A statement such as "AI-assisted drafting was used; generated sections were independently revised and checked against the cited sources, but model-level training provenance is unavailable" would be more accurate than a generic disclosure. It separates what the author verified from what remains opaque. This does not solve attribution collapse, but it prevents concealment of the collapse.

The final aim of generative provenance is to restore epistemic friction. LLM interfaces are designed to reduce friction: immediate answers, fluent prose, rapid synthesis, low effort. But scholarship, law, journalism, and technical accountability require certain forms of friction. Sources must be checked. Claims must be traced. Arguments must be situated. Citations must be meaningful. Provenance systems reintroduce productive friction into generated text. They slow down the movement from output to publication enough to preserve accountability.

This friction should not be viewed as inefficiency. It is the cost of legitimacy. A system that eliminates the labor of attribution also eliminates part of the structure that makes knowledge trustworthy. Fast generation without provenance creates cheap fluency and expensive consequences. Generative provenance imposes a necessary cost on outputs that

claim academic, legal, journalistic, or technical authority. The higher the authority claimed, the stronger the provenance burden.

The proposed framework therefore resolves the central chain of the paper. Part I showed that classical plagiarism collapses under predictive generation because source, copy, and author no longer remain stable. Part II showed that corpus extraction converts human intellectual labor into invisible model capacity. Part III showed that recombinative plagiarism produces apparent novelty from absorbed structures. Part IV showed that referential opacity prevents source lineage from being reconstructed. Part V showed that synthetic originality converts opacity into the appearance of authorship. Part VI showed that academic and legal systems are not equipped to evaluate this condition. Part VII now proposes generative provenance as the minimum corrective architecture.

The conclusion is direct. Structural appropriation cannot be governed by similarity detection alone. It cannot be solved by AI disclosure alone. It cannot be repaired by post-hoc citation alone. It requires a provenance framework capable of tracking, classifying, and disclosing the relation between corpus, model, output, and user. Without such a framework, large language models will continue to convert human intellectual labor into synthetic originality while leaving attribution structurally unavailable.

Generative provenance systems do not abolish the risks of AI-assisted writing. They make those risks visible. They create conditions under which generated text can be evaluated not only for fluency, usefulness, or originality of surface, but for source accountability. In doing so, they restore the central principle that structural appropriation threatens: intellectual production must preserve a traceable relation to the labor that makes it possible. Under predictive conditions, provenance is not an optional supplement to authorship. It is the condition under which authorship can remain legitimate.

References

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 610–623. https://doi.org/10.1145/3442188.3445922

Merton, R. K. (1968). The Matthew effect in science. *Science, 159*(3810), 56–63. https://doi.org/10.1126/science.159.3810.56

Parasuraman, R., & Riley, V. (1997). Humans and automation: Use, misuse, disuse, and abuse. *Human Factors, 39*(2), 230–253. https://doi.org/10.1518/001872097778543886

Startari, A. V. (2025a). *Citation by Completion: LLM Writing Aids and the Redistribution of Academic Credits*. Zenodo. https://doi.org/10.5281/zenodo.17287506

Startari, A. V. (2025b). *Borrowed Voices, Shared Debt: Plagiarism, Idea Recombination, and the Knowledge Commons in Large Language Models*. SSRN. https://doi.org/10.2139/ssrn.5494528

Startari, A. V. (2025c). *Ethos Without Source: Algorithmic Identity and the Simulation of Credibility*. SSRN. https://doi.org/10.2139/ssrn.5313317

Startari, A. V. (2025d). *Indexical Collapse: Reference Disappears, Authority Remains in Predictive Systems*. SSRN. https://doi.org/10.2139/ssrn.5545240

Startari, A. V. (2025e). *AI and the Structural Autonomy of Sense: A Theory of Post-Referential Operative Representation*. SSRN. https://doi.org/10.2139/ssrn.5272361