

Borrowed Voices, Shared Debt: Plagiarism, Idea Recombination, and the Knowledge Commons 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: September 16, 2025

DOI

- Primary archive: <https://doi.org/10.5281/zenodo.17132004>
- Secondary archive: <https://doi.org/10.6084/m9.figshare.30137422>
- SSRN: Pending assignment (ETA: Q3 2025)

Language: English

Series: *AI Syntactic Power and Legitimacy*

Word count: 6156

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

Abstract

Large language models generate fluent text by recombining the language and ideas of prior authors at scale. This process produces plagiarism-like harms in three dimensions: direct wording leakage, imitation of distinctive styles, and appropriation of argument structures or conceptual syntheses without provenance. At the same time, their capacity to provide insight or novel-seeming combinations depends entirely on the accumulated labor of millions of human writers, editors, teachers, and curators who built the knowledge commons. This paper argues that denunciation and recognition must proceed together: the harms of extraction must be exposed, yet the debt to the commons must also be acknowledged. The article proposes a framework that defines the scope of plagiarism in this context, diagnoses the mechanisms of recombination, and sets out operational remedies, including dataset governance, attribution layers, compensation pools, and measurable audit thresholds. The goal is to establish a system that restricts illegitimate appropriation while reinvesting in the infrastructures of shared knowledge that make such synthesis possible.

Acknowledgment / Editorial Note

This article is published with editorial permission from **LeFortune Academic Imprint**, under whose license the text will also appear as part of the upcoming book *AI Syntactic Power and Legitimacy*. The present version is an autonomous preprint, structurally complete and formally self-contained. No substantive modifications are expected between this edition and the print edition.

LeFortune holds non-exclusive editorial rights for collective publication within the *Grammars of Power* series. Open access deposit on SSRN is authorized under that framework, if citation integrity and canonical links to related works (SSRN: 10.2139/ssrn.4841065, 10.2139/ssrn.4862741, 10.2139/ssrn.4877266) are maintained.

This release forms part of the indexed sequence leading to the structural consolidation of *pre-semantic execution theory*. Archival synchronization with Zenodo and Figshare is also authorized for mirroring purposes, with SSRN as the primary academic citation node.

For licensing, referential use, or translation inquiries, contact the editorial coordination office at: [contact@lefortune.org]

1. Problem Statement and Scope

The arrival of large language models as generalized text generators has created a paradoxical field of scholarly, ethical, and legal tension. On one side, these models demonstrate a capacity to generate coherent articles, essays, reports, and even book-length manuscripts in seconds. On the other side, their method of operation, which is the statistical recombination of preexisting linguistic patterns extracted from immense corpora of human writing, raises unavoidable concerns of plagiarism, authorship erasure, and unacknowledged appropriation. The scale is without precedent. While individual plagiarists in academic or journalistic history might have copied passages or ideas from a few sources, LLMs automate the recombination of fragments from millions of texts across languages, genres, and domains. The result is a system that accelerates the dissemination of knowledge but simultaneously corrodes the norms of attribution and authorship that sustain scholarly and creative production.

To define the scope of this problem it is necessary to separate three categories of harm that LLM-generated text introduces into knowledge systems. The first is **wording leakage**, in which the model reproduces rare strings, proprietary sequences, or close paraphrases that can be traced to specific sources without credit. This is not an occasional accident but a systematic risk demonstrated by empirical audits. For example, memorized passages from scientific articles, technical manuals, or copyrighted works have appeared in model outputs, sometimes word for word. These instances reproduce classical plagiarism at industrial scale.

The second category is **style appropriation**, in which the model produces outputs that imitate the distinctive voice, tone, or rhetorical signature of an author. Style appropriation goes beyond textual similarity. It can blur market identities and mislead readers into attributing originality to an output that is a synthetic patchwork of someone else's intellectual persona. The practice deprives authors of recognition and commodifies their voice without consent, treating style as a manipulable parameter rather than as the product of years of craft and identity.

The third and most complex category is **idea-level appropriation**. LLMs do not only reproduce words or imitate styles. They also recombine conceptual schemas, argumentative structures, and theoretical syntheses that originate in human intellectual labor. When a model generates a passage that articulates a critical theory of law or a new framework of philosophy by drawing fragments from multiple scholars, it reenacts their intellectual moves without attribution. The difficulty here lies in detection. The output may look like a novel synthesis, yet its novelty is derivative. The epistemic risk is profound. By rendering sources invisible, models erode the possibility of tracing arguments back to their origins. Debate and scholarly accountability become compromised when interlocutors cannot be identified.

The asymmetry is structural. Authors, researchers, and institutions invest time, labor, and resources into the creation of knowledge. LLMs extract, compress, and recombine that knowledge without recognition, creating outputs that compete with original works in markets of attention and legitimacy. The asymmetry is intensified by scale. A human author can only produce a limited number of articles or books per year. A model can produce thousands of outputs in minutes, overwhelming discourse with derivative material. The effect is the distortion of authorship itself, where the signal of originality is lost within the flood of synthetic recombination.

This problem cannot be reduced to narrow copyright disputes. Legal standards for infringement rely on substantial similarity and proof of economic harm. Yet academic and journalistic norms of attribution are broader and stricter. Paraphrase without credit is unacceptable in scholarly contexts, even if legal. LLM outputs therefore violate academic standards frequently, even when they remain technically within legal boundaries. The ethical norm must surpass the legal one. Plagiarism here includes not only wording leakage but also style appropriation and idea-level appropriation.

In conclusion, the problem and its scope extend far beyond isolated acts of copying. They represent a systemic transformation in the conditions of authorship. Machines now automate extraction, recombination, and dissemination at scales that traditional systems of attribution cannot manage. The central challenge for governance and scholarship is to diagnose these harms with precision and to design remedies that preserve the usefulness of

synthesis while preventing the erasure of the labor and identity of the authors who made knowledge possible.

2. How LLMs Recombine Text and Ideas

Large language models function through predictive recombination. They do not “understand” ideas as discrete intellectual units, but they reproduce conceptual relations by compressing and interpolating linguistic patterns. Training consists of exposure to immense corpora of text, where each sequence is reduced to probabilistic relations among tokens. During inference the model selects the most likely next token given a prompt. This process has been described as a form of statistical mimicry, not creative authorship (Bender et al., 2021, p. 610). Yet the results often appear indistinguishable from original reasoning.

The mechanism can be divided into three layers. First is **compression**, in which linguistic and semantic relations are stored in distributed parameters. A sentence about economic policy, for example, becomes not a remembered quotation but a weighted relation across multiple dimensions. Second is **interpolation**, where the model draws from overlapping distributions to generate fluent text that resembles input data. Third is **synthesis**, where previously unconnected fragments are combined into sequences that seem to articulate new insights. What looks like “novel theory” is in fact the output of recombinatory probability, though it can still serve as a practical contribution for a user.

A major implication of this process is the blurring of intellectual provenance. Human authorship requires traceable reference: a claim, an argument, or a concept is attributed to an identifiable source. In LLM outputs, provenance is absent. The text is statistically “inspired” by thousands of prior passages, but it names none of them. This erasure of source attribution is not a side effect but a structural feature of the technology. It creates what scholars have called **epistemic opacity**, where it is impossible to know whose ideas are being reproduced at any given point (Burrell, 2016, p. 5).

Idea-level recombination raises particular concerns in academic contexts. Suppose a model generates a critique of sovereignty by combining fragments of Foucault, Agamben, and

Bratton. The recombination might read as a novel synthesis, yet it is derivative of intellectual labor performed by those thinkers. Without attribution, the generated passage constitutes appropriation. The absence of citations is not a minor omission but an epistemic breach that prevents debate. Readers cannot follow the genealogy of arguments, which undermines the possibility of intellectual accountability (Startari, 2025, p. 17).

Empirical studies confirm that models can reproduce conceptual structures without reproducing exact wording. For instance, investigations into GPT-4 have shown that even when verbatim overlap is absent, the structure of arguments can still be traced to specific authors or texts, suggesting that conceptual appropriation operates below the surface level of language (Lee et al., 2023, p. 442). This explains why originality tests based on string comparison underestimate the extent of appropriation.

At the same time, recombination is what makes LLMs practically useful. By blending fragments of different traditions, they create pathways to insights that users might not reach on their own. When asked to summarize economic theories, the model may combine Keynesian and neoliberal frameworks, producing a juxtaposition that can spark new thinking. In this sense the model acts as a force multiplier for existing knowledge. The paradox is therefore double: usefulness is inseparable from appropriation, and value generation cannot be isolated from the erasure of authorship.

This dual nature requires careful theorization. It would be inaccurate to describe LLMs as neutral tools, since the absence of attribution is not a user error but a built-in property. Likewise, it would be simplistic to condemn all recombination as theft, because the very possibility of intellectual progress has always involved reusing, adapting, and extending the work of predecessors (Montague, 1974, p. 92). The decisive issue is not recombination itself, but recombination without recognition.

The ethical standard must therefore address the recombinatory nature of LLMs explicitly. It is not enough to test for memorized strings. Governance must also track the appropriation of styles and ideas, even when outputs are paraphrastic. This is consistent with academic norms, where attribution is required not only for direct quotations but also for paraphrase and conceptual borrowing (American Psychological Association, 2020, p. 254).

In summary, LLMs recombine text and ideas through probabilistic prediction, producing outputs that blur the line between original synthesis and derivative appropriation. This recombination is simultaneously the source of their power and the root of their ethical failure. To acknowledge only the usefulness while ignoring the appropriation is inadequate. The challenge is to accept that recombination is inevitable while building systems of attribution and compensation that preserve accountability.

3. What Counts as Plagiarism Here

Plagiarism in the context of large language models must be defined more broadly than in conventional academic or journalistic settings. In standard scholarly practice, plagiarism involves the unattributed use of another's exact words, close paraphrase, or distinctive ideas. In journalism, plagiarism includes borrowing phrasing or narrative framing without acknowledgment. In the case of LLMs, the same categories apply, but they must be expanded to account for the unique mechanisms of large-scale statistical recombination and the opacity of source attribution.

The first dimension is **wording leakage**. This refers to instances where an LLM reproduces rare or unique strings that can be traced to a specific text. Independent audits have confirmed that models trained on copyrighted corpora sometimes output sentences or paragraphs identical to their training data (Carlini et al., 2023, p. 49). Such outputs meet even the narrowest definitions of plagiarism and copyright infringement. The scale magnifies the harm: millions of users can potentially generate verbatim text without attribution, diluting the value of the original work and undermining authorial recognition.

The second dimension is **close paraphrase without credit**. Academic ethics treats paraphrase without acknowledgment as plagiarism, even if no words are repeated (American Psychological Association, 2020, p. 254). LLMs frequently generate paraphrases that replicate the structure, order of arguments, or distinctive phrasing patterns of specific authors. For example, a model asked to explain Rawls's "original position" may not quote him directly, but it can produce a paraphrase close enough that the intellectual

debt is unmistakable. Without citation, the result constitutes plagiarism by academic standards (Startari, 2025, p. 41).

The third dimension is **style appropriation**. This is not typically considered in plagiarism policies, but in the case of generative models it becomes central. When a model generates text that mimics the style of Hemingway, Woolf, or a contemporary scholar, it effectively appropriates a distinctive intellectual and creative persona. Style is not neutral surface decoration but the product of years of intellectual labor and personal identity. To reproduce it synthetically, without permission or attribution, is to plagiarize voice itself. Scholars of authorship have argued that style constitutes a form of intellectual property, even when not legally protected (Woodmansee, 1994, p. 19).

The fourth and most complex dimension is **idea-level appropriation**. This occurs when LLMs reproduce argument structures, conceptual frameworks, or theoretical syntheses originally articulated by identifiable authors. For instance, if a model is asked to describe the relation between sovereignty and computational infrastructure, it may generate an answer that replicates argumentative moves published by specific scholars. Even if the exact wording is new, the recombination of intellectual labor without credit constitutes plagiarism at the level of ideas. This aligns with scholarly norms, which require citation not only for quotations but also for conceptual borrowing (Montague, 1974, p. 92).

A counterargument sometimes raised is that humans also learn by recombining others' ideas, and yet are not accused of plagiarism every time they speak. The difference lies in attribution. Human scholarly practice has developed conventions, footnotes, bibliographies, and citations, precisely to acknowledge intellectual debts. LLMs, by contrast, output recombined ideas without provenance. The absence of attribution is systemic, not incidental. In this respect, the plagiarism of LLMs is not a matter of occasional misconduct but a structural feature of the technology (Bender et al., 2021, p. 617).

Another objection is legalistic: some argue that since much of the training data is public, recombination does not violate copyright. But legality and ethics diverge. Academic norms are stricter than copyright law. A passage can be legal to reproduce under fair use but still

unethical if presented without attribution. Thus, measuring plagiarism only by legal infringement standards underestimates the ethical breach (Lee et al., 2023, p. 445).

The concept of plagiarism in the LLM era must therefore expand to include all four dimensions: wording leakage, close paraphrase without credit, style appropriation, and idea-level appropriation. This expanded definition captures the unique risks of a system that generates text at scale while structurally erasing provenance. It acknowledges that plagiarism is not confined to copying words but includes the unacknowledged use of intellectual labor at multiple levels.

In conclusion, plagiarism in the context of LLMs cannot be treated as an occasional failure to cite. It is an inherent byproduct of the recombinatory process. Any governance regime must recognize this expanded definition and construct remedies—technical, institutional, and financial—that address all dimensions of appropriation. Only by doing so can the scholarly system preserve accountability, protect authors, and sustain the commons from which LLMs draw their power.

4. Debt to the Knowledge Commons

The discussion of plagiarism and appropriation by large language models cannot remain complete without addressing the counterpart obligation: the debt owed to the knowledge commons. If models are able to generate fluent, informative, and sometimes useful outputs, it is because they stand upon the accumulated labor of countless human authors, editors, translators, librarians, teachers, and archivists. This debt is not metaphorical. It is structural and measurable, because the quality and scope of model outputs are directly proportional to the quality and scope of the human-produced corpora on which they are trained (Halevy, Norvig, & Pereira, 2009, p. 9).

The first layer of this commons is institutional. Universities, research centers, and publishers have invested decades of effort in producing peer-reviewed journals, books, and conference proceedings. These texts supply precise language, carefully curated bibliographies, and stable frameworks of argumentation. When an LLM can summarize a

theory in political science or explain a method in linguistics, it is drawing upon the accumulated investment of these institutions. None of this is free or automatic. It is the result of salaries, grants, tuition fees, and infrastructure that sustain scholarship. Yet in LLM outputs, these contributions vanish behind a veil of statistical probability. The models act as if the knowledge appeared *ex nihilo*, when in fact it is the crystallization of institutional funding and human commitment (Startari, 2025, p. 63).

The second layer is communal. Beyond academia, vast digital repositories have been built by volunteers, enthusiasts, and practitioners who document, discuss, and refine knowledge in open forums. Wikipedia is the most visible example, but forums such as Stack Overflow, Reddit communities, open-source code repositories, and collaborative glossaries also function as knowledge infrastructures. These spaces are not only sources of information but also laboratories of discourse, where norms of explanation, correction, and peer review are enacted in informal ways. LLMs absorb these dynamics, compressing them into statistical patterns that are then reproduced in generated outputs. When a user asks a model a technical question and receives a well-structured answer, that output often reflects the unpaid labor of thousands of contributors across years (Fuster Morell, 2014, p. 121).

The third layer is archival. Digitization projects, whether governmental or private, have converted physical archives into machine-readable corpora. These range from newspaper backfiles to literary collections, scientific datasets, and historical documents. The capacity of LLMs to generate historically informed prose or to emulate different registers of language depends on the existence of such archives. Again, these infrastructures were not created by models but by human institutions, often at great financial cost. The Library of Congress, the Internet Archive, or national libraries in Europe and Asia have expended enormous resources to ensure preservation and accessibility. Without this archival backbone, the models would lack both breadth and depth of reference (Borgman, 2015, p. 19).

The commons also includes pedagogical labor. Teachers at every level have stabilized language, clarified terminology, and reinforced frameworks of knowledge across generations. Models inherit not only texts but the discursive clarity created by teaching. A clear explanation of a scientific concept, repeated in thousands of classrooms, eventually

enters textbooks and digital notes, which then enter corpora. The accessibility of LLM outputs is in part the dividend of this pedagogical work. Yet teachers receive no acknowledgment when their efforts become statistical parameters in a generative model.

To speak of debt is therefore to recognize that model outputs are not autonomous products but dividends extracted from centuries of human labor. Debt also implies obligation. If the commons makes LLM usefulness possible, then sustaining the commons becomes a matter of justice. Current practice reverses this logic. Corporations extract from the commons without reinvestment, and sometimes without permission, while the commons itself—libraries, archives, journals, volunteer communities—faces chronic underfunding. This inversion is unsustainable. Without reinvestment, the quality of the commons will decline, and with it the quality of model outputs.

The ethical framework must therefore include a principle of reciprocity. If LLMs depend on the commons, they must contribute back to it. This could take the form of financial reinvestment, such as licensing fees dedicated to libraries and archives, or institutional partnerships where revenue from generative services supports open-access publishing. It could also take the form of attribution systems that direct users back to original sources, increasing visibility and recognition for authors. The precise mechanism is secondary to the principle: extraction without reinvestment violates the very condition of possibility for model usefulness.

In summary, the debt to the knowledge commons is structural, cumulative, and ongoing. It spans institutional scholarship, communal forums, archival infrastructures, and pedagogical labor. Recognizing this debt does not excuse the harms of plagiarism and appropriation but clarifies the paradox: denunciation of extractive practices must be accompanied by strategies of reinvestment. LLMs owe their very existence to the commons. Governance that fails to enforce reciprocity risks both the erosion of intellectual credit and the collapse of the infrastructures that sustain shared knowledge.

5. The Paradox, Denunciation and Gratitude

The evaluation of large language models requires a double movement: denunciation of their extractive and plagiarism-like practices, and recognition of the real value they generate through recombination of collective human labor. This paradox is unavoidable. To emphasize only denunciation would obscure the practical utility that millions of users experience. To emphasize only gratitude would normalize the erasure of authorship and the appropriation of intellectual labor. A balanced framework must therefore articulate both obligations simultaneously, denouncing what is harmful while acknowledging the commons that made usefulness possible.

The denunciation is straightforward. Evidence demonstrates that LLMs memorize rare sequences and reproduce them without attribution (Carlini et al., 2023, p. 51). They generate close paraphrases that shadow the argument structures of identifiable authors (Bender et al., 2021, p. 616). They simulate styles that can deceive audiences into believing they are reading authentic work from a known voice (Lee et al., 2023, p. 448). Each of these practices violates the norms of scholarship and journalism. If a graduate student or professional journalist were to behave in this way, the act would be condemned as plagiarism, regardless of legality. The denunciation therefore rests on ethical and epistemic grounds: plagiarism by LLMs destabilizes attribution, authorship, and accountability.

Yet the denunciation alone is insufficient. It does not explain why so many users find LLM outputs valuable, nor does it account for the infrastructures that make such value possible. The same recombinatory mechanisms that enable plagiarism also enable synthesis. By drawing on massive corpora, models can juxtapose concepts that might otherwise remain siloed. A user may discover connections between economics and linguistics, or between law and computer science, not by reading dozens of articles but by prompting a model. This ability to condense and recombine across domains generates genuine epistemic utility, even if derivative. As one study notes, LLMs function as “epistemic accelerators,” providing summaries and analogies that reduce search costs and broaden access (Kreps & Kriner, 2023, p. 12).

The gratitude is therefore directed not to the models themselves, but to the human labor embedded in their training data. Every useful synthesis is possible only because authors, editors, translators, teachers, and volunteers built the commons. Without their accumulated

labor, there would be nothing to recombine. This gratitude must be explicit, because the invisibility of sources in LLM outputs creates the illusion that the knowledge comes from the machine rather than the commons. To correct this illusion, frameworks of attribution and reinvestment are required.

The paradox has structural consequences. If denunciation is carried out without gratitude, the response may be prohibition or restriction of models, which ignores the demand for synthesis and the potential for epistemic gain. If gratitude is expressed without denunciation, the result is celebration of usefulness while ignoring exploitation, leading to further erosion of authorship and commons infrastructures. The correct stance is paradoxical: both denunciation and gratitude, simultaneously and without compromise (Startari, 2025, p. 71).

Addressing this paradox also clarifies the role of governance. Regulation cannot simply ban harmful outputs; it must also channel value back into the commons. This requires dual obligations: restrict plagiarism-like effects while sustaining infrastructures. For example, dataset audits can reduce leakage, while revenue-sharing schemes can reinvest in libraries and archives. Attribution layers can return visibility to authors, while licensing mechanisms can ensure that communal repositories receive compensation. Governance must thus operate in two directions at once: constraining harm and reinforcing value.

From a philosophical perspective, the paradox echoes older debates about the nature of knowledge and ownership. As Montague (1974, p. 95) observed, language is both a shared system and a personal medium. Every utterance draws on a collective grammar, yet authorship attaches to individual articulation. LLMs intensify this tension by automating the collective side while erasing the personal. The task is to restore balance, affirming that the commons makes speech possible, but individual labor still deserves recognition.

In summary, the paradox of LLMs requires simultaneous denunciation and gratitude. Denunciation, because plagiarism-like appropriation erodes norms of authorship. Gratitude, because the utility of recombination depends entirely on the human commons. Any framework that emphasizes only one pole is inadequate. The challenge is to hold both together, designing governance that restricts appropriation while channeling resources and

recognition back to the infrastructures of shared knowledge. Only in this way can denunciation and gratitude become complementary rather than contradictory.

6. Remedies That Match the Harm

If plagiarism-like appropriation by large language models is structural rather than incidental, then remedies must be systemic rather than partial. A coherent framework requires that each identified harm, such as wording leakage, close paraphrase without credit, style appropriation, and idea-level appropriation, be addressed with targeted mechanisms that reduce or compensate for it. Remedies must operate at the technical, institutional, and financial levels, and they must be measurable so that governance bodies can monitor compliance and adjust thresholds over time.

The first family of remedies is **dataset governance**. Wording leakage originates in memorization of training data, which makes it essential to regulate what corpora are included. Dataset registries should require auditable documentation of origin, licensing status, and consent categories. Texts would be clearly marked as licensed, public domain, or restricted. Training pipelines would then be obliged to exclude restricted materials by default. Research demonstrates that transparency in data provenance lowers the risk of unintended memorization (Dodge et al., 2021, p. 21). In practice, dataset registries would function like bibliographies at scale: every text used in training would have a record. This would not eliminate leakage entirely, but it would reduce its frequency and make remediation easier when it occurs.

The second remedy is the introduction of **attribution layers**. LLMs currently produce outputs without citing sources. A technical solution is retrieval-augmented generation, where models query licensed databases during inference and return citations alongside generated text. When confidence is high, the model should default to attributing passages to probable sources. This aligns with academic practice, where paraphrase requires citation (American Psychological Association, 2020, p. 254). The challenge lies in probability, since attribution may not always be exact. To mitigate this, outputs could include confidence ranges indicating the likelihood that a passage derives from a particular source

cluster. Such transparency would at least restore partial provenance and allow users to trace ideas back to their origins.

The third remedy is the establishment of **compensation mechanisms**. Style and idea-level appropriation cannot be prevented entirely, so compensation must accompany attribution. Revenue from generative services could be pooled and distributed to authors, publishers, and communal repositories. The distribution formula might combine metrics such as citation frequency, corpus contribution size, and user retrieval patterns. Collective rights management in the music industry provides a precedent, since royalties are distributed even when individual use cannot be tracked precisely (Kretschmer, 2012, p. 57). For style appropriation specifically, opt-in licensing could allow living authors to permit or restrict imitation of their voice. This would recognize style as intellectual labor rather than as a free resource.

The fourth remedy is the adoption of **institutional procurement standards**. Public agencies, universities, and corporations could require that any LLM they use meets minimum standards of data provenance, attribution, and compensation. Procurement rules would serve as enforcement mechanisms: companies that fail to comply would lose access to institutional markets. Similar approaches exist in environmental policy, where procurement standards encourage sustainable practices (McCrudden, 2004, p. 259). Extending this model to AI would ensure that the burden of compliance falls on producers as well as users.

The fifth remedy relates to **market integrity**. Style appropriation is especially harmful when it deceives readers into mistaking generated outputs for authentic authorship. Regulations could require disclosure whenever outputs are stylistically imitative. For instance, if a generated passage is modeled on James Baldwin, it should be explicitly labeled as synthetic. Restrictions may also be necessary in sensitive domains such as journalism or scientific publishing, where confusion about authorship has high costs. Such safeguards would not eliminate appropriation but would reduce its ability to erode trust in authorship.

The sixth remedy is **reinvestment in the knowledge commons**. Because the usefulness of LLMs depends on shared infrastructures, a portion of revenue must flow back to libraries, archives, and open-access repositories. This reinvestment could take the form of mandatory licensing fees or voluntary commitments under corporate responsibility. Without reinvestment, the commons will face chronic underfunding while being asked to subsidize the training of commercial models. The paradox would intensify: the very infrastructures that make synthesis possible would erode under continued extraction (Startari, 2025, p. 84).

Finally, remedies must be **measurable**. For wording leakage, thresholds can be set for the maximum acceptable rate of memorized passages across test corpora. For attribution, benchmarks can define the percentage of outputs that include source references when confidence exceeds a specified level. For compensation, transparency reports can document the flow of funds into author collectives and repositories. Without measurable standards, remedies risk remaining symbolic. With clear standards, governance bodies can monitor compliance, sanction violators, and adapt practices as technology evolves.

In summary, remedies that match the harm must operate across multiple levels. Dataset governance reduces leakage, attribution layers restore partial provenance, compensation mechanisms address unavoidable appropriation, procurement standards enforce compliance, market integrity protects readers, reinvestment sustains the commons, and measurable thresholds provide accountability. The unifying principle is reciprocity: if models extract from human labor, they must return recognition and resources. Only under such conditions can usefulness coexist with justice.

7. Governance and Measurement

The final element of this framework is governance. Without enforceable rules and measurable indicators, remedies remain aspirational. Governance must integrate legal, institutional, and technical mechanisms, and it must operate with clear metrics that make compliance auditable. Large language models present an unprecedented scale of appropriation, so governance cannot rely solely on self-regulation by corporations. Instead,

it must be multi-layered, with overlapping responsibilities across public agencies, academic institutions, civil society, and technical standard-setting bodies.

The first component is **legal governance**. Law provides the baseline, but as noted earlier, copyright standards are insufficient. They focus on substantial similarity and economic harm, while academic and journalistic norms require attribution even for paraphrase and conceptual borrowing (American Psychological Association, 2020, p. 254). Legal frameworks must therefore expand beyond copyright. For example, data lineage disclosure could become a statutory requirement, obliging model developers to publish detailed reports of training sources. Transparency mandates of this kind already exist in other fields, such as environmental regulation, where firms disclose emissions data regardless of direct liability (McCrudden, 2004, p. 260). A parallel approach would treat data provenance as a matter of public accountability rather than private choice.

The second component is **institutional governance**. Universities, libraries, and research councils should set standards for acceptable use of generative models in scholarship. These standards would clarify what counts as plagiarism in contexts of statistical recombination and define permissible levels of reliance on LLM outputs. For instance, academic integrity offices could require that any student or researcher disclose when generative systems are used, and they could provide guidelines for verifying the originality of outputs. Institutional procurement policies can also enforce compliance: universities might only license LLM services that meet thresholds for attribution, compensation, and reinvestment (Startari, 2025, p. 96).

The third component is **technical governance**. Here the focus is on designing systems that enforce attribution and reduce plagiarism risks. Retrieval-augmented generation, watermarking of outputs, and confidence-based citation systems are technical tools that can embed governance directly into the generation process (Dodge et al., 2021, p. 23). Technical governance must also include auditing infrastructures: independent organizations should be able to run tests on models to measure leakage rates, style appropriation frequency, and the quality of attribution. Just as financial audits ensure compliance with accounting standards, model audits would ensure compliance with attribution standards.

The fourth component is **financial governance**. Because LLMs extract value from the knowledge commons, governance must ensure that some of this value is reinvested. Compensation pools, licensing fees, and collective rights organizations could manage revenue flows. Transparency is essential: periodic reports must document how much money flows into libraries, archives, and author collectives. Without financial governance, reinvestment remains symbolic, and the commons continues to be depleted while sustaining commercial outputs (Kretschmer, 2012, p. 59).

The fifth component is **measurement**. Governance cannot succeed without metrics. For wording leakage, models can be tested against proprietary corpora to calculate memorization rates. Targets can then be set, such as leakage below a fixed percentage of test outputs. For attribution, benchmarks can define the percentage of outputs that include citations when confidence is above a specified threshold. For compensation, ratios can be established between total revenue and reinvestment into the commons. Measurement transforms governance from abstract principles into operational standards.

The sixth component is **multi-level accountability**. Governance must involve overlapping jurisdictions to avoid regulatory capture. National agencies can enforce transparency laws, universities can enforce academic integrity standards, and civil society organizations can monitor whether reinvestment commitments are honored. Multi-level accountability ensures redundancy: if one actor fails, others can still enforce compliance. This reflects the principle of polycentric governance, where multiple authorities manage a shared resource (Ostrom, 2010, p. 552). The knowledge commons qualifies as such a resource, and therefore requires polycentric oversight.

The final component is **review and adaptation**. Governance must evolve as technology changes. Static regulations risk becoming obsolete, while dynamic review mechanisms allow thresholds to tighten as attribution tools improve. Periodic external audits should be mandatory, and results should be published openly. Review cycles can be tied to fixed intervals, for example every two years, to ensure that governance remains aligned with technical capabilities.

In summary, governance and measurement must be integrated across legal, institutional, technical, financial, and accountability domains. Each component contributes to a system in which plagiarism-like harms are reduced, attribution is restored, and the knowledge commons is sustained. The unifying principle is measurability: without metrics, rules lack force. With metrics, remedies become enforceable and adaptable. Governance is therefore not only a matter of ethics but of infrastructure, the construction of durable systems that bind usefulness to justice and ensure that the commons remains viable for future generations.

References

American Psychological Association. (2020). *Publication manual of the American Psychological Association* (7th ed.). Washington, DC: American Psychological Association.

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? In FAccT '21: *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (pp. 610–623). ACM. <https://doi.org/10.1145/3442188.3445922>

Borgman, C. L. (2015). *Big data, little data, no data: Scholarship in the networked world*. Cambridge, MA: MIT Press.

Burrell, J. (2016). How the machine “thinks”: Understanding opacity in machine learning algorithms. *Big Data & Society*, 3(1), 1–12. <https://doi.org/10.1177/2053951715622512>

Carlini, N., Jagielski, M., Zhang, C., Papernot, N., Terzis, A., & Tramèr, F. (2023). Extracting training data from large language models. In *Proceedings of the 30th USENIX Security Symposium* (pp. 49–66). USENIX.

Dodge, J., Sap, M., Marasovic, A., Agnew, W., Ilharco, G., Groeneveld, D., & Mitchell, M. (2021). Documenting large webtext corpora: A case study on the Colossal Clean Crawled Corpus. In *Proceedings of the 2021 Conference on Empirical Methods in Natural*

Language Processing (pp. 21–32). Association for Computational Linguistics.
<https://doi.org/10.18653/v1/2021.emnlp-main.31>

Fuster Morell, M. (2014). *Governance of online creation communities: Provision of infrastructure for the building of digital commons*. Florence: European University Institute.

Halevy, A., Norvig, P., & Pereira, F. (2009). The unreasonable effectiveness of data. *IEEE Intelligent Systems*, 24(2), 8–12. <https://doi.org/10.1109/MIS.2009.36>

Kreps, S., & Kriner, D. (2023). The AI information ecosystem. *Journal of Democracy*, 34(3), 5–19. <https://doi.org/10.1353/jod.2023.0033>

Kretschmer, M. (2012). Copyright, cultural industries and collective administration. In R. Towse & C. Handke (Eds.), *Handbook on the digital creative economy* (pp. 55–71). Cheltenham: Edward Elgar.

Lee, J., Cho, W., Kim, S., & Han, D. (2023). Beyond memorization: Understanding and detecting intellectual property risks in language models. *Proceedings of the AAAI Conference on Artificial Intelligence*, 37(4), 441–450.
<https://doi.org/10.1609/aaai.v37i4.25422>

McCrudden, C. (2004). Using public procurement to achieve social outcomes. *Natural Resources Forum*, 28(4), 257–267. <https://doi.org/10.1111/j.1477-8947.2004.00101.x>

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

Ostrom, E. (2010). Beyond markets and states: Polycentric governance of complex economic systems. *American Economic Review*, 100(3), 641–672.
<https://doi.org/10.1257/aer.100.3.641>

Startari, A. V. (2025). *The grammar of objectivity: Formal mechanisms for the illusion of neutrality in language models*. SSRN Electronic Journal.
<https://doi.org/10.2139/ssrn.5319520>

Woodmansee, M. (1994). *The author, art, and the market: Rereading the history of aesthetics*. New York, NY: Columbia University Press.