From Uncontrolled Artificial Generation to an Accountable Research Partnership: Methodological Governance of LLMs in Academic Work

Moses Boudourides

School of Professional Studies, Northwestern University
moses.boudourides@northwestern.edu

Abstract

Large Language Models (LLMs) are increasingly integrated into academic workflows, presenting both opportunities and significant challenges to scholarly integrity. While issues of factual accuracy are widely discussed, a more subtle and dangerous failure mode is structural hallucination, where LLM-generated content, despite being factually correct at a sentence level, distorts the relational structure of knowledge, misrepresents bibliographic landscapes, and undermines intellectual attribution. This paper argues for a shift from treating LLMs as uncontrolled artificial generators to incorporating them into an accountable research partnership. We propose a researcher- and instructor-centric governance framework based on a suite of lightweight, individually implementable computational protocols. By combining knowledge graph extraction with social network and bibliometric analysis, our methodology provides a transparent, quantitative toolkit for validating the structural and bibliographic integrity of LLM-assisted work. We demonstrate how network diagnostics, including centrality analysis and modularity, can serve as a "hallucination stress test" to detect conceptual distortions. We further detail protocols for citation integrity and bibliometric benchmarking to ground LLM outputs in real scholarly ecosystems. The paper outlines direct applications of this framework in high-stakes academic practices, including manuscript writing, syllabus design, and student assessment. Ultimately, we argue that methodological governance, rather than outright prohibition, is the most effective path toward a responsible and productive human-LLM collaboration in academia.

1 Introduction

Large Language Models (LLMs) are rapidly becoming everyday collaborators in academic life, assisting in tasks that range from drafting manuscripts and summarizing literature to designing syllabi and supporting student supervision. Their widespread adoption raises not only technical questions but also fundamental epistemic and ethical ones—chief among them, the responsibility to preserve factual accuracy, scholarly reliability, and the fair attribution of intellectual work. The discourse surrounding LLMs in academia has often centered on the prevention of plagiarism and the verification of factual claims. However, this focus on local accuracy, while necessary, is insufficient to address the deeper challenges that LLMs pose to the integrity of scholarly communication.

The present paper argues that the most significant risks associated with LLM use in academia are not isolated factual errors, but rather systemic, structural distortions of knowledge. We introduce the concept of *structural hallucination*, a failure mode in which LLM-generated content, despite being factually correct at a sentence level, fundamentally misrepresents the conceptual and bibliographic landscape of a field [Tang et al., 2025, Pan et al., 2024, Liu et al., 2025, Mustaffa et al., 2025]. Such outputs may elevate peripheral concepts to central status, fabricate or misattribute citations, and construct logically flawed arguments that appear coherent on the surface. These distortions evade simple fact-checking but have profound consequences for the reliability of research and the validity of educational assessment.

To address the underlying challenge, we propose a paradigm shift: from treating LLMs as *uncontrolled artificial generators* to integrating them into an *accountable research partnership*. This requires moving beyond a reactive

stance of error correction to a proactive one of methodological governance. We present a practical, hands-on framework designed for researchers, instructors, and graduate students—not AI specialists—to control and validate LLM outputs within their own workflows. Such an approach is grounded in a suite of simple, individually implementable computational protocols that leverage established techniques from network science and bibliometrics [Leon et al., 2017, Domenichini et al., 2025].

The core of the framework is a multi-layered validation pipeline that translates text—whether produced by an LLM, written by a human with LLM assistance, or drawn from trusted scholarly sources—into structured representations that can be quantitatively compared and evaluated. By combining knowledge graph extraction, citation and reference parsing, and social network analysis, our methodology provides a transparent and inspectable toolkit for assessing the structural coherence and bibliographic grounding of LLM-assisted work [Wen et al., 2024, Zhang et al., 2024, Liu et al., 2024, Stocker et al., 2024]. This paper details these protocols and demonstrates their application in detecting structural hallucinations, validating citation integrity, and ensuring thematic coherence.

It is important to clarify the nature of our contribution. The framework proposed in the current work is not an automated evaluation system, a benchmark, or a computational pipeline intended to replace human judgment. Rather, it is a set of methodological instruments designed to support and discipline human epistemic responsibility when working with LLM-assisted content. The emphasis is therefore not on performance optimization, but on governance: enabling researchers and instructors to retain control over the structural, bibliographic, and attributional properties of the knowledge claims they accept.

We structure our argument across six main sections, mirroring the pedagogical flow of a seminar designed to equip academics with these governance skills. Section 1 establishes the epistemic and ethical foundations of the framework, arguing that factual accuracy is a necessary but insufficient condition for responsible LLM use. Section 2 delves into the problem of structural hallucination, providing a theoretical and practical understanding of a critical failure mode. Section 3 presents the core methodology of the framework: the translation of text into knowledge graphs for evaluation. Section 4 details the use of network diagnostics, including centrality and modularity analysis, as a "hallucination stress test." Section 5 focuses on protocols for ensuring citation integrity and grounding LLM outputs in real scholarly ecosystems through bibliometric benchmarking. Section 6 explores the practical application of the discussed governance framework in research, teaching, and assessment, with a focus on the fair attribution of intellectual work. Finally, the conclusion summarizes the argument for methodological governance as the basis for an accountable and productive human-LLM research partnership.

2 Factual Accuracy, Responsibility, and the Limits of Fluency

The integration of LLMs into academic workflows necessitates a clear articulation of the epistemic and ethical principles that should govern their use. At the forefront of these principles is an unwavering commitment to factual accuracy. In academic practice, factual accuracy is not optional, negotiable, or secondary; it is a foundational moral and epistemic requirement for any form of scholarly communication. No degree of fluency, productivity, or stylistic sophistication can compensate for the introduction of factual error into the scholarly record. Therefore, the first and most basic responsibility of any researcher or instructor using an LLM is to verify the factual correctness of its output.

However, the very power of LLMs to generate fluent, coherent, and stylistically appropriate text creates a significant challenge. The polished nature of LLM-generated content can mask underlying inaccuracies, leading to a false sense of confidence in its reliability. A fact which is compounded by the probabilistic nature of LLMs, which are designed to generate plausible sequences of text rather than to retrieve and report verified facts [Chen et al., 2023, Sansford et al., 2024]. As many have noted, this can lead to the phenomenon of "hallucination," where an LLM generates statements that are entirely fabricated but presented with the same confidence as factual claims [Wysocka et al., 2024, Abishethvarman et al., 2025]. The responsibility for detecting and correcting these hallucinations rests squarely with the human user.

More importantly, the present paper argues that a singular focus on factual accuracy, while essential, is insufficient for the responsible governance of LLMs in academia. The core of academic work lies not merely in the accumulation of facts, but in their organization, interpretation, and synthesis into coherent structures of knowledge. An LLM output may be composed entirely of factually correct statements and yet fundamentally misrepresent the knowledge of a field by distorting the relationships between concepts, misattributing the significance of different works, or constructing a logically flawed argument. The distinction between local factual correctness and global structural integrity is central to the proposed framework.

We therefore advocate for a shift in perspective: from viewing LLMs as tools for *uncontrolled artificial generation* to engaging with them as partners in an *accountable research partnership*. An uncontrolled generator is a system that produces fluent text without any user-side validation of its structural or bibliographic coherence. An accountable partnership, in contrast, is a collaborative process in which the human researcher or instructor actively governs the LLM's contributions through explicit methodological, bibliographic, and evaluative protocols [Shankar et al., 2024, Pan et al., 2025]. Such a shift requires a move beyond simple fact-checking to a more holistic assessment of the knowledge structures that LLM outputs represent.

The normative framework for this partnership is grounded in the core values of academic inquiry: intellectual honesty, rigorous argumentation, and a commitment to building upon a shared and verifiable body of knowledge. The use of an LLM does not absolve the researcher of these responsibilities; it heightens them. The following sections of the current study will detail a practical, computationally-grounded framework for enacting a methodological responsibility, providing tools to assess not just the factual content of LLM outputs, but their structural and bibliographic integrity as well.

3 Structural Integrity and the Problem of Hallucination

While the concept of "hallucination" in LLMs is most commonly associated with the generation of factually incorrect statements, the present paper argues for a broader and more functionally significant understanding of the term. We define *structural hallucination* as the generation of content that, while potentially factually accurate at the sentence level, fundamentally distorts the relational structure of academic knowledge. This type of hallucination is more insidious and often more damaging than simple factual error, as it can mislead researchers, students, and reviewers in ways that are difficult to detect with standard fact-checking procedures [Nonaka and Perry, 2025, Chen et al., 2023].

Academic knowledge is not a flat collection of facts; it is a complex, hierarchical, and deeply relational structure. It is composed of concepts linked by dependencies, causal arguments, methodological hierarchies, and schools of thought. The significance of a particular claim is determined not only by its truth value, but by its position within the underlying structure. Structural hallucination occurs when an LLM fails to accurately represent such a relational context.

Structural hallucination manifests through several interrelated mechanisms that systematically distort the organization of academic knowledge. One prominent mechanism is conceptual re-centering, whereby an LLM assigns disproportionate importance to peripheral, weakly grounded, or even spurious concepts. In such cases, ideas that occupy marginal positions in the scholarly literature are presented as foundational or paradigmatic, effectively reshaping the perceived intellectual core of a field. This re-centering is especially dangerous because it often mirrors the stylistic conventions of authoritative academic writing, making the distortion difficult to detect without structural analysis.

A second manifestation of structural hallucination concerns bibliographic distortion. LLMs may generate citations that appear plausible but do not correspond to any existing publication, misattribute ideas to incorrect authors, or compress complex intellectual genealogies into misleadingly simplified narratives. Even when citations are factually correct, their selection and placement may distort the actual structure of influence within a field, overemphasizing secondary sources while omitting canonical works that define the discipline's conceptual backbone.

A third manifestation involves logical mis-structuring. Here, LLM-generated text may construct arguments that are locally coherent yet globally flawed, relying on false equivalences, invalid causal chains, or the misrepresentation of theoretical dependencies. Such arguments often pass superficial coherence checks because each individual claim appears reasonable, but the overall argumentative structure fails to reflect the epistemic standards of scholarly reasoning.

These failures often evade sentence-level fact-checking because the individual statements that compose the structurally flawed argument may themselves be factually correct. For example, an LLM might correctly state that Author A published a paper on Topic X and that Author B published a paper on Topic Y, but then incorrectly imply that Author B's work was a direct response to Author A's, thereby fabricating a scholarly dialogue that never occurred. Such a type of error is consequential in manuscripts, theses, grant proposals, and assessment contexts, as it can lead to fundamentally flawed research and a distorted understanding of a field [Chen et al., 2023].

All these forms of hallucination are particularly dangerous because they do not trigger the same alarms as obvious factual errors. A structurally hallucinated text can be fluent, well-written, and internally consistent, making it difficult to challenge without deep domain expertise. Something that poses a significant threat to the integrity of peer review, the reliability of student assessment, and the overall progress of scholarly inquiry.

The guiding question for evaluating LLM outputs must therefore be: *Does an output remain factually correct while also fitting coherently, bibliographically, and intellectually within the established knowledge landscape of the field?* To answer the previous question, we need tools that can move beyond the analysis of individual statements to the evaluation of the overall structure of the knowledge being presented. The following sections will introduce a methodology for doing just that, using techniques from network science to make the relational structure of LLM-generated text explicit and available for inspection and validation.

4 From Text to Network: Knowledge Graphs for LLM Evaluation

To address the challenge of structural hallucination, we propose a methodological framework that translates unstructured text into a structured, analyzable format. Our framework provides such a methodology by translating text—whether produced by an LLM, written by a human with LLM assistance, or drawn from trusted reference sources—into lightweight *knowledge graphs*. This allows for the quantitative evaluation of the structural properties of LLM-generated content against a trusted baseline, such as a corpus of peer-reviewed literature or a curated bibliography [Liu et al., 2024].

A knowledge graph is a network representation of knowledge, where concepts, entities, or documents are represented as *nodes*, and the relationships between them are represented as *edges*. Thus, a knowledge graph transforms a document from a linear sequence of sentences into a network of concepts and relationships [Markowitz et al., 2025, Algaba et al., 2025]. Such a structured representation provides a basis for quantitative analysis that is not possible with the raw text alone.

To render structural hallucination empirically detectable, we propose a methodological framework that translates unstructured academic text into explicit network representations. Doing so, such a translation enables the inspection of conceptual and bibliographic structure independently of surface-level fluency.

Knowledge Graph Extraction: The first component of this framework involves the extraction of a knowledge graph from the text. Using established natural language processing techniques, key concepts, entities, and relational predicates are identified and represented as nodes and edges. The resulting graph encodes the conceptual architecture implicit in the document, making visible which ideas function as organizing principles and how subsidiary concepts are positioned relative to them [Wen et al., 2024, Pan et al., 2024]. In particular, from the network–data computational point of view, relationships can be identified in several ways, including: *Co-occurrence:* If two concepts frequently appear in the same sentence or paragraph, an edge can be drawn between them. The weight of the edge can represent the frequency of co-occurrence. *Syntactic Dependency Parsing:* Analyzing the grammatical structure of sentences can reveal more specific relationships, such as subject-verb-object triples, which can be translated into directed edges in the graph (e.g., "Researcher A *proposes* Theory X"). *Semantic Similarity:* Using word embeddings or other semantic models, one can draw edges between concepts that are semantically close, even if they do not co-occur directly in the text.

Bibliometric Network Construction: In parallel, the bibliographic content of the text is used to construct a co-authorship, citation or shared keywords network that represents the intellectual lineage, scholarly context, and semantic positioning of the work. Nodes correspond to authors, cited publications or underlying keywords, while edges encode co-authorship, citation or shared keywords relationships and thematic proximity. Such networks provide a structural map of how the text situates itself within the existing scholarly literature [Leon et al., 2017, Algaba et al., 2025].

Comparative Network Analysis: The power of the final component of the discussed approach lies in comparisons between the knowledge networks induced by the evaluated text and reference networks derived from trusted corpora. To evaluate an LLM-generated text, we first construct a knowledge graph from it. We then compare the knowledge graph to one or more "reference graphs" constructed expert-curated sources [Liu et al., 2024, Sansford et al., 2024]. The reference graphs can be created from various sources that might include: *Textbooks and Review Articles:* These sources provide a canonical representation of the established knowledge structure of a field. *Expert-Curated Outlines:* A syllabus or the bibliography suggested for a graduate-level course, for example, represents an expert's view of the important concepts in a field and their logical relationships. *Bibliographic Databases:* Various bibliometric networks, as previously mentioned, can be treated themselves as knowledge graphs. Structural divergence between these networks serves as an indicator of conceptual distortion or bibliographic misalignment, thereby operationalizing the assessment of structural integrity.

Thus, the considered framework provides a practical and scalable approach to validating the structural integrity of LLM-assisted work, moving beyond a simple reliance on surface-level fluency and local factual accuracy.

5 Network Diagnostics as a Hallucination Stress Test

Once text has been translated into knowledge graphs, a rich toolkit of network analysis techniques becomes available for the diagnosis of structural distortions in LLM-generated content. These quantitative measures allow us to move beyond subjective impressions of coherence and to pinpoint specific, falsifiable claims about the structure of the knowledge being presented. We designate this comparative, structure-based diagnostic procedure, providing quantitative metrics for the evaluation of the structural integrity of LLM-generated content, as a "hallucination stress test," since it examines the stability of LLM-induced knowledge representations under comparison with trusted reference structures [Sansford et al., 2024, Liu et al., 2024]. The hallucination stress test procedure involves: *Identifying Core Concepts:* From the reference graph, identify the top-ranked concepts according to various centrality measures. These are the canonical concepts of the field. *Checking for Displacement:* Examine the centrality rankings in the LLM-generated graph. Are the canonical concepts still central? Or have they been displaced by other concepts? *Investigating Upwardly Mobile Concepts:* Identify any concepts that have a significantly higher centrality rank in the LLM graph than in the reference graph. These are potential "hallucinations." Investigate their provenance. Are they real but peripheral concepts that have been over-emphasized? Or are they entirely fabricated?

Key network diagnostics of hallucination stress include:

Centrality Analysis: Centrality measures offer a direct way to identify which concepts function as structural hubs within the graph. When applied comparatively, discrepancies in centrality rankings between an LLM-generated graph and a reference graph reveal cases in which peripheral or fabricated concepts are artificially elevated. A centrality-based comparison may probe whether the conceptual center of gravity of the text aligns with that of the field [Sansford et al., 2024, Tang et al., 2025]. Various centrality measures capture different aspects of importance: *Degree Centrality:* The number of connections a node has. In a knowledge graph, a concept with high degree centrality is one that is connected to many other concepts, suggesting it plays a broad, integrative role. *Betweenness Centrality:* A measure of how often a node lies on the shortest path between other nodes. A concept with high betweenness centrality acts as a bridge, connecting different clusters of ideas. Foundational theories or methods often have high betweenness centrality. *Eigenvector Centrality:* A measure of a node's influence, which takes into account the centrality of its neighbors. A concept is important if it is connected to other important concepts.

Modularity: Modularity provides a complementary measure by quantifying the strength of thematic separation within the graph [Tang et al., 2025, Nonaka and Perry, 2025]. Low modularity in an LLM-generated knowledge graph suggests a breakdown of conceptual boundaries, often reflecting generic or overly homogenized representations of complex intellectual landscapes. Together, these diagnostics transform qualitative concerns about coherence into inspectable, quantitative signals.

Community Detection: Community structure analysis further refines the assessment by examining how concepts cluster into thematic or methodological groups. In well-structured academic texts, such clusters correspond to recognizable subfields or lines of argumentation. Significant deviations in clustering patterns may indicate that an LLM has conflated unrelated themes or constructed artificial bridges between conceptually distinct areas [Sansford et al., 2024, Tang et al., 2025, Chen et al., 2023].

In addition, we can assess the overall structural similarity between the LLM graph and the reference graph using measures of *graph isomorphism* or *graph edit distance* [Sansford et al., 2024, Liu et al., 2024, Tang et al., 2025]. These measures provide a single, quantitative score for the degree of structural alignment. Furthermore, by generating multiple outputs from near-identical prompts and comparing the resulting knowledge graphs, we can assess the stability of the LLM's knowledge representation. High variance in graph structure across similar prompts is a signal of unreliable or opportunistic generation [Shankar et al., 2024, Tang et al., 2025].

6 Citation Integrity and Bibliometric Grounding

Beyond the conceptual structure of an argument, a critical dimension of academic integrity is its *bibliographic grounding*. Scholarly claims are not made in a vacuum; they are situated within a landscape of existing research, and such a situatedness is made explicit through citation. LLMs pose a significant threat to the broader ecosystem, as they are capable of generating citations that are fabricated, misattributed, or structurally implausible. The present section details a set of protocols for validating the citation integrity of LLM-generated content and for grounding it within real scholarly ecosystems using bibliometric analysis [Algaba et al., 2025, Mustaffa et al., 2025, Liu et al., 2025].

The most basic form of bibliographic failure is the *fabricated reference*, where an LLM generates a citation to a

paper that does not exist. While tools like CrossRef and Google Scholar can be used to check for the existence of a cited work, a more subtle problem is the *implausible citation*. An LLM might cite a real paper by a real author, but in a context where that paper is entirely irrelevant. For example, it might cite a paper on quantum physics in the middle of a literature review on 19th-century French poetry. While the citation itself is "real," its use is a form of bibliographic hallucination.

Citation Validation: A first step in governing LLM-generated content is the systematic parsing of all cited references and their automatic verification through bibliographic APIs. After constructing the LLM-assisted citation network, we need to compare it to established citation networks from databases (like Web of Science, Scopus, Dimensions, OpenAlex, Google Scholar etc.). This comparison allows us to detect several forms of bibliographic distortion: *Missing Canonical Works:* By comparing the citation network of the LLM-generated text to the citation network of a field, we can identify whether the LLM has omitted crucial, highly-cited works that should be present in any competent review of the topic. *Distorted Citation Neighborhoods:* We can analyze the "citation neighborhood" of the papers cited by the LLM. Do the papers that cite them, and the papers that they cite, form a coherent intellectual community? Or has the LLM plucked a paper from one context and inserted it into a completely unrelated one? *Implausible Author-Topic Pairings:* By analyzing the broader publication record of the authors cited by the LLM, we can detect cases where an author is cited for a topic on which they have never published. Thus, researchers may establishe a minimal baseline check for fabricated or non-existent references. Beyond simple existence, bibliographic integrity constitutes a central axis of methodological governance. Given the documented tendency of LLMs to generate plausible but fictitious citations or to misattribute scholarly work, all cited sources must be verified against authoritative bibliographic databases to ensure both their existence and the correctness of their metadata. The validation step is essential for maintaining the integrity of scholarly ecosystems and for preventing the propagation of spurious references in academic manuscripts, theses, and evaluations [Algaba et al., 2025, Mustaffa et al., 2025]. When integrated with the structural comparison procedures described above, citation validation functions as a bibliographic component of the hallucination stress test, revealing cases in which relational coherence at the conceptual level is accompanied by distortions in scholarly attribution.

Bibliometric Benchmarking: Beyond mere existence checks, bibliometric benchmarking situates LLM-generated citation patterns within the empirically observed structure of a scholarly field. By comparing citation frequency, author prominence, and journal representation—together with network-level properties such as degree distributions, clustering coefficients, and related topological measures—between the citation graph induced by an LLM-generated text and reference graphs derived from established bibliometric baselines, it becomes possible to identify systematic distortions. These include the overrepresentation of marginal sources, the suppression of canonical works, or abnormal patterns of intellectual concentration. Such deviations are best understood not as isolated citation errors, but as indicators of deeper structural hallucinations affecting the representation of the field's knowledge landscape [Leon et al., 2017, Algaba et al., 2025, Tang et al., 2025, Mustaffa et al., 2025].

Author and Journal Analysis: Finally, the distribution of authors, publication venues, and intellectual traditions referenced in a text can be analyzed to uncover systematic biases, thematic narrowing, or distortions in scholarly representation. Patterns such as the disproportionate concentration on a limited set of authors or journals, or the systematic omission of established venues, may indicate that an LLM-generated citation profile fails to reflect the empirical structure of the field [Leon et al., 2017, Algaba et al., 2025, Mustaffa et al., 2025].

By subjecting LLM-assisted work to the same structural and bibliographic validation applied to any other academic output, we ensure that it meets disciplinary standards regardless of how it was produced. Responsibility for the validation process remains with the human author, and the act of performing and documenting it constitutes a substantive intellectual contribution in its own right. In such a way, LLM-assisted scholarship remains anchored in the actual ecology of scientific communication rather than devolving into a self-consistent but fictitious approximation of scholarly knowledge [Shankar et al., 2024, Wysocka et al., 2024, Chen et al., 2023].

7 Attribution, Assessment, and Accountable Research Partnerships

More broadly, the use of LLMs in the production of scholarly work raises complex questions of *attribution*. When a researcher uses an LLM to assist in drafting a manuscript, the boundary between the researcher's own intellectual contribution and the model's generative output may become blurred. We may address such issues by offering a partial response to the key challenge by shifting attention away from the internal process of text production and toward the *verifiable properties* of the final scholarly artifact [Shankar et al., 2024, Wysocka et al., 2024, Chen et al., 2023].

The ultimate goal of the governance framework proposed in the current study is to enable a responsible and productive partnership between humans and LLMs in academic work. The final section presents now a practical toolkit for governing the use of LLMs across three core practices of academic life: manuscript writing, syllabus design, and student assessment.

Assessment of Student Work: In student evaluation and academic assessment, one needs a principled basis for evaluating LLM-assisted student work. By analyzing conceptual structure and citation patterns, instructors can distinguish genuine intellectual engagement from fluent but structurally hollow reproduction. This is a shift which is particularly important in addressing concerns about plagiarism, which in the context of LLM use cannot be reduced to text reuse alone but must also account for uncritical delegation of intellectual labor and the reproduction of unvalidated model outputs. Such an approach enables a more pedagogical fair attribution of originality and creativity without relying on crude detection or prohibition strategies [Wysocka et al., 2024, Shankar et al., 2024]. Rather than simply trying to determine *if* a student used an LLM, instructors may use certain protocols to assess *how well* they used it. A student who submits a piece of LLM-generated work that is structurally flawed and bibliographically ungrounded has failed to engage in the kind of critical validation that is central to academic work [Chen et al., 2023]. In contrast, a student who uses an LLM as a starting point but then applies the validation protocols described in the present paper to refine, correct, and properly situate the model's output has demonstrated a high level of scholarly competence. This approach allows for the design of more robust and meaningful assessment practices [Domenichini et al., 2025]: *Evaluating LLM-Assisted Homework and Projects:* Instead of simply grading the final output, instructors can ask students to submit a "validation report" detailing the steps they took to verify the structural and bibliographic integrity of their work. Within such an assessment framework, critical engagement with the LLM's output is explicitly incorporated [Wysocka et al., 2024, Shankar et al., 2024]. *Fair and Transparent Grading:* By grounding assessment in objective, verifiable properties of the submitted work (e.g., the structural alignment of its knowledge graph with a reference model, the integrity of its citations), instructors can develop grading rubrics that are more fair and transparent [Chen et al., 2023, Domenichini et al., 2025]. *Supervision of Theses and Dissertations:* In the context of graduate supervision, institutions may provide a concrete set of tools for mentors to guide students in the responsible use of LLMs in their research. The validation protocols can become a regular part of the feedback and revision process [Wysocka et al., 2024, Shankar et al., 2024].

Manuscript Writing: In research writing, the proposed governance framework enables the systematic validation of literature reviews, conceptual frameworks, and interpretive sections by assessing whether their structural and bibliographic organization faithfully reflects the field they claim to engage. Rather than treating fluency, stylistic coherence, or rhetorical sophistication as proxies for scholarly quality, the framework evaluates whether central concepts, intellectual lineages, and argumentative dependencies are positioned in ways that align with established disciplinary structures. In doing so, concerns about plagiarism in LLM-assisted writing can be reframed, shifting attention away from surface textual similarity and toward the substantive question of whether borrowed formulations, arguments, or structures have been explicitly acknowledged, critically validated, properly attributed, and intellectually integrated. By translating both LLM-assisted drafts and trusted reference corpora into comparable knowledge and citation networks, researchers can identify conceptual re-centering, omitted foundational work, or distorted patterns of influence that would otherwise remain hidden beneath polished prose. Accordingly, evaluation shifts decisively away from surface-level textual polish toward epistemic alignment, intellectual contribution, and accountable engagement with the scholarly record [Chen et al., 2023, Liu et al., 2024, Sansford et al., 2024, Algaba et al., 2025].

Syllabus Design: In instructional contexts, the framework supports the design and evaluation of syllabi by examining whether selected readings, topics, and learning sequences form a coherent, representative, and pedagogically meaningful structure. Instead of accepting LLM-generated curricula at face value, instructors can assess whether foundational works are appropriately emphasized, whether conceptual progressions reflect disciplinary logic, and whether major subfields are proportionally represented. By modeling syllabi as structured knowledge and bibliometric networks, it becomes possible to detect thematic gaps, excessive homogenization, or the marginalization of core traditions. This approach allows instructors to use LLMs as generative aids while retaining expert control over curricular integrity, ensuring that course design remains anchored in the actual structure and evolution of the field rather than in a generic or flattened approximation of it [Leon et al., 2017, Domenichini et al., 2025, Stocker et al., 2024].

By integrating these governance protocols into their workflows, researchers, instructors, and students can move towards a more accountable and productive partnership with LLMs, one that harnesses their power while upholding the core values of academic inquiry.

8 Conclusion

The present paper has argued for a fundamental reorientation in how Large Language Models are integrated into academic work—an integration that brings with it both substantial opportunities and serious risks. Rather than adopting a primarily reactive stance focused on the detection of factual errors or plagiarism, we have advocated a proactive framework of methodological governance. Our central claim was that the most consequential risks posed by LLMs are not isolated inaccuracies, but systemic forms of structural hallucination that distort the conceptual organization and bibliographic grounding of scholarly fields. Addressing these risks requires a shift from treating LLMs as uncontrolled artificial generators to establishing an accountable research partnership grounded in explicit methodological oversight.

To this end, we have presented a practical, researcher-centric framework for the governance of LLM-assisted academic work. The framework is grounded in a set of individually implementable computational protocols that combine knowledge graph extraction, citation and reference parsing, and network-based analysis. Together, these techniques provide a transparent and inspectable means of evaluating the structural coherence and bibliographic integrity of LLM-generated or LLM-assisted texts. By equipping researchers, instructors, and students with tools to assess not only factual correctness but also conceptual alignment and scholarly situatedness, the framework shifts responsibility decisively back to the human user while preserving the productive potential of these technologies.

The transition from uncontrolled artificial generation to an accountable research partnership is not merely technical; it is normative. It requires a renewed commitment to intellectual rigor, a willingness to engage critically and methodologically with emerging tools, and a reaffirmation of core academic values—honesty, transparency, responsibility, and a commitment to verifiable knowledge. In such a view, governance is not an external constraint imposed on creativity, but an enabling condition for meaningful and trustworthy scholarly contribution.

The tools and protocols outlined in the current work are not intended as a final or exhaustive solution. Rather, they are offered as a tentative starting point for a broader conversation about how LLMs can be responsibly incorporated into research, teaching, and evaluation without eroding the foundations of academic inquiry. Future work will involve refining these validation procedures, developing more robust benchmarks for structural and bibliographic integrity, and exploring how such governance frameworks might be integrated into editorial practices, peer review, and institutional assessment.

The challenge of integrating LLMs into academic life is ongoing and complex. Yet it is a challenge that can be met neither through fear-driven prohibition nor through uncritical and overstated enthusiasm driven by hype, but through the same spirit of rigorous, reflective, and methodologically grounded inquiry that has long defined serious scholarly work.

Data Availability Statement

The present work is conceptual and methodological in scope; it does not report original empirical data, but instead proposes a governance framework intended to inform and support future empirical and computational studies.

References

[Abishethvarman et al., 2025] Abishethvarman, V., Sabrina, F., and Kwan, P. (2025). Knowledge integrity in large language models: A state-of-the-art review. *Information*, 16(12):1076.

[Algaba et al., 2025] Algaba, A., Holst, V., Verbeken, B., Tori, F., Wenmackers, S., Mobini, M., and Ginis, V. (2025). How deep do large language models internalize scientific literature and citation practices? *arXiv preprint arXiv:2504.02767*.

[Chen et al., 2023] Chen, L., Deng, Y., Bian, Y., Qin, Z., Wu, B., Chua, T.-S., and Wong, K.-F. (2023). Beyond factuality: A comprehensive evaluation of large language models as knowledge generators. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 6325–6341.

[Domenichini et al., 2025] Domenichini, D., Strauß, S., Gombert, S., Rummel, N., Drachsler, H., Neumann, K., Chiarello, F., Fantoni, G., and Kubsch, M. (2025). Leveraging ai and network analysis to uncover learning trajectories of energy to foster knowledge-in-use in science education. *Disciplinary and Interdisciplinary Science Education Research*, 7:28.

[Leon et al., 2017] Leon, R.-D., Rodríguez-Rodríguez, R., Gómez-Gasquet, P., and Mula, J. (2017). Social network analysis: A tool for evaluating and predicting future knowledge flows from an insurance organization. *Technological Forecasting and Social Change*, 114:103–118.

[Liu et al., 2025] Liu, X., Wang, C., Yue, Y., Guo, Q., Hu, X., Tang, X., et al. (2025). Survey on factuality in large language models. *ACM Computing Surveys*.

[Liu et al., 2024] Liu, X., Wu, F., Xu, T., Chen, Z., Zhang, Y., Wang, X., and Gao, J. (2024). Evaluating the factuality of large language models using large-scale knowledge graphs. *arXiv preprint arXiv:2404.00942*.

[Markowitz et al., 2025] Markowitz, E., Galiya, K., Ver Steeg, G., and Galstyan, A. (2025). KG-LLM-bench: A scalable benchmark for evaluating LLM reasoning on textualized knowledge graphs. *arXiv preprint arXiv:2504.07087*.

[Mustaffa et al., 2025] Mustaffa, N. E., Lai, K. E., Preece, C. N., and Wong, F. Y. (2025). A bibliometric review of large language model hallucination. *International Journal of Research and Innovation in Social Science*, 9(9):5025–5037.

[Nonaka and Perry, 2025] Nonaka, H. and Perry, K. E. (2025). Evaluating llm story generation through large-scale network analysis of social structures. *arXiv preprint arXiv:2510.18932*.

[Pan et al., 2024] Pan, S., Luo, L., Wang, Y., Chen, C., Wang, J., and Wu, X. (2024). Unifying large language models and knowledge graphs: A roadmap. *IEEE Transactions on Knowledge and Data Engineering*, 36(7):3580–3599.

[Pan et al., 2025] Pan, S., Luo, L., Wang, Y., Chen, C., Wang, J., and Wu, X. (2025). Knowledge graphs and their reciprocal relationship with large language models. *MAKE: Multimodal AI Knowledge Engineering*, 7(2):38.

[Sansford et al., 2024] Sansford, H., Richardson, N., Petric Maretic, H., and Nait Saada, J. (2024). Grapheval: A knowledge-graph based llm hallucination evaluation framework. In *Workshop on Knowledge-infused Learning co-located with 30th ACM KDD Conference*. arXiv:2407.10793.

[Shankar et al., 2024] Shankar, S., Zamfirescu-Pereira, J. D., Hartmann, B., Agrawala, M., and Parameswaran, A. (2024). Who validates the validators? aligning llm-assisted evaluation of llm outputs with human preferences. In *Proceedings of the 37th ACM Conference on User Interface Software and Technology*, pages 3654777–3676450. ACM.

[Stocker et al., 2024] Stocker, M., Oelen, A., Karras, O., and Auer, S. (2024). Leveraging large language models for semantic scholarly knowledge summarization in the open research knowledge graph. *Information*, 15(6):328.

[Tang et al., 2025] Tang, L., Zhang, X., Zhang, X., Yu, X., and Cheng, H. (2025). Statistical network analysis for llms via knowledge graphs. *TechRxiv preprint*. Posted on 10 Dec 2025.

[Wen et al., 2024] Wen, Y., Wang, Z., and Sun, J. (2024). Mindmap: Knowledge graph prompting sparks graph of thoughts in large language models. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 10370–10388.

[Wysocka et al., 2024] Wysocka, M., Wysocki, O., Delmas, M., Mutel, V., and Freitas, A. (2024). Large language models, scientific knowledge and factuality: A framework to streamline human expert evaluation. *Journal of Biomedical Informatics*, 158:104724.

[Zhang et al., 2024] Zhang, Y., Chen, Z., Guo, L., Xu, Y., Zhang, W., et al. (2024). Making large language models perform better in knowledge graph completion. In *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management*, pages 3664647–3681327. ACM.