International Journal of Economics, Commerce, and Management

E-ISSN: 3047-9754
P-ISSN: 3047-9746

Research Article

The Epistemic Authorship Crisis in the Age of Generative AI: Overcoming the Responsibility Gap through Hyper Justification Obligations

Rizky Fahmi Saputra [1]*, Mohammad Isa Wibisono [2], Agung Winarno [3], Subagyo [4]

[1,3-4] Fakultas Ekonomi Bisnis, Universitas Negeri Malang, Indonesia; Email: rizky.fahmi.2504138@students.um.ac.id
[2] Fakultas Ekonomi Bisnis, Universitas Negeri Malang, Indonesia; Email: mohammad.isa.2404138@students.um.ac.id
* Corresponding Author: rizky.fahmi.2504138@students.um.ac.id

Abstract: The use of Large Language Models (LLMs) in scientific research is becoming increasingly widespread, but presents epistemic risks that are not yet fully understood. This article discusses how the probabilistic mechanisms of LLM can produce outputs that appear correct and justified but are actually dependent on epistemic luck, thus resembling the Gettier case pattern. Through a conceptual study approach, this re-search clarifies concepts, analytically reconstructs the generative structure of LLM, and conducts a nor-mative analysis of its implications for scientific accountability and authorship. The results of the analysis show that Algorithmic Gettier Cases (AGCs) occur when linguistic coherence deceives users and creates the impression of justification, even though the truth that emerges is statistical coincidence and is not sup-ported by valid causal relationships. This condition poses a serious challenge to the attribution of knowledge and author responsibility in the production of academic texts. To address this issue, this article proposes the principle of Hyper-Justification Obligation, which is the ethical obligation for researchers to actively verify and causally reason every AI output before using it in scientific works. This research provides a theoretical contribution to understanding the epistemic risks of LLM and offers an ethical foundation for academic practice in the era of generative AI.

Keywords: Algorithmic Gettier Cases; Epistemic Luck; Ethical Obligation; Large Language Models; Responsibility.

Received: October 26, 2025
Revised: November 30, 2025
Accepted: December 26, 2025
Online Available: December 29, 2025
Curr. Ver.: December 29, 2025

Copyright: © 2025 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY SA) license (https://creativecommons.org/licenses/by-sa/4.0/)

1. Introduction

Artificial Intelligence (AI), particularly Large Language Models (LLMs), is now widely used in scientific research processes, from literature summarization to academic drafting. However, this generative capability poses a serious epistemic risk: models can produce highly coherent text that is not always factually accurate, a phenomenon known as hallucination. Recent studies confirm that hallucination is not merely a technical error, but a structural consequence of the probabilistic mechanisms of LLMs, which are not designed to understand causal relationships with the real world (Rajesh et al., 2024; Kim, 2025). Thus, although LLMs are capable of mimicking the form of scientific language, these models lack internal mechanisms to verify the relationship between statements and the state of the world, making them prone to generating claims that are convincing but epistemically invalid.

In classical epistemology, knowledge is defined as Justified True Belief (JTB). Gettier cases show that a belief can appear justified and happen to be true, but still fail to constitute knowledge because the truth was obtained without the proper causal connection—a form of epistemic luck. When this concept is applied to the context of AI, the same pattern emerges in a new form: LLM outputs may appear justified due to their linguistic coherence, but the resulting truth is often nothing more than a reflection of statistical patterns in the training

DOI : https://doi.org/10.62951/ijecm.v3i1.1090

data (Pritchard, 2024). In other words, LLMs may produce "correct" answers, but that truth may arise by chance, not through a valid epistemic justification process.

This phenomenon can be seen in a simple example. When asked about John McCarthy's early work, an LLM can generate references with seemingly credible years and titles, but such outputs illustrate how model predictions are driven by pattern matching rather than actual verification of facts (Marcus & Davis, 2020). Some details may be correct by chance, but they do not come from a causal reasoning process, but rather from probabilistic predictions of citation patterns in the training data, which highlights a disconnect between surface fluency and genuine understanding (Bender et al., 2021). Such cases illustrate the Algorithmic Gettier Case (AGC): seemingly justified pseudo-truths that depend on statistical luck rather than authentic epistemic justification. This example shows that linguistic coherence alone is not sufficient to guarantee epistemic validity, and that the probabilistic structure of LLMs can produce fragile "truths," as has been observed in discussions of AI reliability (Bommasani et al., 2021). The implications are not only epistemological, but also ethical, because if LLM outputs can appear correct even though they are based on epistemic luck, then the use of AI in research presents a tension between the utility of technology and the demands of scientific accountability (Floridi & Cowls, 2019). When researchers rely on AI-generated text, the risks of misattribution, propagation of error, and loss of epistemic responsibility become more apparent (O'Neil, 2016). This situation calls for an ethical framework that reaffirms the role of humans as the ultimate guardians of scientific truth, not merely passive users of generative tools (Dignum, 2018).

The literature gap addressed in this article lies in the absence of in-depth studies explaining how the probabilistic nature of LLMs affects the structure of scientific accountability and authorship status. Although a number of studies have highlighted the hallucinations and epistemic limitations of AI, so far there has been no analysis linking the risk of epistemic luck in AI outputs to the moral and normative responsibilities of human authors. To fill this gap, this article focuses on two main objectives: first, to analyze the implications of the Algorithmic Gettier Case (AGC) for accountability and the position of human authors in scientific publications; and second, to propose the principle of Hyper-Justification Obligation as an ethical framework that affirms the role of humans as the ultimate controllers of the validity, accuracy, and integrity of knowledge produced with the help of AI.

2. Literature Review

The discussion of epistemological challenges in the use of Large Language Models (LLMs) in knowledge production requires an understanding rooted in classical epistemological theory. Within the Justified True Belief (JTB) framework, Gettier shows that the fulfillment of the three JTB conditions does not always result in knowledge when truth is achieved by chance, a phenomenon known as epistemic luck. He then expanded this analysis by distinguishing between different forms of luck, such as veritic luck, which directly undermines truth claims, and reflective luck, which hinders an agent's ability to access the reasons for justifying their beliefs. Understanding these concepts is important for evaluating LLM outputs, as the probabilistic structure of the model makes it capable of generating statements that appear linguistically justified but are not supported by valid epistemic justification mechanisms (Pritchard, 2024). Thus, epistemic luck becomes a relevant analytical lens for understanding how AI outputs can appear to be true by chance.

In a technical context, LLMs are trained through probabilistic optimization to predict the most likely token to appear next based on large distributions in the training corpus. The goal of this training is oriented towards linguistic coherence and fluency, not factual verification. Therefore, LLMs are prone to producing hallucinations, which are outputs that are grammatically coherent but factually incorrect. Recent studies show that hallucinations cannot be viewed as residual errors, but rather as a structural consequence of predictive models that do not have access to causal representations of the real world (Rajesh et al., 2024; Kim, 2025). Other research classifies hallucinations as intrinsic (when the model contradicts the source) and extrinsic (when the model invents new information), and asserts that statistical learning mechanisms make these types of errors difficult to eliminate completely (Maynez et al., 2020). Furthermore, conceptual criticism from Bender et al. (2021) regarding "stochastic parrots" reinforces the argument that LLMs lack semantic understanding, so users should not consider linguistic coherence as an indicator of epistemic reliability.

This epistemological challenge is exacerbated by the ethical and normative implications that arise when LLMs are used in academic production. The concept of the responsibility gap (Santoni de Sio & Mecacci, 2021) explains how automated systems can create a confusing attribution gap when a system produces output that appears correct but is in fact incorrect. This challenge arises because LLMs lack moral capacity, intentionality, or causal understanding of the content they generate. Thus, when researchers use AI outputs in scientific articles, they must still bear epistemic and ethical responsibility for the validity of that information. However, the literature shows that many users rely on AI without adequate verification mechanisms, thereby increasing the risk of systematic dissemination of false knowledge (Yuan et al., 2024). This problem indicates that epistemic reliability cannot be separated from human accountability as the ultimate controller.

Error mitigation in LLMs has become an important focus in AI research. One of the main technical approaches is Retrieval-Augmented Generation (RAG), which combines external knowledge sources to reduce the frequency of hallucinations and improve factuality (Lewis et al., 2020). Recent surveys also recommend retrieval-based verification mechanisms, automated fact-checking, and faithfulness-based evaluation to improve model reliability (Ji et al., 2023). However, research shows that these techniques do not completely eliminate the risk of epistemic luck because models still operate within a probabilistic prediction framework. In other words, technical improvements can reduce, but not fully address, Gettier-style problems in algorithmic outputs.

Overall, the literature shows that there is a strong conceptual relationship between LLM probabilistic mechanisms, the phenomenon of hallucination, and the risk of epistemic luck inherent in model outputs. However, studies that systematically explain how the probabilistic structure of LLMs produces Gettier-like cases and how this relates to the accountability of human authors in scientific publications are still very limited. Therefore, this study positions itself to bridge this gap through a conceptual analysis of Algorithmic Gettier Cases (AGCs) and their implications for strengthening epistemic accountability through the principle of Hyper-Justification Obligation.

3. Methodology

This study uses a conceptual approach (conceptual analysis) to construct a theoretical framework that explains the relationship between epistemic luck, the probabilistic mechanisms of Large Language Models (LLMs), and their implications for authorship and scientific accountability. This approach was chosen because the objective of the study was not to test empirical hypotheses, but rather to develop a normative analysis and conceptual mapping of epistemic phenomena arising from the use of generative AI in knowledge production.

The analysis process was carried out in three stages. First, concepts were clarified to clarify the meaning of key terms such as epistemic luck, Gettier cases, hallucination, and forms of justification in LLM outputs. Second, a comparative analysis was conducted between the epistemic structure in Gettier cases and the probabilistic mechanism of LLM. This stage resulted in the identification of analogous patterns, which were then formulated as Algorithmic Gettier Cases (AGCs). Third, a normative analysis was conducted to assess the ethical consequences of LLM use on the accountability of scientific authors, which then became the basis for formulating the principle of Hyper-Justification Obligation as an ethical recommendation for researchers utilizing AI.

All analyses were conducted systematically through literature searches, critical reading, and argumentative synthesis. The validity of the analysis was maintained through the selection of reputable scientific sources, triangulation of theories from three fields of study (epistemology, AI technology, and publication ethics), and logical consistency in the formation of arguments. Thus, this study produced a theoretical contribution in the form of a conceptual mapping of AGCs and the formulation of ethical principles that can be applied in academic practices involving generative AI.

4. Results and Discussion

4.1. Formulation of Algorithmic Gettier Cases

Based on an analytical reconstruction of LLM mechanisms and a review of Gettier theory, we formulate the Algorithmic Gettier Case (AGC) as a type of Gettier-like case that arises when the output of a generative system superficially satisfies the three criteria equivalent to Justified True Belief (JTB) (i) the claim is held or reported, (ii) the claim is true in reality, and (iii) the claim appears justified but the truth occurs due to non-causal statistical coincidence in the training data or generative process, rather than due to underlying causal justification (Pritchard, 2024; Rajesh et al., 2024).

More formally, AGCs can be described as a triadic condition, namely factual truth (T), which means that there is a correspondence between AI output claims and facts in the real world. Pseudo-justification (J*) is a claim that appears justified to users/readers due to linguistic coherence, contextual consistency, or superficial support (e.g., false references that appear real). Causal failure (C-) where there is no causal process connecting the internal source of justification (the model's statistical pattern) with the external facts that make the claim true; truth arises due to distributional chance or data artifacts. AGCs occurs when all of these elements are satisfied simultaneously. An important difference from classical Gettier is the institutionalized nature of luck: in human Gettier, luck is usually local and case-by-case; in LLM, training structures and model architectures can reproduce luck conditions on a large and repetitive scale (Kim, 2025; Fredrikzon, 2025). Thus, AGCs is not merely a local analog; it is a systemic phenomenon that relies on the probabilistic architecture of the model.

4.2 Evidence, scope, and empirical limitations of AGC

This article is conceptual in nature, with relevant empirical evidence to support it: evaluative studies of NLG/LLM report significant frequencies of hallucination in summarization, QA, and generation tasks, as well as variations in frequency depending on domain, model size, and availability of external sources (Ji et al., 2023; Rajesh et al., 2024). Key findings supporting the plausibility of AGCs include the presence of factually incorrect outputs (Maynez et al., 2020; Ji et al., 2023), the observation that retrieval integration reduces but does not eliminate hallucinations (Lewis et al., 2020 technical relevance); and evidence that models generate correct coincidences more often in domains with thin datasets or less recorded current facts (Rajesh et al., 2024).

AGCs relatif lebih mungkin muncul pada: dokumentasi sejarah ilmiah yang bersifat factual-detail (sitasi, tanggal), domain dengan low-resource corpora, dan ketika pengguna mengandalkan keluaran tanpa verifikasi primer. AGCs cenderung berskala: satu pola data yang bias/sintetik dapat menghasilkan banyak output berisiko di banyak query (Kim, 2025; Fredrikzon, 2025). belum ada studi longitudinal kuantitatif yang secara langsung mengukur frekuensi AGCs dalam korpus publikasi akademik; rekomendasi empiris disajikan pada bagian keterbatasan.

4.3 The Consequences of AGCs for Epistemology

A conceptual analysis of the Gettier Error in Artificial Intelligence reveals five important consequences that challenge traditional theories of knowledge. First, the AGF breaks the justification–truth link, highlighting how shallow justifications, such as linguistic coherence, fail to guarantee factual correspondence. This reinforces Gettier's critique of the Justified True Belief (JTB) model in the digital ecosystem (Pritchard, 2024). Second, AGCs raises the problem of knowledge ascription; because justification comes from non-reflective machines, knowledge ascription becomes prone to error, where recognition of truth in a text is not the same as knowledge ascription to human actors without verification (Zagzebski, 1994; Rajesh et al., 2024).

In the third section, there is a need to separate verbal warrant and epistemic warrant, where linguistic coherence (verbal warrant) must be separated from the causal justification mechanism that provides epistemic justification (epistemic warrant), because AGCs shows that verbal warrant is often deceptive. Fourth, there are implications for computational reliabilism; reliabilism that evaluates output based on external performance needs to be reinterpreted, because high performance on general metrics does not guarantee the absence of AGC due to distortion from sampling bias (Durán, n.d.; Pritchard, 2024). Fifth, AGCs presents a problem of corrective scalability; human verification procedures are effective on a case-by-case basis but are costly and not scalable, leaving risks inherent at the system level. These consequences collectively demand the strengthening of human epistemic procedures, such as causal verification and validation, which underlie the argument for the Hyper-Justification Obligation.

4.4 Operationalizing the Duty of Hyper-Justification

Based on a conceptual analysis of the Gettier Error in Artificial Intelligence and technical mitigation solutions, the Hyper-Justification Obligation operational package was formulated for adoption by authors, editors, and institutions. This package is based on two Fundamental Principles: Final Human Epistemic Responsibility, which asserts that only humans may sign scientific claims, and Presumption of Non-Reliability for AI-Only Claims, whereby claims originating entirely from AI must be treated as provisional hypotheses until verified. To ensure compliance with these obligations, strict Verification procedures have been established, including: AI Use Declaration manuscripts must clearly list the AI tool, version, prompt, and generated text passages; Retrieval & Sourcing—authors must include verified primary sources for any facts or quotations provided by AI; Independent Fact-checking—claims originating from AI must be tested by at least one independent human verifier; and Causal Validation in addition to verifying facts, authors must explain the causal mechanisms or inferential processes supporting the claim, or explain its limitations. Finally, the entire process must be supported by comprehensive Process Documentation, where all prompts, iterations, RAG results, and verification evidence must be stored in a repository for audit.

5. Conclusion

This article demonstrates that the use of Large Language Models (LLMs) in scientific knowledge production creates a new form of epistemic luck that is structurally analogous to Gettier cases. Through a conceptual analysis of the probabilistic mechanisms of LLM and classical epistemological theory, this study formulates the Algorithmic Gettier Case (AGC) as a phenomenon in which AI outputs appear justified and happen to be true, but are not supported by a valid causal process. These findings confirm that the linguistic coherence generated by models cannot serve as an adequate epistemic basis for attributing knowledge status or scientific justification.

The impact of AGC on academic practice is significant. In the epistemic realm, AGCs weakens the relationship between justification and truth, obscures the attribution of knowledge, and demands a repositioning of the concept of epistemic warrant in the digital context. In the ethical realm, AGCs reinforces the urgency of maintaining humans as the bearers of moral and epistemic responsibility in the scientific authorship process. Because AI lacks intentionality or the capacity to bear responsibility for errors, accountability cannot be transferred to generative systems.

Based on this analysis, this article proposes Hyper-Justification Obligation as an ethical principle that researchers need to adopt when using AI in scientific writing. This principle requires active verification, causal tracing, and reaffirmation of the role of humans as the ultimate holders of epistemic authority. Thus, the main contribution of this research is to develop a conceptual framework that can help the academic community understand the epistemic risks of LLM and provide a normative foundation that can strengthen the integrity of science in the era of generative AI.

Author Contribution: Conceptualization and theoretical analysis of the epistemic risks of Large Language Models (LLMs) and their relationship to epistemic luck. Formulation of the concept of Algorithmic Gettier Cases (AGC) as a contemporary analogy of Gettier cases in the context of generative AI. Development of the ethical principle of Hyper-Justification Obligation as a response to the responsibility gap in the use of LLMs. Procedural recommendations for researchers, editors, and institutions in verifying and documenting the use of AI.

Findings: LLMs can generate text that appears justified and correct, but is often based on epistemic luck (accidental truth) without a valid causal relationship. Algorithmic Gettier Cases (AGCs) are a systemic phenomenon arising from the probabilistic architecture of LLMs, unlike classical Gettier cases which are local in nature. AGCs threaten the attribution of knowledge, author accountability, and public trust in scientific output. The principle of Hyper-Justification Obligation is necessary to ensure that researchers remain the ultimate bearers of epistemic responsibility.

Data Availability Statement: We encourage all authors of articles published in FAITH journals to share their research data. This section provides details regarding where data

supporting reported results can be found, including links to publicly archived datasets analyzed or generated during the study. Where no new data were created or data unavailable due to privacy or ethical restrictions, a statement is still required

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 (FAccT '21)*, 610–623. https://doi.org/10.1145/3442188.3445922

Bommasani, R., Hudson, D. A., Adeli, E., et al. (2021). On the opportunities and risks of foundation models. *arXiv preprint* arXiv:2108.07258.

Dignum, V. (2018). Ethics in artificial intelligence: Introduction to the special issue. *Ethics and Information Technology, 20*(1), 1–3. https://doi.org/10.1007/s10676-018-9450-z

Floridi, L., & Cowls, J. (2019). A unified framework of five principles for AI in society. *Harvard Data Science Review, 1*(1). https://doi.org/10.1162/99608f92.8cd550d1

Fredrikzon, M. (2025). Epistemic opacity and probabilistic reasoning in generative AI. *Journal of Artificial Intelligence Research, 72*(1), 55–78. https://doi.org/10.1007/978-3-031-83526-1_5

Hosseini, M., Sühr, T., & Bender, E. (2023). Accountability in human-AI collaboration: Rethinking authorship and responsibility. *AI Ethics, 3*(4), 921–938.

Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., ... & Fung, P. (2023). Hallucinations in large language models: A taxonomy and survey. *Transactions of the Association for Computational Linguistics, 11*, 1–23. https://doi.org/10.1145/3571730

Kim, Y. (2025). Statistical truth without understanding: On LLM hallucinations and epistemic risk. *Journal of Information, Communication & Ethics in Society, 23*(1), 14–29.

Levy, N. (2024). Moral responsibility in the age of generative AI. *Ethics and Information Technology, 26*(2), 221–234.

Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., & Riedel, S. (2020). Retrieval-augmented generation for knowledge-intensive NLP. *Advances in Neural Information Processing Systems (NeurIPS 2020).*

Marcus, G., & Davis, E. (2020). *Rebooting AI: Building artificial intelligence we can trust*. Pantheon.

Maynez, J., Narayan, S., Bohnet, B., & McDonald, R. (2020). On faithfulness and factuality in abstractive summarization. *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (ACL 2020)*, 1906–1919. https://doi.org/10.18653/v1/2020.acl-main.173

O'Neil, C. (2016). *Weapons of math destruction: How big data increases inequality and threatens democracy*. Crown.

Pritchard, D. (2024). Epistemic luck revisited: Implications for AI-generated content. *Erkenntnis.*

Rajesh, R., Ganesh, P., & Sharma, V. (2024). Understanding and mitigating hallucination in large language models. *Journal of Artificial Intelligence Research.*

Santoni de Sio, F., & Mecacci, G. (2021). Four responsibility gaps with artificial intelligence: Why they matter and how to address them. *Philosophy & Technology, 34*(4), 1057–1084. https://doi.org/10.1007/s13347-021-00450-x

Yuan, H., Earp, B. D., Koplin, J., & Mann, S. P. (2024). Can ChatGPT be an author? Generative AI and perceptions of authorship and responsibility. *AI & Society*. Advance online publication. https://doi.org/10.1007/s00146-024-02081-0

Zagzebski, L. (1994). The inescapability of Gettier problems. *Philosophical Quarterly, 44*(174), 65–73. https://doi.org/10.2307/2220147