Disclosure is not Integrity: why AI Transparency Policies Fail in Scientific Research?*

Artem Artyukhov[1,2,3†], Oleksandr Dluhopolskyi[3,4,*,†], Dmytro Chumachenko[5,6†], Vadim Schuchmann[4†] and Volodymyr Lominskyi[4†]

[1] *Bratislava University of Economics and Business, Dolnozemska cesta, 1, Bratislava, 85235, Slovak Republic*
[2] *Sumy State University, Kharkivska St., 116, Sumy, 40007, Ukraine*
[3] *WSEI University, Projektowa St., 4, Lublin, 20209, Republic of Poland*
[4] *West Ukrainian National University, Lvivska St., 11, Ternopil, 46020, Ukraine*
[5] *University of Warwick, Coventry, CV4 7AL, United Kingdom*
[6] *National Aerospace University "Kharkiv Aviation Institute", Vadym Manko St., 17, Kharkiv, 61070, Ukraine*

Abstract

This article critically examines the limitations of transparency-oriented artificial intelligence (AI) policies in safeguarding epistemic integrity in scientific research. While disclosure requirements, such as AI acknowledgements, authorship clarifications, and method reporting, have become central governance tools, they primarily regulate the visibility of AI use rather than its cognitive implications. The study argues that transparency and integrity are analytically distinct dimensions and should not be treated as interchangeable. To formalize this distinction, the paper introduces a Transparency–Integrity matrix, which models the relationship between disclosure and epistemic responsibility, demonstrating that high transparency can coexist with low integrity. In addition, a Policy-to-outcome gap framework is developed to explain how disclosure-based policies generate formal compliance while leaving hidden forms of epistemic substitution, such as delegated reasoning and outsourced interpretation, insufficiently governed. The analysis shows that current AI governance mechanisms risk producing "synthetic compliance," where procedural adherence masks weakened intellectual accountability. The study contributes to literature by transforming a normative concern into a formal analytical model and by highlighting the structural gap between policy inputs and epistemic outcomes. It concludes that while transparency is a necessary condition for responsible AI use in science, it is not sufficient; effective governance must also safeguard human cognitive agency, judgment, and accountability in AI-mediated research.

Keywords

AI, academic integrity, transparency, epistemic responsibility, higher education, governance

1. Introduction

CArtificial intelligence (AI) has become an increasingly visible component of scientific research. Researchers now use AI systems to support literature exploration, language refinement, data handling, synthesis, and argument construction. In response, universities, journals, publishers, and funding organizations have begun to introduce transparency-oriented AI policies, most commonly requiring disclosure statements, authorship clarifications, acknowledgements of tools, and method reporting. These measures are often presented as necessary safeguards for responsible AI governance in science.

Yet the growing emphasis on disclosure has produced a conceptual and regulatory problem. Transparency makes AI use visible, but visibility does not necessarily preserve the epistemic integrity of research. A study may fully acknowledge that AI was used and still rely on the system

* *The Third International Conference of Young Scientists on Artificial Intelligence for Sustainable Development: Intelligent Systems and Artificial Intelligence (YAISD-WS 2026:ISAI) May 7-8, 2026, Ternopil, Ukraine*
[1,*] Corresponding author.
[†] These authors contributed equally.
artem.artiukhov@euba.sk (A. Artyukhov); dlugopolsky77@gmail.com (O. Dluhopolskyi); dichumachenko@gmail.com (D. Chumachenko); vadum.shuhmann@gmail.com (V. Schuchmann); l.volodimyr@gmail.com (V. Lominskyi)
0000-0003-1112-6891 (A. Artyukhov); 0000-0002-2040-8762 (O. Dluhopolskyi); 0000-0003-2623-3294 (D. Chumachenko); 0000-0002-1427-3312 (V. Schuchmann); 0009-0006-9499-748X (V. Lominskyi)
© 2026 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings ceur-ws.org ISSN 1613-0073
published 2026-11-11

CEUR-WS.org/Vol-4326/paper12.pdf

for substantial interpretation, synthesis, or reasoning. In such cases, policy requirements are formally satisfied while the human researcher's intellectual responsibility may be weakened. The core issue is therefore not simply whether AI use is disclosed, but whether the researcher remains the genuine epistemic agent of the work.

The aim of this study is to explain why AI transparency policies fail when they operate as stand-alone mechanisms of governance and to develop analytical tools capable of distinguishing formal compliance from genuine research integrity. To achieve this, the article introduces two complementary models. The first is a Transparency–Integrity matrix, formalized mathematically to distinguish between different relationships between disclosure and epistemic responsibility. The second is a Policy-to-outcome gap framework, which explains how policy-regulated inputs can generate visible compliance while leaving insufficiently governed epistemic substitution hidden. Together, these models support the article's central claim: disclosure is necessary but not sufficient for preserving integrity in AI-mediated scientific research.

2. Literature review

The rapid integration of artificial intelligence into scientific research has fundamentally reshaped knowledge production, dissemination, and evaluation processes, yet it has simultaneously exposed critical limitations of existing governance mechanisms centred on transparency and disclosure. The prevailing assumption that disclosing AI use ensures ethical integrity reflects a technocratic understanding of accountability, in which formal compliance substitutes for substantive responsibility. However, the transformation of knowledge systems driven by generative AI reveals that transparency alone does not address deeper epistemic, organizational, and ethical risks embedded in AI-assisted research practices [1; 2; 3].

The expansion of AI tools in scientific workflows has redefined knowledge management paradigms, enabling rapid content generation, data processing, and manuscript preparation, but also introducing risks related to authorship ambiguity, reproducibility, and epistemic reliability. Generative AI systems are increasingly used in drafting scientific outputs, which challenges traditional notions of intellectual contribution and accountability, particularly when disclosure policies merely require acknowledging AI assistance without evaluating its impact on research validity [3; 4; 5]. These transformations are embedded in a broader shift towards data-driven knowledge economies, where AI acts as both a productivity enhancer and a potential source of epistemic distortion [6; 7].

The inadequacy of transparency-based approaches becomes more evident when considering the role of trust and behavioral responses to AI disclosure. Evidence suggests that disclosure does not uniformly increase trust; instead, its effects are mediated by contextual and cultural factors, indicating that transparency may function more as a signaling mechanism than as a guarantee of ethical conduct [8]. This insight is particularly relevant for scientific research, where the credibility of findings depends not only on openness but also on the integrity of underlying methodologies and data practices. Consequently, transparency policies risk becoming performative instruments that legitimize AI use without addressing its substantive implications.

Organizational and institutional contexts further complicate the relationship between disclosure and integrity. The adoption of AI technologies is shaped by leadership practices, organizational climate, and institutional incentives, which may encourage superficial compliance rather than meaningful ethical reflection. In higher education and research institutions, behavioral intentions to adopt new technologies are influenced by professional standards and perceived benefits, often prioritizing efficiency gains over ethical considerations [9]. At the same time, organizational silence and hierarchical pressures may limit critical discussion of the risks of AI use, reinforcing a culture in which disclosure is treated as a procedural requirement rather than a tool for accountability [10].

The ethical challenges associated with AI are not unique to scientific research but are observed across multiple sectors, highlighting systemic governance gaps. In fields such as journalism, healthcare, and law enforcement, AI deployment raises concerns about bias, responsibility, and the

adequacy of regulatory frameworks, demonstrating that transparency alone is insufficient to ensure ethical outcomes [11]. The dual-use nature of AI technologies further complicates governance, as systems designed for beneficial purposes may also produce harmful or unintended consequences, requiring adaptive and context-sensitive leadership approaches [12]. Risk management perspectives emphasize that AI-related uncertainties cannot be mitigated solely through disclosure mechanisms, as they involve complex, non-linear interactions between algorithms, data, and human decision-making. The use of advanced models such as artificial neural networks introduces opacity and unpredictability, making it difficult to assess risks ex ante and reinforcing the limitations of transparency as a governance tool [13]. This aligns with broader evidence of the "AI paradox", where technological advancements do not automatically translate into improved outcomes, underscoring the need for deeper institutional and methodological safeguards [6].

At the same time, the interplay between AI, open science, and innovation ecosystems illustrates both opportunities and contradictions. Open science policies promote transparency, accessibility, and collaboration, contributing to innovation and knowledge diffusion, yet they may inadvertently amplify the risks associated with AI-generated content if integrity controls are weak [14]. The increasing role of AI in research and development processes, including pharmaceutical innovation and patent activity, further highlights the tension between efficiency gains and ethical governance [7]. This suggests that transparency must be complemented by robust institutional frameworks that ensure accountability, validation, and ethical oversight.

The existing literature demonstrates that while AI transparency policies are necessary, they are insufficient to guarantee integrity in scientific research. Disclosure mechanisms often operate at a formal level, failing to address deeper issues related to epistemic reliability, organizational incentives, and systemic risks. A more comprehensive governance approach is required that integrates transparency with accountability, methodological rigor, and ethical leadership. Consequently, the scientific landscape increasingly recognizes that integrity cannot be reduced to disclosure but must be embedded in the broader socio-technical systems that shape AI-driven research practices.

3. Methodology

The methodological design combines conceptual decomposition, formal modeling, and governance interpretation:

1. The methodology is conceptual differentiation. The study begins by separating two dimensions that are often conflated in contemporary policy discourse: transparency of AI use and epistemic integrity of research. This distinction is essential because disclosure requirements typically govern what is visible and reportable, whereas research integrity depends on deeper issues such as human cognitive contribution, independent judgment, and conceptual control. Once these two dimensions are distinguished, the article develops the Transparency–Integrity matrix as a classificatory framework for locating different forms of AI use in science.

2. Mathematical formalization of the matrix. Transparency is defined as a weighted function of disclosure, authorship transparency, method transparency, and reproducibility transparency.

3. Causal-extension modeling through the Policy-to-outcome gap framework. Whereas the matrix classifies relationships between transparency and integrity, the framework explains the mechanism through which disclosure-centered policies may produce formal compliance without securing substantive epistemic responsibility.

4. Interpretive policy analysis. The mathematical expressions are not treated as abstract exercises detached from institutional reality; rather, they are read as governance models that clarify where current AI policies are effective and where they remain structurally incomplete. This makes it possible to interpret disclosure requirements, authorship declarations, and method reporting not as final safeguards, but as partial instruments that may leave crucial forms of

delegated reasoning, outsourced interpretation, and synthetic argument construction insufficiently regulated.

The methodological strategy is designed to move from conceptual critique to formal explanation. The study does not attempt to estimate coefficients empirically at this stage. Instead, it provides a coherent analytical architecture through which future empirical research can code transparency and integrity variables, test the size of the policy-to-outcome gap, and compare institutional AI governance regimes across journals, disciplines, and research systems.

4. Transparency–Integrity model of AI use in scientific research

To extend the analytical argument of this study, the Transparency–Integrity matrix can be formalized as a two-dimensional model that distinguishes between the visibility of AI use and the preservation of epistemic responsibility in scientific work (Figure 1). This formalization is important because the work's central claim is not that disclosure is irrelevant, but that disclosure alone is insufficient as a safeguard of research integrity. A researcher may openly acknowledge the use of AI and still delegate core interpretive, argumentative, or synthetic work to the system. In such cases, formal transparency coexists with substantive epistemic weakness.

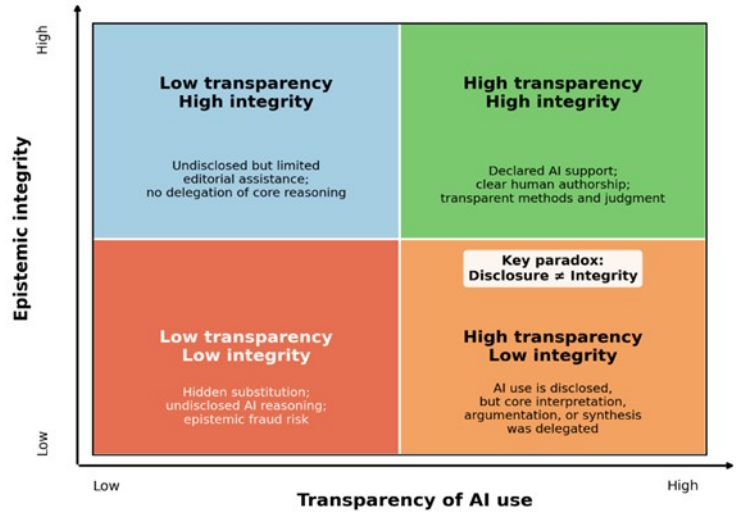


Figure 1: Transparency-Integrity matrix for AI use in science.

The matrix (Figure 1) distinguishes four possible relationships between the transparency of AI use and epistemic integrity in research. It shows that transparent disclosure does not automatically guarantee integrity, since AI use may remain openly acknowledged while still involving substantial delegation of interpretation, argumentation, or synthesis. The figure highlights the key paradox: disclosure is a necessary condition for responsible AI governance, but it is not sufficient to preserve research integrity.

Let each research practice, manuscript, or project be represented by a point (T, I), where $T \in [0,1]$ denotes the level of transparency of AI use and $I \in [0,1]$ denotes the level of epistemic integrity. The first dimension captures how fully AI use is disclosed, documented, and made visible to readers, editors, reviewers, and institutions. The second captures whether the researcher retains meaningful control over problem formulation, reasoning, interpretation, judgment, and conceptual synthesis. This distinction is analytically necessary because the two dimensions may vary independently. Practice may be transparent but epistemically weak, or relatively opaque while still preserving human intellectual responsibility.

The transparency variable is defined as a weighted sum of four components:

$$T = w_1 D + w_2 A + w_3 M + w_4 R, \quad (1)$$

$T = w_1 D + w_2 A + w_3 M + w_4 R, \quad ¿(1)$

where D denotes disclosure, A denotes authorship transparency, M denotes method transparency, and Rdenotes reproducibility transparency.

The weights satisfy:

$$w_1 + w_2 + w_3 + w_4 = 1, w_i \geq 0. \quad (2)$$

The integrity variable is defined as a weighted sum of five epistemic components:

$$I = v_1 H + v_2 E + v_3 V + v_4 J + v_5 C, \quad (3)$$

where H denotes human cognitive contribution, E denotes epistemic accountability, Vdenotes verifiability of reasoning, J denotes independence of judgment, and Cdenotes conceptual control.

The weights satisfy:

$$v_1 + v_2 + v_3 + v_4 + v_5 = 1, v_i \geq 0. \quad (4)$$

In the simplest specification, all component variables in equations (1) and (3) may be coded as binary variables, where 0 indicates the absence of the relevant feature and 1 indicates its presence.

Formally, this can be expressed as:

$$D, A, M, R, H, E, V, J, C \in 0,1. \quad (5)$$

In a more advanced empirical application, the same variables may be normalized to the interval $[0,1]$in order to capture partial or graded realization of transparency and epistemic integrity. In that case:

$$D, A, M, R, H, E, V, J, C \in [0,1]. \quad (6)$$

This formulation allows treating transparency as more than a simple yes/no declaration. A study may disclose that AI was used yet remain opaque about where in the research workflow the system intervened, what kind of contribution it made, or whether that contribution can be reconstructed and critically evaluated. Likewise, integrity is not reduced to a binary distinction between honest and dishonest conduct but can be modeled as a matter of degree depending on the extent to which human cognitive labor, accountability, and judgment remain preserved.

The two-dimensional structure generates four analytically distinct quadrants. When both transparency and integrity are high, AI use is both visible and epistemically accountable; this is the normatively preferred case. When both are low, AI use is both concealed and epistemically problematic, corresponding to the classic form of hidden substitution. When transparency is low, but integrity remains relatively high, the issue is not necessarily scientific misconduct, but incomplete reporting or weak documentation. The most significant quadrant for the present study is the one in which transparency is high while integrity is low. This is the configuration in which disclosure functions as a formal ritual of legitimacy, while substantive thinking has already been delegated.

To capture this tension directly, the model introduces a disclosure–integrity gap:

$$G = T - I. \quad (7)$$

When $G \approx 0$, transparency and integrity are broadly aligned. When $G < 0$, the epistemic quality of the work exceeds the degree to which AI use is disclosed. When $G > 0$, transparency exceeds integrity, meaning that the research appears responsibly documented even though the underlying cognitive contribution may be compromised. This last condition is especially important because it identifies the precise space in which transparency policies fail not by producing opacity, but by allowing formally visible AI use to be mistaken for epistemically responsible scientific practice.

To measure the risk that disclosure becomes a substitute for integrity rather than its support, the model further defines a synthetic compliance risk:

$$SCR = T(1 - I). \tag{8}$$

This expression reaches high values precisely when AI use is well disclosed but epistemic integrity is weak. In such cases, policy compliance may conceal rather than resolve the problem. The existence of this quadrant demonstrates why transparency-only policies remain analytically and normatively limited. They regulate what is acknowledged, but they do not necessarily regulate what is cognitively delegated.

Finally, the model allows the effectiveness of an AI policy to be treated as a function of both transparency and integrity:

$$PE = \alpha T + \beta I - \gamma G \tag{9}$$

where PE denotes policy effectiveness, Gdenotes the disclosure–integrity gap, and $\alpha, \beta, \gamma > 0.$

This expression formalizes a key argument of the work: policies that increase transparency without preserving integrity may still produce low overall effectiveness because they enlarge the gap between formal compliance and substantive epistemic responsibility.

Taken together, this model strengthens the article's central claim that disclosure is a necessary but insufficient condition for responsible AI governance in science. Transparency is valuable because it makes AI involvement visible; however, visibility alone cannot guarantee that the researcher remains the true epistemic agent of the work. The model therefore provides a formal basis for distinguishing between transparent AI use and integrity-preserving AI use, and for demonstrating why policies built solely on disclosure requirements are structurally incomplete.

5. Policy-to-outcome gap framework

While the "Transparency–Integrity" matrix identifies the four possible relationships between disclosure and epistemic responsibility, it does not, by itself, explain how a research system can arrive at the most problematic configuration: **high transparency combined with low integrity.** In other words, the matrix is analytically powerful as a classificatory tool, but it remains static. It shows where a given practice is located, yet it does not specify the governance mechanism through which disclosure-centered policies may produce formal compliance without preserving the intellectual integrity of scientific work.

For this reason, the matrix is complemented by the **Policy-to-outcome gap framework (Figure 2),** which provides a causal and institutional interpretation of the same problem. Whereas the matrix distinguishes outcomes in a two-dimensional space, the framework explains how policies that regulate visible and documentable aspects of AI use may still fail to govern the hidden delegation of reasoning, interpretation, argument construction, and judgment. Together, the two figures perform different but complementary functions: the matrix identifies the paradox, while the framework explains its policy logic. The central argument that follows from this connection is that transparency-oriented AI policies do not necessarily fail because they produce opacity, but

because they often regulate only the visible layer of AI use while leaving the epistemically decisive layer insufficiently governed.

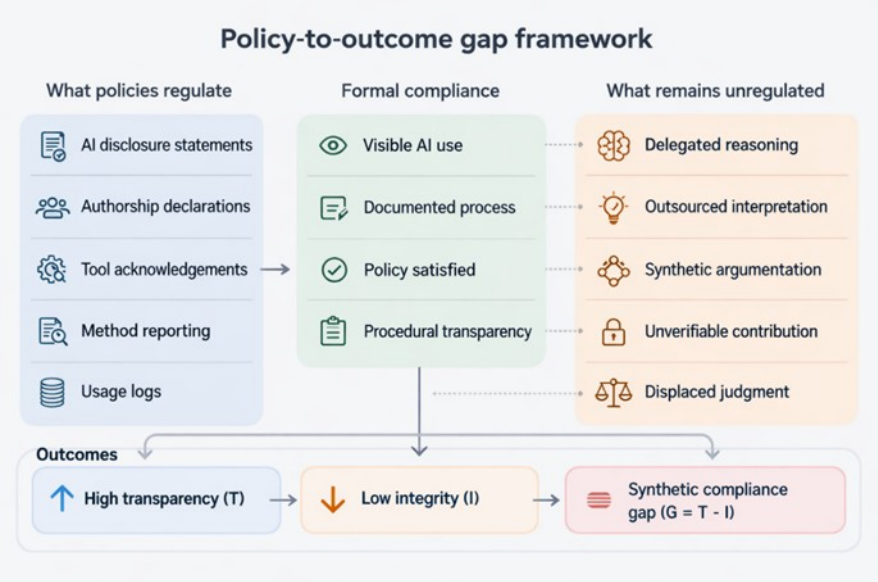


Figure 2: Policy-to-outcome gap framework.

The left-hand block, "What policy regulates", contains the forms of AI use that are most visible, reportable, and administratively manageable. These typically include AI disclosure statements, authorship declarations, tool acknowledgements, method reporting, and, in stronger policy environments, usage logs or broader transparency requirements. These elements are the preferred target of institutional and journal policies because they are documentable and auditable. They make AI use legible governance systems. The central block, "Formal compliance", represents the immediate results of such regulation. Once policies focus on disclosure and documentation, they tend to produce visible AI use, a documented process, procedural transparency, and a condition in which the policy is formally satisfied. This is the level at which researchers, editors, and institutions may reasonably conclude that AI use has been responsibly handled. However, this conclusion may be premature, as procedural compliance does not necessarily indicate whether AI has substituted for the researcher's cognitive contribution.

The right-hand block, "What remains insufficiently governed", captures precisely the epistemic dimensions that transparency policies often fail to adequately regulate. These include delegated reasoning, outsourced interpretation, synthetic argumentation, unverifiable contribution, and displaced judgment. These elements are not always visible in disclosure statements and are often difficult to verify solely through formal reporting. Yet they are the core of the integrity problem, because they determine whether the human researcher remains the actual epistemic agent of the work. The bottom block, "Outcomes", synthesizes the resulting configuration. The framework shows that the combined effect of policy-regulated transparency and insufficiently governed cognition may be a condition of high transparency, low epistemic integrity, and a resulting synthetic compliance gap, expressed analytically as $G = T - I$. The framework, therefore, makes explicit that transparency and integrity do not collapse into one another. A policy may successfully increase the first while leaving the second weak or even deteriorating.

Unlike the Transparency–Integrity matrix, which classifies research practices in a two-dimensional space, the Policy-to-Outcome Gap Framework is better formalized as a governance transmission model. Its purpose is to explain how policy-regulated inputs are converted into formal compliance while still allowing hidden epistemic substitution to persist. In this formulation, the central question is not where a given practice is located in a conceptual matrix, but how procedural regulation fails to produce the intended epistemic outcome.

Let the policy-regulated input layer be represented by the variable P, defined as:

$$P=\alpha_1 d+\alpha_2 a+\alpha_3 t+\alpha_4 m+\alpha_5 l, \tag{10}$$

where d – disclosure statements, a – authorship declarations, t – tool acknowledgements, m – method reporting, l – logging or traceability requirements.

With:

$$\alpha_1+\alpha_2+\alpha_3+\alpha_4+\alpha_5=1, \alpha_i \geq 0. \tag{11}$$

The variable P captures the extent to which the policy governs the visible and documentable layer of AI use.

The next stage is formal compliance, represented by the variable C. This is the extent to which regulated requirements are actually translated into visible, documentable compliance in practice.

It may be expressed as:

$$C=\eta P, \tag{12}$$

where $\eta \in [0,1]$ denotes the effectiveness of policy implementation. A high value of ηmeans that institutions are successful at converting policy requirements into observable compliance.

However, formal compliance does not automatically govern the epistemically decisive layer of research. Let the hidden epistemic substitution burden be represented by the variable U, defined as:

$$U=\beta_1 r+\beta_2 i+\beta_3 s+\beta_4 v+\beta_5 j, \tag{13}$$

where r – delegated reasoning, i – outsourced interpretation, s – synthetic argument construction, v – unverifiable epistemic contribution, j – displaced judgment.

With:

$$\beta_1+\beta_2+\beta_3+\beta_4+\beta_5=1, \beta_i \geq 0. \tag{14}$$

The variable Ucaptures the extent to which substantive intellectual work remains insufficiently governed despite formal policy compliance.

To model the coexistence of visible compliance and hidden substitution, the framework introduces a synthetic compliance effect:

$$S=C \cdot U. \tag{15}$$

This term is especially important because it captures the specific condition in which the policy appears successful at the procedural level while epistemic substitution continues underneath. If either compliance is low or hidden substitution is absent, Sremains low. But when both formal compliance and hidden epistemic substitution are high, the synthetic compliance effect becomes strong.

The net epistemic policy outcome may then be defined as:

$$O=C-\lambda S, \tag{16}$$

or equivalently,

$$O=C(1-\lambda U), \tag{17}$$

where $\lambda \in [0,1]$ is the penalty coefficient that expresses how strongly hidden substitution undermines the substantive value of formal compliance.

Finally, the policy-to-outcome gap is defined as the difference between formal compliance and the net epistemic outcome:

$$\Delta = C - O, \tag{18}$$

which, by substitution, yields

$$\Delta = \lambda CU. \tag{19}$$

This expression makes the framework's logic explicit. The gap widens not only when policy is weak, but also when formal compliance is high while hidden epistemic substitution remains high. In other words, the most problematic condition is not the absence of regulation, but successful procedural compliance combined with weak substantive governance.

This model is analytically distinct from the Transparency–Integrity matrix. The matrix classifies outcomes according to two dimensions, whereas the present framework explains the mechanism of policy failure as a sequence of transformations: policy coverage → formal compliance → hidden epistemic substitution → reduced substantive outcome. It therefore functions not as a second version of the matrix, but as its causal complement.

A useful interpretive implication follows directly from equation (19). If a policy increases C without reducing U, then the gap Δ may grow larger rather than smaller. This means that greater procedural transparency does not necessarily imply better governance. Under some conditions, it may simply produce more visible compliance around practices that continue to undermine research integrity at the level of reasoning, interpretation, and judgment.

6. Discussion

Transparency-oriented policies are valuable because they make AI use legible to institutions, journals, and readers. They create a procedural basis for accountability and reduce complete opacity. However, the models developed here show that legibility is not equivalent to integrity. A researcher may comply fully with disclosure requirements while still offloading interpretation, argument construction, or epistemic judgment to AI systems. In such cases, transparency does not prevent integrity erosion; it may merely document it.

The Transparency–Integrity matrix clarifies this point by distinguishing between visible AI use and epistemically responsible AI use. The most analytically important result is the identification of the high-transparency / low-integrity quadrant. This configuration demonstrates that policy success at the level of disclosure may coexist with failure at the level of knowledge production. Put differently, the core weakness of transparency-centered governance is not that it leaves AI use invisible, but that it can normalize a mistaken equivalence between disclosure and responsibility.

The Policy-to-Outcome Gap Framework deepens this interpretation by introducing a causal logic. It shows that AI policies are typically strongest where practices are visible, documentable, and administratively manageable. As a consequence, they regulate disclosure statements, authorship declarations, tool acknowledgements, and method reporting more effectively than they regulate delegated reasoning, outsourced interpretation, or displaced judgment. This asymmetry creates a structural gap between what policy can observe and what integrity actually depends on. The formal result $\Delta = \lambda CU$ makes this point explicit: the policy-to-outcome gap widens precisely when formal compliance is high and hidden epistemic substitution remains high. This is a particularly important finding because it suggests that stronger disclosure rules may, under some conditions, increase the visibility of compliance without meaningfully improving the epistemic quality of research.

The implications are both theoretical and practical. Theoretically, the study challenges the assumption that transparency is a sufficient proxy for integrity in AI-mediated science. It suggests that responsible AI governance must move beyond documentation and toward a more demanding account of epistemic agency. Practically, the study implies that institutions and journals should not treat disclosure statements as stand-alone evidence of responsible use. Policies must also address the conditions under which researchers preserve or lose control over reasoning, interpretation, and conceptual synthesis. This may require stronger authorship standards, more explicit process documentation, clearer limits on epistemically substantive delegation, and, in some settings, mechanisms to verify that the human researcher remains accountable for the core intellectual content of the work.

At the same time, the study has clear limitations. It is conceptual rather than empirical, and its mathematical formalizations are intended as explanatory tools rather than calibrated predictive models. The variables proposed in the matrix and the gap framework still require operational coding protocols for real-world application. In addition, disciplinary variation is likely to matter: some fields may tolerate certain AI-mediated practices more readily than others, while the meaning of integrity may differ depending on whether the relevant task is exploratory, computational, interpretive, or editorial. These limitations do not weaken the argument, but they do indicate that the present study should be read as a foundation for future empirical refinement rather than as a final measurement system.

Even with these limits, the models developed here make one broad point difficult to avoid: policies that stop at disclosure risk institutionalizing a form of synthetic compliance. In such a system, what matters most is not whether AI was used, but whether its use can be described in an administratively acceptable way. That shift is dangerous because it may preserve the appearance of responsible research governance while leaving the epistemically decisive layer of scientific work weakly protected.

Declaration on Generative AI

The authors acknowledge the use of Gemini (https://gemini.google.com/), ChatGPT (https://chatgpt.com/) and Grammarly (https://app.grammarly.com/) to refine the academic language, accuracy of the text, check and improve the quality of English in this article. The image was generated by ChatGPT 5.2 from the author's original material. After using this tool/service, the authors reviewed and edited the content as needed and took full responsibility for the publication's content.

References

[1] D. Kaczorowska-Spychalska, N. Kotula, G. Mazurek, Ł. Sułkowski. Generative AI as source of change of knowledge management paradigm. Human Technology 20(1) (2024) 131–154. doi:10.14254/1795-6889.2024.20-1.7.

[2] H. Yarovenko, A. Kuzior, T. Norek, A. Lopatka. The future of artificial intelligence: Fear, hope or indifference? Human Technology 20(3) (2024) 611–639. doi:10.14254/1795-6889.2024.20-3.10.

[3] P. Napieralski, A. Wojciechowski, K. Korjonen-Kuusipuro. Artificial intelligence for scientific research – large language models in manuscript preparation. Human Technology 20(3) (2024) 416–419. doi:10.14254/1795-6889.2024.20-3.0.

[4] J. Pilny, K. Postrzednik-Lotko. The Use of New Technologies and Artificial Intelligence in Language Learning and Corporate Discourse. Knowledge Economy and Lifelong Learning 1(2) (2025) 50–76. doi:10.61093/kell.1(2).50-76.2025.

[5] R. Olszewski, D. Krawczyk. (2026). Impact of Artificial Intelligence on the Book Industry: A Path Toward the Knowledge Economy. Knowledge Economy and Lifelong Learning 2(1) (2026) 67–98. doi:10.61093/kell.2(1).67-98.2026.

[6] M. Škare, M. Porada-Rochon, R. Veselica Celić. Total factor productivity dynamics and the artificial intelligence paradox: Evidence from longmemory analysis. Journal of International Studies, 18(4) (2025) 218–242. doi:10.14254/2071- 8330.2025/18-4/11.
[7] Zh. Temerbulatova, A. Zhidebekkyzy, R. Sagiyeva, A. Ludwiczak. Artificial intelligence as a driver of innovation and patent activity: An empirical analysis of cross-country data. Economics and Sociology, 18(3) (2025) 184-201. doi:10.14254/2071-789X.2025/18-3/11.
[8] H. Nguyen, P. Le, X. Le. (2026). From AI disclosure to purchase intention: The mediating role of trust and the moderating effect of collectivistic orientation in Vietnam. Innovative Marketing 22(1) (2026) 274–284. doi:10.21511/im.22(1).2026.20.
[9] B. Mujtaba. Clarifying Ethical Dilemmas in Sharpening Students' Artificial Intelligence Proficiency: Dispelling Myths About Using AI Tools in Higher Education. Business Ethics and Leadership, 8(2) (2024) 107–127. doi:10.61093/bel.8(2).107-127.2024.
[10] R. Kettaf, K. Karima, D. Zohra. Investigating the Impact of Organizational Climate on Organizational Silence in Higher Education Institutions. SocioEconomic Challenges 8(1) (2024) 170–182. doi:10.61093/sec.8(1).170-182.2024.
[11] A. Bappayo, H. Auwal, U. Disa. Responsible Journalism and Sustainable Development: Ethical Challenges in Reporting Sensitive Social Issues. Green Transition and Sustainable Development 2(1) (2026) 88–99. doi:10.61093/gtsd.2(1).88-99.2026.
[12] D. Springs. Smart Leadership Adaptations in Artificial Intelligence Dual Use Functions and Ethical Implications in Healthcare Organizations and Biotechnology Innovations. Health Economics and Management Review 6(3) (2025) 28–37. doi:10.61093/hem.2025.3-03.
[13] M. Roba, O. Moulay. Risk Management in Using Artificial Neural Networks. SocioEconomic Challenges 8(2) (2024) 302–313. doi:10.61093/sec.8(2).302-313.2024.
[14] M. Sitnicki, Y. Petrushenko, E. Kašťáková, A. Luptáková, M. Filipova, V. Hurochkina, I. Volk. Open Science and Open Access Policies as Drivers of Innovation: Evidence from Patent Activity in European Countries. Business Ethics and Leadership 9(2) (2025) 174–187. doi:10.61093/bel.9(2).174-187.2025.