

AI Disclosure without Accountability: Paper Compliance and the Governance Limits of Transparency in Scientific Research

Victor Frimpong¹

¹ Management Department, SBS Swiss Business School, Kloten-Zurich, Switzerland

Correspondence: Victor Frimpong, Management Department, SBS Swiss Business School, Kloten-Zurich, Switzerland.

Received: April 27, 2026

Accepted: June 2, 2026

Available online: June 15, 2026

doi:10.11114/ijsss.v14i3.8747

URL: <https://doi.org/10.11114/ijsss.v14i3.8747>

Abstract

This paper argues that institutionalising artificial intelligence (AI) disclosure in scientific research has led to a form of compliance that prioritises symbolic transparency over genuine accountability. Although journals and publishers are progressively mandating that authors disclose their use of AI, these policies remain fragmented, non-standardised, and largely unverifiable. Drawing on an exploratory review of 80 recent academic articles, the study identifies patterns suggesting that explicit AI disclosure remains limited and, when present, is primarily symbolic or narrative rather than verifiable. The paper explains these patterns by introducing the concept of paper compliance and the AI Disclosure Integrity Gap (AIDG), defined as the discrepancy between reported AI use and the actual epistemic influence of AI on research outputs. The analysis argues that this gap emerges from a misalignment between disclosure-oriented governance frameworks and the iterative, opaque, and often irreproducible characteristics of AI-assisted knowledge production. The study develops testable propositions and introduces the AI Use Traceability Framework (AUTF), a process-oriented governance model that emphasises traceability, auditability, and transparency. While recognising the practical, technical, and institutional constraints associated with implementing traceability mechanisms, the framework outlines proportionate pathways for strengthening accountability in AI-assisted research. The paper contributes to AI governance and research integrity by positioning disclosure as a necessary but insufficient mechanism and by highlighting the risk that current practices may create a false sense of transparency.

Keywords: AI disclosure, AI governance, research integrity, paper compliance, AI disclosure integrity gap, traceability, transparency, accountability, generative AI, scientific publishing

1. Introduction

1.1 Introduce the Problem

The swift integration of artificial intelligence (AI), particularly generative AI, is transforming the processes of scientific knowledge creation and dissemination. This evolution presents significant structural challenges for transparency, authorship attribution, and epistemic accountability in research (Stokel-Walker & Van Noorden, 2023; UNESCO, 2022). AI is now commonly used for literature reviews, hypothesis development, data analysis, and manuscript writing, which is redefining authorship roles. In response, academic publishers, journals, and professional organisations are implementing governance frameworks to uphold research integrity. A key part of these efforts is the requirement for authors to disclose the use of AI tools in their research and writing (Elsevier, 2025; Yoo, 2025).

Major publishing platforms such as Elsevier, Springer Nature, Wiley, and Taylor & Francis now require authors to disclose the AI tools they used, their purpose, and the level of human oversight involved. This information is often included in specific sections of the manuscript (Elsevier, 2025, Taylor & Francis. (2024). Key principles established by leading journals and editorial bodies state that AI tools must be disclosed transparently, cannot be listed as authors, and do not hold responsibility for the scientific output, which rests solely with human researchers (Stone, 2025). This convergence indicates that AI disclosure is becoming a key part of transparency in modern research governance.

The increase in disclosure requirements has not been accompanied by advancements in defining, standardising, or verifying these disclosures. Guidelines differ widely across journals and fields, with no agreement on what constitutes meaningful disclosure, the required level of detail, or how compliance is evaluated (Cleland et al., 2025). Additionally, evidence shows a significant gap between policy and practice; even with widespread disclosure mandates, the actual rates of AI disclosure in published research are very low, highlighting a disconnect between expected and actual

transparency (He & Bu, 2025).

The current approach to AI transparency, primarily through disclosure statements, fails to ensure accountability (Suchikova et al., 2025). These statements are often self-reported, non-standardised, and lack independent verification, offering limited insight into how AI influences research outcomes. They do not facilitate the reproducibility of AI-assisted processes, such as prompt design or model impact analysis. Consequently, disclosure acts more as a compliance signal than a genuine oversight mechanism.

This paper argues that the institutionalisation of AI disclosure has led to “paper compliance”, a superficial approach to governance where meeting transparency requirements replaces real accountability. Authors often provide minimal or vague disclosures under paper compliance that meet editorial standards but do not accurately represent the role or impact of AI in their research. Consequently, transparency becomes more about appearances than actual clarity.

The introduced AI Disclosure Integrity Gap (AIDG) serves as a metric for assessing the discrepancy between the reported use of AI technologies and their actual impact on knowledge production. The AIDG highlights issues such as under-disclosure, oversimplification of AI’s role, and the challenges in verifying AI-assisted processes afterwards. This gap indicates not only inconsistencies in behaviour but also a key limitation of current disclosure systems: they track AI use but lack validation mechanisms.

We argue that disclosure-based governance represents a structurally inadequate paradigm under conditions of AI-induced epistemic opacity, thereby contributing to the literature on AI in scientific practice in three key ways. First, it introduces the concept of paper compliance to understand the performative aspects of AI governance in research. Second, it introduces the AI Disclosure Integrity Gap (AIDG) as a tool to analyse the gap between formal transparency and its actual impact on knowledge. Third, it advocates moving from disclosure-based models to stronger frameworks that emphasise traceability, auditability, and verifiable accountability.

The paper challenges the common belief that simply disclosing AI involvement in research is enough to ensure integrity. It argues that without proper mechanisms to verify these disclosures, we risk creating a false sense of transparency that obscures AI’s actual role in scientific research.

2. The Emergence and Fragmentation of AI Disclosure Regimes

The rapid rise of artificial intelligence (AI) tools in scientific research has prompted major academic publishers, including Elsevier, Springer Nature, Wiley, and Taylor & Francis, to adopt new disclosure policies. These guidelines require authors to reveal the use of generative AI in their manuscripts. This push for transparency aligns with global efforts to ensure accountability in AI systems (OECD, 2019; UNESCO, 2022). Policies emphasise the need to disclose AI’s role in tasks such as text generation, data analysis, and visualisation, while stating that human authors are still responsible for the integrity of their work (Elsevier, 2024; Wiley, 2024).

Editorial organisations like the Committee on Publication Ethics state that AI tools cannot be listed as authors and must be transparently documented to ensure accountability in scholarly communication (COPE, 2023). This reflects a growing expectation for AI disclosure in research governance. However, AI disclosure implementation is inconsistent. Policies differ significantly in detail and enforcement across journals and disciplines. Some journals require comprehensive descriptions of AI use, while others only ask for minimal statements or general guidance. Often, disclosure is merely recommended, leading to variations in reporting practices and editorial oversight (Stokel-Walker & Van Noorden, 2023).

Disclosure regimes face a significant limitation due to the lack of standardised reporting frameworks for AI use in research. There is no widely accepted format for documenting key aspects like prompt design, iteration processes, model versioning, and AI’s impact on analytical outcomes. This results in disclosures that are often non-comparable and provide little insight into AI’s role in knowledge production (Dwivedi et al., 2023; Kasneci et al., 2023).

Current policies rely heavily on self-reporting, with minimal verification or auditing processes. Authors are expected to disclose their use of AI when evaluating the accuracy or completeness of these disclosures. This reliance on unverified information creates vulnerabilities in the governance framework, especially given the increasing sophistication and invisibility of AI processes (Bender et al., 2021; Floridi et al., 2018).

Recent evidence shows a significant gap between AI policy and practice. Although disclosure requirements are widespread, actual rates of AI disclosure in research are low, indicating inconsistent compliance. This issue arises from unclear policy expectations and a lack of enforcement mechanisms to catch non-disclosure (Liang et al., 2025). Studies in academic settings also reveal high levels of non-disclosure, demonstrating that formal mandates alone don’t guarantee transparency (Perkins et al., 2024).

The current policy landscape in general presents a paradox: while AI disclosure aims to enhance transparency and

uphold research integrity, its implementation suffers from fragmentation, weak standardisation, and limited verifiability. Instead of ensuring accountability, existing disclosure practices allow for minimal or vague reporting to count as compliance.

The current governance structure fosters a culture of paper compliance. Without standardised disclosure requirements, verification, and enforcement, organisations are more likely to focus on meeting formalities rather than achieving genuine transparency. This policy environment not only fails to discourage superficial compliance but actually enables it.

2.1 Exploratory Probe: Disclosure Practices in AI-Assisted Research

2.1.1 Data and Sampling Approach

An exploratory review of 80 academic articles published between 2024 and 2026 was conducted to examine current AI disclosure practices. The articles were selected from journals with publicly stated AI disclosure policies and were indexed in Scopus and Web of Science. This review aimed to identify common disclosure patterns and evaluate alignment with the concepts discussed in this paper, without attempting to estimate the overall prevalence of AI disclosure in scientific literature.

Articles were selected through purposive sampling based on three criteria: (1) publication in journals with clear AI disclosure requirements; (2) relevance to generative AI tools based on topic, methodology, or publication period; and (3) availability of full disclosure statements. The sample includes various disciplines such as business, medicine, computer science, and social sciences.

The authors used a structured classification framework to categorise disclosures into four types: no disclosure, symbolic disclosure, narrative disclosure, and traceable disclosure. Classification decisions were based on predefined criteria related to specificity, detail, and traceability. Examples of each category can be found in Appendix A. As this review was exploratory and aimed at building theory, formal inter-coder reliability testing was not conducted. Therefore, the findings should be viewed as illustrative evidence for conceptual development rather than precise estimates of disclosure prevalence in academic publishing.

2.1.2 Prevalence of AI Disclosure

Only 14 of the 80 papers included explicit AI disclosure, whereas 66 did not provide any disclosure statement. This aligns with growing evidence that the use of AI in scientific research is often under-disclosed (Liang et al., 2025), highlighting concerns in governance and editorial policy discussions. Table 1 summarises the prevalence and variation in AI disclosure observed in the sample.

Table 1. Prevalence and Depth of AI Disclosure (Exploratory Sample) ($N = 80$)

Category	Description	Number of Papers	Percentage
No Disclosure	No mention of AI use despite likely applicability	66	82.5%
Symbolic Disclosure	Minimal, generic statement (e.g., "AI used for language editing")	8	10.0%
Narrative Disclosure	Expanded description of AI use, but non-standardised and non-verifiable	5	6.25%
Traceable Disclosure	Structured, process-oriented disclosure enabling partial reconstruction	1	1.25%
Total		80	100%

Note: Classification is based on the content of disclosures. Categories indicate increasing levels of detail and potential for verification.

The percentages in Table 1 should be interpreted with caution. They reflect patterns from an exploratory sample, not overall disclosure rates in the population. The analysis aims to showcase variation in AI disclosure practices rather than establish their prevalence across all scientific publications, highlighting issues of paper compliance and the AI Disclosure Integrity Gap (AIDG).

The Table shows that although AI disclosure policies are common, meaningful disclosure remains uncommon. The prevalence of non-disclosure and symbolic reporting highlights the issue of paper compliance and the ongoing AIDG. It also shows two main patterns: non-disclosure is common despite institutional requirements, and most reported disclosures are superficial and symbolic.

2.1.3 Depth Variation in Disclosure

The 14 disclosed papers showed considerable variation in both depth and informational value, as shown in Table 2.

Table 2. Classification of disclosure papers ($N = 14$)

Type of Disclosure	Description	Frequency
Symbolic disclosure	Minimal, generic statement with no detail	8
Narrative disclosure	Expanded description of AI use, but non-standardised	5
Traceable disclosure	Structured, process-oriented detail (rare)	1

The distribution of disclosure types in AI reporting shows a clear hierarchy in depth and value. Symbolic disclosure (57%) is the most common, featuring vague statements that acknowledge AI use without details on tools or contributions, serving mainly as a compliance signal. Narrative disclosure (36%) offers more detail, with authors describing AI use and mentioning specific tools, but these accounts are self-reported and unverified, lacking true transparency. Conversely, traceable disclosure (7%) is rare, involving some documentation such as model identification and workflows, but remains underdeveloped and allows only limited reconstruction of AI involvement.

The distribution shows that disclosure is on the rise but remains mostly at low levels, revealing only slight progress towards traceability and verifiable accountability.

Appendix A includes examples of these disclosure types to show differences in depth and traceability.

2.1.4 Evidence of Paper Compliance and Implications for the AIDG

The observed patterns offer a clear empirical illustration of paper compliance:

- High non-disclosure rate (82.5%) indicates weak enforcement.
- Dominance of symbolic disclosures (57%) shows minimal compliance strategies.
- Lack of standardisation hinders study comparison.

Even where disclosure is present, it:

- Does not enable verification.
- Fails to represent epistemic influence.
- Lacks support for reproducibility.

This confirms that disclosure often serves as a formal indication of compliance instead of a means of ensuring accountability.

The findings also provide evidence for the AI Disclosure Integrity Gap in three clear ways:

1. Non-disclosure despite probable AI involvement (concealed AI influence).
2. Under-specified disclosures (misrepresentation of extent).
3. Absence of process documentation (irretrievable knowledge pathways).

The empirical probe reveals a clear gap between reported AI use and its actual impact, as evidenced by current publication practices.

Even though AI disclosure policies are becoming more common, most existing literature continues to focus on transparency-based governance models. These models assume that simply disclosing AI usage is enough for accountability. This is evident in editorial and policy guidelines that emphasise reporting AI use (e.g., COPE, 2023; Elsevier, 2024; Wiley, 2024). However, these approaches depend on conditions that are no longer practical in AI-driven research. They assume that (i) research processes can be easily observed and reconstructed, (ii) the role of tools can be clearly defined, and (iii) disclosure statements accurately reflect the underlying research activities.

Recent studies on large language models and AI systems reveal their lack of transparency, unpredictability, and complex interaction dynamics (Bender et al., 2021; Floridi et al., 2018; Dwivedi et al., 2023). However, current disclosure frameworks do not adequately address these issues, leading to a gap between governance practices and the realities of AI-driven knowledge production. While the existing literature highlights risks related to transparency and accountability, it fails to offer a clear framework for understanding how disclosure may become inadequate or merely performative amid epistemic opacity.

3. Paper Compliance in AI Disclosures

To overcome this limitation, we need to move beyond transparency as the main principle of AI governance in research. While institutional theory helps us understand the gap between formal compliance and actual practices (Meyer & Rowan, 1977; Power, 2003), these theories were developed in environments where organisational processes are generally observable. However, AI-assisted research introduces a new challenge: epistemic opacity, in which knowledge-generating processes are often misrepresented and sometimes irrecoverable. This shift indicates that the

issue is not only about incomplete disclosure but that transparency, as a governance tool, may fundamentally fail to address AI's role in knowledge creation.

We redefine AI disclosure as more than just a transparency issue; it's a governance failure stemming from the gap between what is reported and the complex processes underlying knowledge production. We introduce the concepts of paper compliance and the AIDG to highlight this structural misalignment.

3.1 Defining Paper Compliance in AI Disclosure

The paper defines Paper Compliance as:

The formal fulfilment of governance or reporting requirements through symbolic or minimal adherence, without corresponding mechanisms for verification, accountability, or substantive transparency.

In AI disclosure, paper compliance occurs when authors make statements that meet editorial or institutional requirements but fail to accurately reflect the role of AI in their research. These often consist of short, standardised declarations, such as "AI tools were used for language editing," which are hard to assess or validate.

This issue is not limited to AI; it mirrors a wider trend in regulatory and organisational settings where compliance actions diverge from their original purposes, functioning instead as legitimacy signals (Power, 2003; Meyer & Rowan, 1977). In AI disclosure, the problem is heightened by the complexity and unpredictability of AI processes, making it challenging to spot superficial compliance.

3.2 Layers of Paper Compliance

AI disclosure compliance involves multiple layers that show how disclosure requirements can be met without improving accountability. Three separate, but overlapping layers are identified:

1) Symbolic Compliance

Disclosure often serves as a simple checkbox exercise, where authors acknowledge AI use in vague or minimal terms. These statements typically:

- Lacks specificity concerning tools, prompts, or outputs.
- Do not distinguish between levels of AI involvement.
- Mainly serve to indicate compliance with policy.

Symbolic compliance turns disclosure into a mere ritual of transparency, focusing more on having a statement than on its actual meaning.

2) Narrative Compliance

At a more advanced level, authors provide more detailed descriptions of AI use in their methods or acknowledgements sections. These narratives are more detailed than basic disclosures. Nonetheless:

- They remain self-reported and non-standardised.
- They cannot be independently verified, and
- AI is often framed as unimportant, despite its real influence.

Narrative compliance creates an illusion of depth but does not enhance traceability or reproducibility.

3) Strategic Compliance

At a complex level, disclosure is strategically adjusted in response to perceived risks, incentives, or norms. Authors may:

- Minimise the acknowledgement of AI involvement to prevent scrutiny or stigmatisation.
- Exaggerate minimal usage to indicate transparency.
- Selectively disclose specific applications while omitting others.

Strategic compliance means that disclosure is controlled to prioritise reputation over accurate information.

3.3 Mechanisms Enabling Paper Compliance

The continued reliance on paper compliance in AI disclosure is not just a matter of behaviour; it is supported by the structure of existing governance frameworks. Three key mechanisms contribute to this issue:

a) Self-Reporting Without Verification

AI disclosure regimes depend mainly on author declarations without effective verification of:

- Whether AI was utilised
- How it was applied
- The extent of its impact

AI disclosures lack enforceable audit trails, making it impossible to verify compliance, unlike data availability statements or methodological appendices.

b) Absence of Standardisation

To date, there is limited convergence around a standardised taxonomy or schema for reporting AI use in research. As a result:

- Disclosure content varies significantly across different journals and academic disciplines.
- Key dimensions such as prompt design, model version, and iteration processes are seldom documented.
- Comparability between studies is limited.

The lack of standardisation allows minimal compliance to be interpreted as adequate.

c) Epistemic Opacity of AI Systems

AI systems, especially large language models, operate through intricate, opaque processes that are hard to reverse-engineer afterwards, even with complete transparency. Even with full disclosure:

- Outputs cannot be reliably replicated.
- Intermediate steps, like prompt refinement, are often not documented.
- Model behaviour may change over time.

The lack of clarity reduces the connection between disclosure and verifiability, which supports the idea of paper compliance.

3.4 Paper Compliance as Performative Transparency

AI disclosure is evolving into a form of performative transparency, where it appears open but does not allow for real scrutiny. This type of transparency is more about checking boxes than providing actionable insights. Key implications of this shift include:

- Compliance is now judged by the mere existence of disclosures rather than their actual content.
- Reviewers and readers are forced to interpret incomplete and inconsistent information regarding AI's role.
- Disclosure can end up confusing rather than clarifying the basis of research.

The problems with paper compliance are not just about implementation failures; they reveal a deeper disconnect between disclosure governance and the actual use of AI in knowledge production.

3.5 Paper Compliance and the AI Disclosure Integrity Gap (AIDG)

Paper compliance serves as the foundation for the AIDG discussed in the following section. If disclosure is mainly a performative signal, then the gap between stated AI use and its actual impact is a structural result, not an anomaly. Thus, paper compliance is not just a description; it is the mechanism that creates and sustains the AIDG. Table 3 summarises the distinctions between these constructs and related frameworks.

Table 3. Concept Differentiation and Boundary Clarification

Construct	Superficially Similar To	Key Difference	Why Existing Frameworks Are Insufficient
Paper Compliance	Symbolic compliance; Institutional decoupling (Meyer & Rowan, 1977; Power, 2003)	Focuses on epistemic representation in AI-assisted research, where disclosure not only signals compliance but obscures the role of AI in knowledge production	Traditional frameworks assume underlying processes remain intact but misaligned; they do not account for AI-induced opacity, where processes themselves become partially irrecoverable
AI Disclosure Integrity Gap (AIDG)	Reporting bias; Measurement error	Captures structural divergence between declared AI use and actual epistemic influence, not just inaccuracies in reporting	Conventional approaches assume observable and verifiable underlying processes; AIDG arises because AI processes are iterative, distributed, and non-reconstructable ex post
AI Use	Reproducibility	Introduces process-level	Existing reproducibility models assume stable

Traceability Framework (AUTF)	frameworks; Data provenance	traceability across input, interaction, model, and output stages	inputs and deterministic outputs, which do not hold in AI-assisted research environments
-------------------------------	-----------------------------	--	--

The Table clarifies the analytical boundaries of the paper's core constructs by distinguishing them from adjacent concepts in institutional theory, research integrity, and reproducibility literature. This differentiation establishes that the proposed constructs address governance challenges specific to AI-mediated knowledge production that existing frameworks do not capture.

4. The AI Disclosure Integrity Gap (AIDG)

AI disclosure regimes are being established, but they have significant limitations that result in superficial compliance. This section introduces the concept of the AIDG, which emphasises the discrepancy between reported information about AI's involvement in research and its actual influence on knowledge production.

We define the AI Disclosure Integrity Gap as:

The difference between the declared extent of AI use in a research output and the actual epistemic influence exerted by AI on that output.

This definition emphasises the importance of disclosure in terms of its epistemic impact, meaning how AI systems influence the content, structure, interpretation, or conclusions of research. The AIDG extends beyond intentional misrepresentation; it highlights a broader issue where disclosure frameworks fail to accurately capture or verify AI's role in research.

Current disclosure practices have several shortcomings: they are declarative (relying on self-reports), static (providing only snapshots), and non-verifiable (lacking audit mechanisms). In contrast, AI-assisted research is iterative (involving multiple prompt-response cycles), distributed (integrated throughout the research process), and opaque (hard to reconstruct afterwards).

The AIDG highlights a disconnect between the governance of AI use and its actual application in research.

The AIDG can be summarised conceptually as:

$$AIDG = f(D - A)$$

Where:

- D = Declared AI use (as reported in disclosure statements)
- A = Actual AI influence (epistemic contribution to the research output)

The function $f(\cdot)$ captures the multidimensional aspects of the gap, including the extent, function, and process dimensions. Such as:

- Linguistic patterns of AI-generated text
- Inconsistencies between methods and outputs
- Absence of traceable research processes

This formulation identifies the AIDG as a latent governance risk rather than a directly measurable variable in its current state.

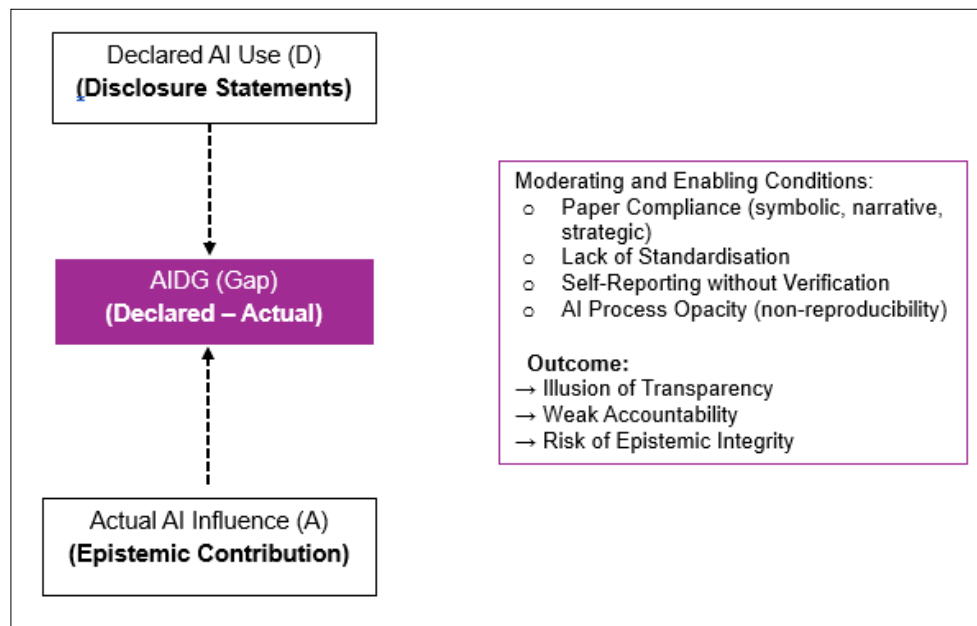


Figure 1. The AI Disclosure Integrity Gap (AIDG) Framework

Developed by the Author, 2026

Figure 1 illustrates the gap between reported AI usage (D) and its actual impact (A) on research outputs, referred to as the AIDG. This gap arises from compliance issues, lack of standardisation, reliance on self-reporting, and unclear AI processes, leading to a false sense of transparency that threatens accountability and integrity in scientific research.

4.1 From Transparency to Integrity Risks

The AIDG alters our understanding of transparency in AI-assisted research. Currently, transparency is seen as a binary choice—either AI use is disclosed, or it is not. However, the AIDG shows that transparency is actually a spectrum, with disclosure being just one aspect of visibility.

This shift highlights a significant risk: the illusion of transparency. When disclosure statements exist, they can create a false sense of accountability without providing verifiable information. Thus, the AIDG reveals that disclosure can turn from a protective measure into a misleading source of trust.

The AIDG arises from the paper compliance dynamics discussed in Section 3. When disclosure is treated as a formality rather than a genuine process, a gap between declared and actual AI use becomes entrenched. Paper compliance acts as the behavioural driver, while the AIDG reflects this dynamic analytically.

4.2 Theoretical Propositions

The paper proposes testable propositions, grounded in the concepts of paper compliance and the AIDG, to guide future research. These propositions convert our theoretical claims into practical relationships and are grounded in existing studies on institutional compliance, epistemic opacity, and AI governance (Meyer & Rowan, 1977; Power, 2003; Bender et al., 2021; Floridi et al., 2018).

P1 — Dominance of Symbolic Compliance

Proposition 1: Without standard reporting and verification, AI disclosures are more symbolic than substantive, process-oriented reporting.

P2 — Incentive-Driven Misrepresentation

Proposition 2: Researchers tend to downplay AI involvement when disclosure harms perceptions of originality, authorship, or credibility.

P3 — Structural Emergence of the AIDG

Proposition 3: The AI Disclosure Integrity Gap widens as the divergence between declaration methods and process-based AI use increases.

P4 — Process Opacity as Core Driver

Proposition 4: Higher process opacity in AI workflows correlates with lower verifiability of disclosure statements and

greater AIDG magnitude.

P5 — Traceability as Corrective Mechanism

Proposition 5: Implementing traceability mechanisms such as logs, model IDs, and interaction records reduces AIDG impact and increases the likelihood of disclosure.

P6 — Illusion of Transparency

Proposition 6: Without traceability and verification, AI disclosure statements boost perceived transparency but do not improve actual epistemic accountability.

Table 4 links core propositions to measurable variables and potential empirical approaches, laying the groundwork for future validation.

Table 4. Operationalisation of Core Theoretical Propositions

Proposition	Key Construct(s)	Measurable Variable(s)	Proxy/Indicator	Empirical Approach (Example)
P1: Symbolic compliance dominance	Disclosure depth; Paper compliance	Disclosure depth score (ordinal: none → symbolic → narrative → traceable)	Content coding of disclosure statements across articles	Manual coding or NLP-based classification of journal publications
P2: Incentive-driven misrepresentation	Perceived reputational risk; Disclosure distortion	Disclosure accuracy vs perceived evaluation risk	Survey responses (e.g., perceived stigma, originality concerns, career risk)	Survey or experimental vignette study
P3: Structural emergence of AIDG	AIDG magnitude	Difference between declared AI use and inferred AI involvement	AI-generated text probability vs disclosure presence	Text analysis (AI detection tools) combined with disclosure comparison
P4: Process opacity as driver	Process opacity; Verifiability	Presence/absence of process documentation	Availability of prompt logs, iteration records, methodological detail	Content analysis of methods sections and Supplementary Materials
P5: Traceability as corrective mechanism	Traceability level; AIDG magnitude	Traceability index vs disclosure gap	Presence of model versioning, prompt logs, output attribution	Regression analysis: traceability → reduction in AIDG
P6: Illusion of transparency	Perceived transparency; Epistemic accountability	Perceived transparency vs actual disclosure quality (AIDG proxy)	Survey ratings vs objectively coded disclosure depth	Experimental study comparing reader perception with measured disclosure quality

The operationalisation focuses on observable indicators that are measured directly or can be inferred from differences between disclosed and actual AI involvement. Proposition 6 highlights a key issue: disclosure can create a false sense of transparency without increasing accountability, thereby increasing governance risks tied to mere compliance.

5. The Limits of Disclosure-Based Governance: Towards Traceability

Disclosure-based governance assumes that transparency can be achieved through author declarations, but this is increasingly questioned in AI-assisted research. In this context, outputs are generated through complex, non-reproducible processes (Bender et al., 2021; Floridi et al., 2018). Consequently, disclosure provides a static view of a dynamic process, overlooking prompt evolution and model variability. Similar challenges exist in broader AI governance discussions, indicating that transparency alone is insufficient without proper verification and accountability mechanisms (Dwivedi et al., 2023).

5.1 From Transparency to Traceability

To address the above-mentioned limitations, we propose shifting from transparency to traceability. Transparency means declaring the use of AI, while traceability allows us to verify and reconstruct how AI contributed to research outcomes.

The importance of documenting the entire lifecycle of knowledge production is emphasised in research on reproducibility and data provenance (Stodden et al., 2016). However, large-scale replication efforts show that transparency alone does not ensure reproducibility, as about half of social science findings do not replicate in controlled studies (Jones, 2026; Nosek et al., 2022).

Traceability differs from disclosure; it focuses on real-time tracking and monitoring rather than simply providing

historical information. It is:

- Process-oriented (captures sequences of interaction)
- Verifiable (allows external scrutiny)
- Reconstructive (enables partial reproducibility)

This shift connects AI governance to established practices such as financial reporting, audit trails, and data lineage tracking in scientific workflows.

5.2 The AI Use Traceability Framework (AUTF)

To operationalise this shift, we introduce AUTF, a model that links declared AI utilisation with actual accountability.

The framework consists of four interrelated components:

(1) Input Traceability: It captures:

- Prompt logs (initial and iterative)
- Task framing and constraints
- Human–AI interaction sequences

Purpose: Show how inputs influence outputs.

(2) Process Traceability: Records and documents AI interaction, such as:

- Iteration History
- Generated and discarded alternative outputs
- Human editing and refinement

Purpose: To highlight the dynamic and non-linear aspects of AI-assisted work, reflecting the iterative nature of computational research workflows (Stodden et al., 2016).

(3) Model Traceability: Specifies the technical configuration of the AI system used, including:

- Model type and version
- Training updates (where relevant)
- Tool-specific parameters

Purpose: Tackle reproducibility issues caused by model variability and system updates (Bender et al., 2021).

(4) Output Attribution: Connects AI contributions to specific sections of the research output.

- Sections Affected by AI
- Differentiation between AI-generated and human-edited content
- Level of reliance on AI outputs

Purpose: Define the epistemic role of AI and address issues of authorship and accountability in AI-assisted research (Dwivedi et al., 2023).

Figure 2 shows the AI Use Traceability Framework (AUTF), which outlines the four interconnected stages for documenting and partially reconstructing AI involvement.

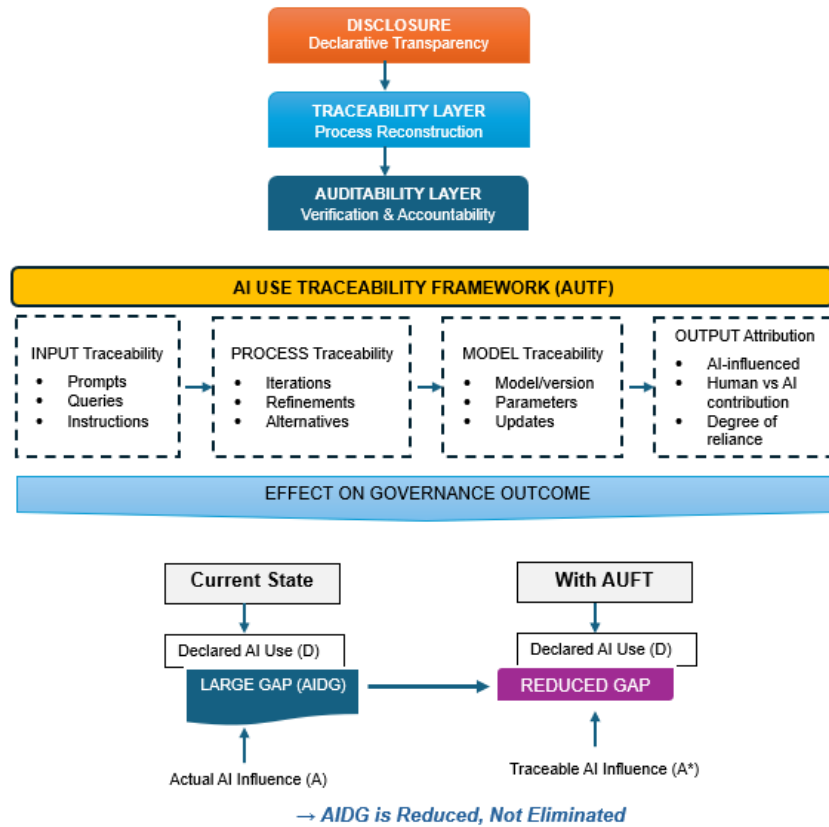


Figure 2. The AI Use Traceability Framework (AUTF)

Developed by the Author (2026)

The Figure illustrates the transition from disclosure-based governance to the prioritisation of traceability and auditability in AI-assisted research. The AUTF identifies four stages of AI involvement: input, process, model, and output, which enables partial reconstruction of AI contributions. This method addresses the AI Disclosure Integrity Gap (AIDG) by improving accountability and ensuring epistemic integrity.

5.3 AUTF as a Governance Layer

The AUTF creates a governance framework centred on three essential components: Disclosure, Traceability, and Auditability. This strategy promotes transparency and accountability, addressing the need for enforceable responsibility in AI ethics beyond just transparency (Floridi et al., 2018).

5.4 Implementation Pathways

AUTF offers a model for improving accountability, but its implementation will vary across different research contexts. The shift from disclosure-based practices to traceability-oriented governance will happen gradually, influenced by institutional capacity, technical feasibility, and incentives. We view this implementation as a staged process rather than a simple transition.

Figure 3 shows the implementation pathway, which transitions from minimal disclosure practices to advanced traceability and auditability. It highlights the evolution of governance mechanisms from symbolic compliance to structured, process-oriented accountability.

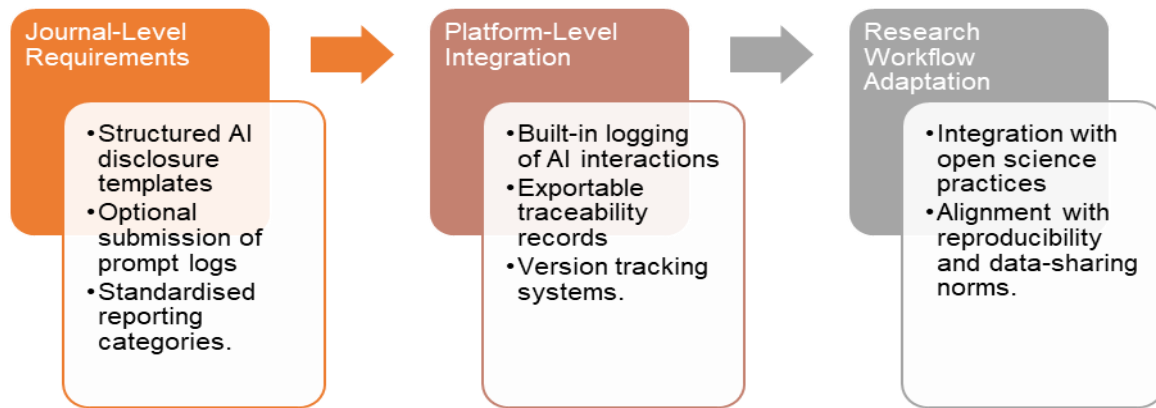


Figure 3. Implementation pathway

Developed by the Author (2026)

The implementation pathway has three key points. First, moving through stages depends on institutional incentives and enforcement, not automatic processes. Second, early interventions, such as structured disclosure templates, help reduce barriers to adoption and facilitate gradual progress. Third, full traceability is more of an ideal goal than a fixed endpoint, constrained by technical issues, privacy concerns, and the unpredictability of AI systems. The pathway stresses the importance of feasibility in governance design, highlighting that accountability should be developed gradually rather than forced all at once.

5.5 Feasibility and Adoption Constraints

The AI Use Traceability Framework (AUTF) aims to enhance accountability, but should not be seen as a call for comprehensive oversight of all AI-assisted research. Complete process traceability can impose high costs that may outweigh benefits in many cases.

Keeping detailed logs of prompts, interactions, and model documentation adds administrative burdens for researchers. Unlike traditional disclosures that can be completed later, traceability requires ongoing documentation, which can be challenging for those with limited time or resources.

Extensive traceability could hinder creativity. Researchers might hesitate to explore AI tools early in the idea development stage if all interactions must be documented and reviewed. Intrusive traceability could stifle the exploratory use of AI, which is valuable in research.

Journals currently lack the capability to efficiently verify detailed AI interaction records. Expecting editors or reviewers to check complete prompt histories for every submission would be impractical and significantly increase costs and review times.

Technical limitations hinder complete traceability. AI systems, particularly large language models, produce non-deterministic outputs and evolve over time, making full reconstruction of processes difficult. Current publishing systems also lack the capability to store process-level data, such as prompt histories, in standardised formats.

Privacy and intellectual property issues further complicate matters. Prompt logs may include sensitive data or proprietary ideas, raising concerns about data exposure. Mandating full traceability without safeguards could pose risks to researchers and institutions.

The AUTF should be viewed as a flexible governance framework rather than a strict requirement. Its purpose is to establish reasonable mechanisms to improve accountability without having to reconstruct every AI interaction.

5.6 Lightweight Verification Pathways

Due to the limitations of comprehensive auditing, journals might adopt lightweight verification methods that ensure accountability while remaining practical.

One method is structured AI disclosure forms. Authors would fill out standardised fields detailing the AI tool used, its purpose, the research stage affected, and the degree of human review. This structured reporting enhances comparability without adding a significant burden on authors or editors.

Another method is optional traceability documentation. Journals could allow authors to voluntarily provide examples of prompts, model identifiers, workflow summaries, or interaction logs, similar to supplementary materials. This wouldn't require routine verification but would enable targeted reviews if issues arise.

A risk-based approach may be more practical than universal auditing. High-impact, technical, or AI-heavy studies could face stricter disclosure requirements, while lower-risk tasks, such as language editing, might require only basic reporting. This model ties accountability to the potential impact of AI on research outcomes.

Contribution to Literature

The paper outlines testable propositions for analysing disclosure practices, the AI Disclosure Integrity Gap (AIDG), and traceability mechanisms. It makes three main contributions to the literature on AI in research and governance: (1) it introduces "paper compliance," which shows how disclosures in AI-assisted research can obscure the true epistemic processes. (2), it defines the AIDG, highlighting the difference between reported AI use and its real epistemic impact, stressing the need to understand the limits of representability in AI-driven knowledge. Finally, it proposes AUTF, a governance model for navigating the complexities of AI-assisted workflows, advancing beyond traditional methods for reproducibility and data provenance.

These contributions, outlined in Table 3, are essential responses to governance challenges posed by artificial intelligence in scientific research, rather than mere extensions of existing frameworks. This paper contributes to theory development by introducing new constructs, defining their boundaries, and proposing relationships among them. The concepts of paper compliance, the AI Disclosure Integrity Gap (AIDG), and the AI Use Traceability Framework (AUTF) aim to provide frameworks for future empirical research rather than being fully validated constructs at this stage (Whetten, 1989; Corley & Gioia, 2011). The propositions set the groundwork for further testing, refinement, and expansion around AI governance and research integrity.

6. Implications

Our findings have significant implications for academic publishing, research integrity, and AI governance in knowledge production.

In academic publishing, the analysis indicates that current AI disclosure requirements can create a misleading sense of transparency. When disclosure is mandatory but not verified, it burdens reviewers and readers with the task of deducing AI's role from incomplete or inconsistent information. Therefore, journals should adopt structured reporting standards and traceable submission processes. This could include optional prompt logs, standardised templates, and appendices detailing AI use.

Second, the introduction of the AI Disclosure Integrity Gap (AIDG) identifies a new type of epistemic risk in research integrity. Even with disclosure, the impact of AI on research output may remain unclear, challenging existing views on authorship, originality, and accountability. This necessitates a reassessment of how credibility is evaluated in research that involves AI.

Thirdly, the paper shows that while disclosure-based methods may seem appealing for policy and governance, they are inadequate without enforceable mechanisms. A shift toward traceability and auditability imposes new responsibilities on publishers, platforms, and research institutions. They must create infrastructure to log AI interactions, develop standardised reporting protocols, and enable partial verification of AI-assisted processes.

Fourth, future research on the AIDG should explore a range of empirical and methodological approaches. Researchers can investigate the gap through linguistic analysis, workflow reconstruction, or experiments comparing disclosed and undisclosed AI use. Additionally, it's important to systematically examine the feasibility, costs, and behavioural impacts of implementing the AUTF.

Limitations

The exploratory review has key limitations. The sample was purposefully selected and does not represent wider scientific literature. Additionally, the disclosure classifications were created for theory-building and were not validated through formal inter-coder reliability checks. Future research should use larger, systematic samples, clear search protocols, and involve multiple coders to more rigorously assess the prevalence and determinants of disclosure practices.

The governance framework in this paper is conceptual and has not been empirically tested in journal environments. While traceability can enhance accountability, it also introduces challenges related to administrative burden, privacy, intellectual property, and researcher autonomy. The feasibility of verification differs widely among disciplines, publication models, and institutions.

The paper, therefore, does not support mandatory logging of all AI interactions. Instead, it suggests using traceability as a flexible governance principle, tailored to the extent of AI involvement and associated risks. Future research should assess the costs, behavioural impacts, and institutional feasibility of different verification methods.

7. Conclusions

The growing institutionalisation of AI disclosure in scientific research indicates that artificial intelligence is changing knowledge production. However, current disclosure practices are inadequate for ensuring true accountability. Instead of fostering real transparency, they often lead to “paper compliance,” where formal adherence replaces effective oversight.

This study introduces the concepts of paper compliance and the AI Disclosure Integrity Gap (AIDG), framing AI disclosure as a structural governance issue rather than just an ethical one. It highlights that the gap between reported AI use and its actual impact is a systemic issue stemming from self-reporting, a lack of standardisation, and the absence of verification mechanisms.

To overcome current limitations, the paper presents AUTF. It emphasises reconstructive, verifiable accountability rather than mere transparency. This framework aims to create governance models that align better with the iterative, distributed, and opaque nature of AI-assisted research.

In conclusion, without accountability, AI disclosure creates a false sense of transparency in scientific knowledge production. The empirical evidence from this study suggests that non-traceable disclosure practices persist, and the proposed theoretical frameworks provide a basis for further investigation into AI disclosure behaviour and governance. Moreover, the challenges of ensuring traceability highlight that moving from disclosure to accountability is not just a technical issue; it’s an institutional process that requires gradual adaptation. As AI becomes more integrated into research, effective governance will hinge on what can be traced and held accountable, rather than just what is stated.

Acknowledgements

Not Applicable.

Author contributions

Not applicable.

Artificial Intelligence (AI) Disclosure

The author used ChatGPT 5.2 solely for language improvement and clarity during editing. The author reviewed all content and takes full responsibility for the manuscript.

Funding

Not applicable.

Competing interests

Not applicable.

Informed consent

Obtained.

Ethics approval

The Publication Ethics Committee of the Redfame Publishing.

The journal’s policies adhere to the Core Practices established by the Committee on Publication Ethics (COPE).

Provenance and peer review

Not commissioned; externally double-blind peer reviewed.

Data availability statement

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

Data sharing statement

No additional data are available.

Open access

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

References

- Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). *On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? In Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (FAccT'21)* (pp. 610-623). Association for Computing Machinery. <https://doi.org/10.1145/3442188.3445922>
- Cleland, J., Driessen, E., Masters, K., Lingard, L., & Maggio, L. A. (2026). When and how to disclose AI use in academic publishing: AMEE Guide No.192. *Medical Teacher*, 48(4), 542-553. <https://doi.org/10.1080/0142159X.2025.2607513>
- COPE (2023). COPE: *Committee on Publication Ethics*. Retrieved from <https://publicationethics.org/news-opinion/cope-2023>
- Corley, K. G., & Gioia, D. A. (2011). Building Theory about Theory Building: What Constitutes a Theoretical Contribution? *Academy of Management Review*, 36(1), 12-32. <https://doi.org/10.5465/amr.2009.0486>
- Dwivedi, Y. K. (2023). "So What If ChatGPT Wrote it?" Multidisciplinary Perspectives on opportunities, Challenges and Implications of Generative Conversational AI for research, Practice and Policy. *International Journal of Information Management*, 71(0268-4012), 102642. <https://doi.org/10.1016/j.ijinfomgt.2023.102642>
- Elsevier. (2024). [Www.elsevier.com](https://www.elsevier.com). Retrieved from <https://www.elsevier.com/about/policies-and-standards/generative-ai-policies-for-journals>
- Elsevier. (2025). [Www.elsevier.com](https://www.elsevier.com). Retrieved from <https://www.elsevier.com/about/policies-and-standards/generative-ai-policies-for-journals?utm>
- Floridi, L., Cows, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., ... Vayena, E. (2018). AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations. *Minds and Machines*, 28(4), 689-707. <https://doi.org/10.1007/s11023-018-9482-5>
- He, Y., & Bu, Y. (2025). *Academic journals' AI policies fail to curb the surge in AI-assisted academic writing*. ArXiv.org. Retrieved from <https://arxiv.org/abs/2512.06705?utm>
- Jones, N. (2026). *Half of social-science studies fail replication test in years-long project*. *Nature*. <https://doi.org/10.1038/d41586-026-00955-5>
- Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., & Gasser, U., et al. (2023). ChatGPT for good? on Opportunities and Challenges of Large Language Models for Education. *Learning and Individual Differences*, 103(102274). <https://doi.org/10.1016/j.lindif.2023.102274>
- Liang, W., Zhang, Y., Wu, Z., Lepp, H., Ji, W., Zhao, X., ... Zou, J. (2025). *Quantifying large language model usage in scientific papers*. *Nature Human Behaviour*. <https://doi.org/10.1038/s41562-025-02273-8>
- Meyer, J. W., & Rowan, B. (1977). Institutionalised Organisations: Formal Structure as Myth and Ceremony. *American Journal of Sociology*, 83(2), 340-363. Retrieved from <http://www.jstor.org/stable/2778293>
- Nosek, B. A., Hardwicke, T. E., Moshontz, H., Allard, A., Corker, K. S., Dreber, A., ... Vazire, S. (2022). Replicability, Robustness, and Reproducibility in Psychological Science. *Annual Review of Psychology*, 73(1), 719-748. <https://doi.org/10.1146/annurev-psych-020821-114157>
- OECD. (2019). *The OECD Artificial Intelligence (AI) Principles*. [Oecd.ai; OECD. https://oecd.ai/en/ai-principles](https://oecd.ai/en/ai-principles)
- Perkins, M., Furze, L., Roe, J., & MacVaugh, J. (2024). The Artificial Intelligence Assessment Scale (AIAS): A Framework for Ethical Integration of Generative AI in Educational Assessment. *Journal of University Teaching and Learning Practice*, 21(06). <https://doi.org/10.53761/q3azde36>
- Power, M. K. (2003). Auditing and the Production of Legitimacy. *Accounting, Organisations and Society*, 28(4), 379-394. [https://doi.org/10.1016/s0361-3682\(01\)00047-2](https://doi.org/10.1016/s0361-3682(01)00047-2)
- Stodden, V., McNutt, M., Bailey, D. H., Deelman, E., Gil, Y., Hanson, B., Heroux, M. A., Ioannidis, J. P. A., & Taufer, M. (2016). Enhancing reproducibility for computational methods. *Science*, 354(6317), 1240-1241. <https://doi.org/10.1126/science.aah6168>
- Stokel-Walker, C., & Van Noorden, R. (2023). What ChatGPT and generative AI mean for science. *Nature*, 614(7947), 214-216. <https://doi.org/10.1038/d41586-023-00340-6>
- Stone, J. A. M. (2025). AI Disclosure Policies in Scientific Publishing and Medical Acupuncture. *Medical Acupuncture*, 37(5), 339-340. <https://doi.org/10.1177/19336586251384551>

- Suchikova, Y., Tsybuliak, N., Teixeira da Silva, J. A., & Nazarovets, S. (2025). GAIDeT (Generative AI Delegation Taxonomy): A taxonomy for humans to delegate tasks to generative artificial intelligence in scientific research and publishing. *Accountability in Research*, 1-27. <https://doi.org/10.1080/08989621.2025.2544331>
- Taylor & Francis. (2024, August 20). *AI Policy*—Taylor & Francis. Retrieved from <https://taylorandfrancis.com/our-policies/ai-policy/?utm>
- UNESCO. (2022). *Recommendation on the Ethics of Artificial Intelligence*. Unesco.org. Retrieved from <https://unesdoc.unesco.org/ark:/48223/pf0000381137>
- Whetten, D. A. (1989). What Constitutes a Theoretical Contribution? *Academy of Management Review*, 14(4), 490-495. <https://doi.org/10.5465/amr.1989.4308371>
- Wiley. (2024). *AI guidelines for researchers* | Wiley.com. Retrieved from <https://www.wiley.com/en-nl/publish/article/ai-guidelines/>
- Yoo, J. H. (2025). Defining the Boundaries of AI Use in Scientific Writing: A Comparative Review of Editorial Policies. *Journal of Korean Medical Science*, 40(23). <https://doi.org/10.3346/jkms.2025.40.e187>

Appendix A: Illustrative Examples of AI Disclosure Practices

This appendix presents three illustrative examples of AI disclosure practices observed in the exploratory review. The examples demonstrate variation in disclosure depth and traceability and are presented as representative paraphrases rather than verbatim excerpts from individual articles.

Sample 1: Symbolic Disclosure

QUOTE: “AI tools were used to improve the clarity and readability of the manuscript.”

Assessment:

- o No tool identified
- o No indication of scope or sections affected
- o No distinction between editing and generation
- o No process or iteration information

Interpretation: This represents symbolic compliance, where disclosure satisfies formal requirements but provides no insight into the epistemic role of AI.

Sample 2: Narrative Disclosure

QUOTE: “ChatGPT (OpenAI) was used to assist in drafting portions of the introduction and discussion. All generated content was reviewed and edited by the authors to ensure accuracy and originality.”

Assessment:

- o Tool identified
- o General functional role described
- o Human oversight asserted

Interpretation: This represents narrative compliance, where disclosure appears substantive but is neither verifiable nor reproducible.

Sample 3: Partial Traceability Disclosure

QUOTE: “We used GPT-4 (March 2025 version) to generate initial summaries of selected literature. Prompts included structured queries specifying inclusion criteria and thematic focus. Outputs were iteratively refined through follow-up prompts, and final text was substantially modified by the authors. Selected prompt examples are available in the Supplementary Materials.”

Assessment:

- o Model and version specified
- o Prompt structure described
- o Iterative interaction acknowledged
- o Partial documentation provided

Interpretation: This represents early-stage traceability, demonstrating movement toward accountability, but still insufficient for full reconstruction or verification.”

End of sample

These examples show that while disclosures exist, their level of detail varies greatly, and traceability is uncommon. The shift from symbolic to narrative to traceable disclosure indicates a move toward greater transparency, yet it highlights that existing practices do not allow for proper verification. This variation supports the idea of paper compliance and emphasises the presence of the AI Disclosure Integrity Gap (AIDG).