education sciences

MDPI

Article

PETMALU-AI: A Reporting Checklist for Transparent and Accountable Generative AI-Assisted Research Writing

Manuel B. Garcia [1,2,3]

[1] Educational Innovation and Technology Hub, FEU Institute of Technology, Manila 1015, Philippines; manuelgarciaph@yahoo.com or mbgarcia@feutech.edu.ph
[2] College of Education, University of the Philippines, Quezon City 1101, Philippines
[3] Graduate School of Education, Korea University, Seoul 02841, Republic of Korea

Abstract

The rapid integration of artificial intelligence (AI) tools into academic research and writing has introduced new challenges related to transparency, accountability, and reproducibility. While publishers increasingly permit AI use with disclosure, there is currently no standardized reporting framework guiding how such use should be documented. This study introduces PETMALU-AI (Principles for Ensuring Transparency in Machine-assisted Authorship, Logging, and Use of Artificial Intelligence), a reporting checklist designed to support more consistent disclosure and documentation of AI-assisted research practices. The development of PETMALU-AI followed a multi-phase design framework, including (1) a narrative synthesis of existing AI policies and reporting guidelines, (2) item generation grounded in methodological transparency principles, and (3) modified Delphi-based expert validation involving an interdisciplinary panel. The Delphi process produced a 33-item checklist organized across nine domains: disclosure, human accountability, system transparency, prompting and interaction processes, data provenance, validation, ethics, limitations, and reproducibility. All retained items achieved expert consensus, with item-level content validity indices ranging from 0.85 to 1.00 and final agreement rates ranging from 85% to 100%. Reliability analyses further supported the stability of expert judgments, with Fleiss' kappa and the intraclass correlation coefficient indicating strong agreement and high consistency. The checklist emphasizes proportional reporting, human accountability, and transparent documentation of AI-assisted research practices. PETMALU-AI offers a framework for documenting AI involvement in ways that are visible, assessable, and accountable, providing a foundation for more transparent research reporting as scholarly writing becomes increasingly shaped by human–AI collaboration.

Keywords: artificial intelligence; generative AI; research reporting; reporting guideline; research integrity; transparency; academic writing; accountability; AI ethics

Academic Editor: Will W. K. Ma

Received: 26 April 2026
Revised: 10 July 2026
Accepted: 13 July 2026
Published: 15 July 2026

1. Introduction

The integration of artificial intelligence (AI) into research writing has rapidly moved from novelty to normalized scholarly practice. Recent evidence shows that generative AI tools are now being used across multiple stages of academic work, including idea generation, outlining, literature synthesis, drafting, coding, editing, and data-related tasks, largely because they improve speed, reduce cognitive load, and support language production (Dorta-González et al., 2024; Garcia, 2025; Khalifa & Albadawy, 2024; Shimray & Subaveerapandiyan, 2025). At the same time, the literature consistently warns

that these benefits are accompanied by risks involving factual inaccuracy, fabricated citations, hidden bias, and overreliance on machine-generated prose, all of which can compromise the integrity of the research record if left insufficiently documented (Acut et al., 2025; Alduais et al., 2025; Calderon & Herrera, 2025). Scholars have therefore argued that AI use in research should not be treated as a mere writing convenience but as a methodological intervention that can shape how knowledge is produced, represented, and interpreted.

Yet the current governance landscape remains fragmented. Although publishers and professional bodies increasingly require disclosure, their policies remain uneven in scope, terminology, and level of procedural detail. For instance, Elsevier (2025) requires a separate declaration describing the AI tool used, its purpose, and the extent of author oversight, while Wiley (2026) requires authors to disclose AI use and ensure that generated content is properly reviewed and verified. Similarly, Springer Nature (2026) mandates transparency in AI use, and prohibits assigning authorship to AI systems. These policies reflect a growing consensus around accountability but stop short of providing a procedural framework comparable to established reporting standards. Empirical evidence further illustrates the extent of this inconsistency. In a comparative analysis of high-impact cardiovascular journals, Laffaye et al. (2025) found substantial variability in AI-related editorial policies. Among the 17 journals examined, 82% had some form of AI policy. However, only a minority provided clear or specific requirements for disclosure, and none mandated a standardized disclosure section or language. These findings highlight both the lack of standardization in policy design and the disconnect between the existence of policies and their effective implementation in reporting practice. Despite the growing emphasis on disclosure, standardized reporting guidance for AI-assisted research writing remains limited, fragmented, and unevenly applicable across disciplines. This gap underscores the need for a structured reporting framework that can guide both authors and journals in documenting AI involvement systematically and transparently.

Against this backdrop, the present paper introduces PETMALU-AI (Principles for Ensuring Transparency in Machine-assisted Authorship, Logging, and Use of Artificial Intelligence), a reporting checklist for the transparent and accountable documentation of AI-assisted research writing. Drawing on the reporting logic of PRISMA 2020 and the broader tradition of structured reporting guidelines, PETMALU-AI provides a framework through which authors can specify where AI was used, what role it played, who remains accountable, how outputs were validated, and what information readers and reviewers require to evaluate the credibility of AI-assisted scholarship. This work contributes to the field by moving beyond fragmented policy directives toward a coherent reporting instrument capable of promoting more consistent manuscript preparation, strengthening peer review, and enabling future meta-research on the role of AI in academic writing.

2. Materials and Methods

2.1. Study Design

This study employed a reporting-checklist development and validation design to construct PETMALU-AI, a framework for documenting the use of generative AI in research writing. The development process followed established methodological principles for reporting guideline construction, including conceptual grounding, item generation, framework organization, and expert-based validation. The organizational logic of PRISMA 2020 and prior reporting guideline development approaches informed the structure of the checklist, particularly the use of itemized reporting domains to promote transparency, completeness, and reproducibility (Moher et al., 2010; Page et al., 2021).

The study was conducted in four sequential phases, consistent with established guidance for developing reporting guidelines and checklist-based research standards. First, a

targeted narrative synthesis of publisher policies, organizational guidance, and scholarly literature was undertaken to identify recurring concerns related to AI disclosure, authorship accountability, system transparency, prompting, validation, ethics, and reproducibility. Second, candidate checklist items were generated using a deductive-inductive approach. Deductively, items were derived from recurring principles in the literature and policy review; inductively, additional items were developed to address emerging concerns that were not sufficiently captured in existing reporting guidance. Third, the candidate items were organized into domains reflecting the major stages and reporting needs of AI-assisted research writing. Fourth, the preliminary checklist was evaluated through a modified Delphi process involving an interdisciplinary expert panel.

This design was selected because reporting checklists require both conceptual rigor and consensus-based refinement. Conceptual grounding was necessary to ensure that the checklist responded to documented gaps in AI-related reporting practice, while expert validation was necessary to assess the clarity, relevance, and practical interpretability of the proposed items. Consequently, the overall design combined literature- and policy-informed item development with iterative expert judgment (Moher et al., 2010). This approach was intended to produce a checklist that is conceptually grounded, methodologically transparent, and usable across scholarly and editorial contexts.

2.2. Conceptual Grounding and Evidence Base

The initial development phase involved a targeted narrative synthesis of AI-related editorial policies, ethical guidance documents, publisher instructions, reporting standards, and scholarly literature published between 2022 and 2025. Rather than serving as a formal systematic review, this phase was conducted as a structured evidence-mapping exercise to identify recurring principles, unresolved issues, and reporting gaps relevant to AI-assisted research writing. The synthesis was informed by scoping-oriented principles, particularly the mapping of concepts, practices, and policy positions across a rapidly evolving area of scholarly communication (Arksey & O'Malley, 2005). The objective was not to comprehensively evaluate or synthesize empirical evidence, but to establish a conceptually grounded foundation for checklist development by identifying convergent themes across policy, methodological, and scholarly sources.

Searches were conducted in Google Scholar, Scopus, Web of Science, and publisher or organizational policy webpages. Search terms included combinations of "generative AI," "artificial intelligence," "academic writing," "research writing," "research integrity," "AI disclosure," "AI policy," "AI authorship," "AI-assisted writing," "reporting guideline," "transparency," "reproducibility," and "scholarly publishing." Because AI-related editorial guidance often appears outside bibliographic databases, targeted searches were also conducted on the websites of major publishers and scholarly organizations with publicly available policies on AI use, authorship, disclosure, and publication ethics.

Sources were considered relevant if they addressed at least one of the following areas: disclosure of AI use in research or writing, authorship and accountability in AI-assisted scholarship, transparency or reproducibility of AI-assisted workflows, validation or verification of AI-generated outputs, data privacy and confidentiality, bias mitigation, or procedural guidance for reporting research practices. Editorial and policy documents from major publishers and organizations were included when they contained explicit expectations regarding AI use, disclosure, human responsibility, authorship, or compliance. Scholarly sources were included when they addressed AI-assisted authorship, research integrity, methodological transparency, reproducibility, or disclosure practices relevant to academic research writing (Arksey & O'Malley, 2005; Peters et al., 2015).

Records were first screened at the title, abstract, policy-summary, or webpage-summary level to remove sources that were outside the scope of the study. Materials were excluded if they focused solely on AI in education, clinical practice, technical model development, or general AI capabilities without addressing research writing, authorship, disclosure, transparency, accountability, ethics, validation, or reproducibility. Potentially relevant sources were then reviewed in full to determine whether they contributed substantive principles, requirements, concerns, or reporting gaps relevant to checklist development. Duplicate documents, inaccessible materials, opinion pieces without clear relevance to reporting practice, and sources without direct implications for AI-use documentation were excluded (Levac et al., 2010; Peters et al., 2015).

For each retained source, key information was extracted on the type of document, issuing body or publication context, AI-related disclosure expectations, authorship or accountability provisions, treatment of privacy and data protection, validation or verification requirements, bias or error-related concerns, and reproducibility-related guidance. Extracted information was compared across sources to identify convergent principles, inconsistencies, and gaps in existing guidance. This process showed that although many publishers and organizations required some form of AI disclosure, policies differed in terminology, specificity, reporting location, and procedural expectations (Ganjavi et al., 2024). More importantly, no unified checklist was identified that could systematically document AI use across manuscript development, prompting practices, data-related tasks, output verification, ethical safeguards, and reproducibility constraints.

The recurring concerns identified through this synthesis informed the initial domains and candidate items of PETMALU-AI. Publisher and organizational policies primarily informed items related to AI-use disclosure, authorship compliance, human accountability, and policy adherence. Scholarly literature on AI-assisted writing, research integrity, methodological transparency, and reproducibility informed items related to system transparency, prompting and interaction processes, data provenance, validation, limitations, and open practices. Because the purpose of the synthesis was to inform checklist development rather than evaluate policy outcomes, studies focused solely on the implementation, enforcement, or effectiveness of AI disclosure policies were not treated as a primary evidence category unless they directly identified reporting requirements, documentation gaps, or practical barriers relevant to checklist construction. This scope was adopted to maintain alignment between the evidence base and the study objective, which was to develop and validate a reporting framework rather than assess the impact of existing AI policies. Nevertheless, evidence on policy implementation was considered conceptually relevant when it illuminated the gap between policy existence and reporting practice.

2.3. Item Generation and Initial Refinement

Item generation was conducted using a deductive-inductive approach. Deductively, initial candidate items were derived from recurring principles identified in the literature and policy review, including transparency of AI involvement, human accountability, traceability of AI-assisted processes, ethical compliance, validation of AI-generated outputs, and reproducibility of AI-assisted workflows (Moher et al., 2010; Resnik & Hosseini, 2026a)—for example, publisher and editorial policies requiring authors to disclose AI use informed items on AI-use disclosure, scope of AI involvement, and policy compliance. Authorship guidance prohibiting the attribution of authorship to AI systems informed items on human–AI contribution delineation, authorship compliance, and accountability assignment. Literature emphasizing reproducibility and methodological transparency informed items on model or version disclosure, prompting approach description, interaction record availability, materials availability, and reproducibility constraints.

Inductively, additional items were formulated in response to emerging reporting concerns that were not adequately captured in existing policies or conventional reporting frameworks. These included prompt design responsibility, iterative human–AI interaction, output curation, AI-generated or synthetic data, system evolution, and limitations arising from the non-deterministic behavior of generative AI systems. For instance, although many policies required general disclosure of AI use, they rarely specified how authors should document prompt construction, repeated prompt-output cycles, or the selection and rejection of AI-generated outputs. This gap informed items on prompt design responsibility, iterative use disclosure, output curation, and interaction record availability. Similarly, concerns about changing model versions, platform updates, and variable outputs informed items on reproducibility constraints and system evolution disclaimers.

The preliminary item pool was then reviewed iteratively to assess conceptual overlap, wording clarity, scope, and practical applicability. During this internal refinement phase, redundant items were merged, ambiguous items were reworded, and items judged to be overly narrow, duplicative, or insufficiently actionable were revised or removed (Gattrell et al., 2024). Particular attention was given to balancing comprehensiveness with usability because reporting guidelines must capture essential information without imposing unnecessary burden on authors, reviewers, or editors (Moher et al., 2010). The initial development process generated 42 candidate reporting items. During pre-Delphi internal refinement, 9 overlapping or overly narrow candidate items were consolidated into broader reporting items, resulting in 33 checklist items that entered Round 1 of the Delphi process. This process produced the preliminary PETMALU-AI checklist that was subsequently submitted for expert evaluation through the modified Delphi process. The preliminary checklist formed part of a broader research program examining ethical governance and reporting practices for generative AI in scholarly research. Within this program, the present study focused on the systematic refinement and validation of the reporting framework through a modified Delphi process. Although the preliminary framework was evidence-informed, all candidate items underwent independent expert review, iterative revision, and consensus-based validation before inclusion in the final checklist.

2.4. Framework Structuring

Following item generation and initial refinement, the retained checklist components were organized into domains that reflected the logical progression of AI-assisted research writing and the major reporting concerns identified during the conceptual grounding phase. The domain structure was intended to make the checklist usable across manuscript preparation, peer review, and editorial assessment by grouping related reporting items into coherent sections rather than presenting them as an undifferentiated list. At this stage, the framework was also structured around three dimensions: *authorship*, *logging*, and *use*. Authorship refers to the attribution of responsibility, human contribution, and accountability for AI-assisted scholarly work. Logging refers to the documentation of systems, prompts, interactions, workflows, and reproducibility-related information. Use refers to the substantive functions performed by AI in generating, transforming, validating, or shaping research content. These overarching dimensions served as organizing principles during framework structuring rather than as post hoc categories applied after validation.

This process resulted in nine reporting domains: (a) disclosure and scope of AI use, (b) human accountability and contribution, (c) AI system transparency, (d) prompting and interaction process, (e) data and AI-generated content, (f) validation and quality assurance, (g) ethics and compliance, (h) limitations and reflexivity, and (i) reproducibility and open practices. These domains were arranged to capture the main points at which AI use may require documentation, from initial declaration and author responsibility to system-

level transparency, interaction records, data-related use, validation procedures, ethical safeguards, limitations, and reproducibility constraints. Although some domains intersect across more than one dimension, each was assigned to the dimension that best reflected its primary reporting function. For example, ethics and compliance were aligned with authorship because privacy safeguards, policy adherence, bias mitigation, and responsible use ultimately depend on human accountability for AI-assisted research practices, even when they also involve technical or procedural considerations. Similarly, limitations and reflexivity were aligned with use because they concern how AI involvement may shape interpretation, framing, and the boundaries of the reported work.

The domain structure was also designed to preserve conceptual distinctiveness while recognizing that some reporting concerns are interrelated. For example, disclosure and accountability are closely connected, but they serve different reporting functions: disclosure identifies the presence and scope of AI use, while accountability clarifies human responsibility for AI-assisted outputs. Similarly, validation, limitations, and reproducibility are related but distinct. Validation concerns how AI-assisted outputs were checked, limitations concern the constraints introduced by AI use, and reproducibility concerns whether the AI-assisted workflow can be traced or approximated by others.

Instead of treating AI disclosure as a single standalone statement, PETMALU-AI distributes reporting expectations across the broader research workflow. This organization reflects the premise that AI involvement may influence multiple stages of scholarly production and therefore requires documentation that is systematic, proportionate, and sensitive to the specific functions AI performs within a given study. To support transparency in framework construction, the relationship between the evidence base, recurring reporting concerns, final domains, and checklist item numbers is summarized later in the paper.

2.5. Alignment with Existing Reporting Standards

PETMALU-AI was developed in alignment with the broader methodological principles that underpin established reporting guidelines, particularly transparency, completeness, replicability, and clear attribution of responsibility (Moher et al., 2010; Page et al., 2021). Similar to widely adopted reporting frameworks, the checklist uses a structured, itemized format to promote consistency in manuscript preparation and to support practical use by authors, reviewers, and editors. Its domain-based organization likewise reflects the logic of established reporting standards by grouping related reporting expectations into coherent sections rather than treating disclosure as an isolated statement.

The alignment with existing reporting standards was reflected in three design decisions. First, PETMALU-AI emphasizes transparency by requiring authors to document how AI was used, which system was involved, and how AI-assisted outputs were reviewed or validated. Second, it supports accountability by distinguishing AI assistance from human intellectual contribution and by requiring identification of the human author or authors responsible for AI-assisted content. Third, it advances reproducibility by encouraging authors to report prompts, workflows, system versions, access conditions, and other contextual details that may influence AI-generated outputs. In this sense, PETMALU-AI extends the reporting logic of established frameworks from conventional methodological procedures to machine-assisted research-writing practices. At the same time, PETMALU-AI was designed to address dimensions that are not adequately covered by conventional reporting standards. These include prompt-level transparency, the documentation of iterative human–AI interaction, disclosure of AI-generated or AI-augmented data, and acknowledgement of system evolution as a constraint on reproducibility. Such additions were necessary because AI systems may shape the generation, refinement, and presentation of scholarly content in ways that carry methodological and epistemic implications (Dwivedi et al., 2023; Park, 2025).

Accordingly, PETMALU-AI should be understood as complementary to existing reporting guidelines, not as a replacement for them. Authors using PETMALU-AI should still follow the reporting standards appropriate to their study design, such as PRISMA for systematic reviews or CONSORT for randomized trials. PETMALU-AI adds a cross-cutting layer of documentation for AI-assisted research writing by specifying how AI involvement should be disclosed, traced, validated, and contextualized. This alignment ensures that the framework remains anchored in established reporting principles while addressing the specific realities of generative AI-assisted scholarship.

2.6. Delphi-Based Validation and Reliability Assessment

A modified Delphi method was employed to evaluate the content validity, clarity, relevance, and structural coherence of the preliminary PETMALU-AI checklist. The modified Delphi approach was selected because it is widely used in the development of reporting guidelines, consensus statements, and evaluation frameworks where expert judgment is required to refine complex or emerging constructs (Diamond et al., 2014; Hsu & Sandford, 2007). The procedure involved iterative rounds of anonymous expert evaluation combined with structured feedback and quantitative consensus assessment. Across three rounds, panelists reviewed the candidate checklist items, assessed their relevance and clarity, provided qualitative recommendations, and participated in progressive refinement of the framework. The final procedure included expert recruitment based on predefined eligibility criteria, panel composition analysis, quantitative consensus thresholds using the item-level content validity index (I-CVI) and percentage agreement, and reliability assessment using Fleiss' kappa and the intraclass correlation coefficient. This process was intended to ensure that the final framework was both conceptually rigorous and practically interpretable across different scholarly and editorial contexts.

2.6.1. Participants

The PETMALU-AI checklist was validated through a three-round modified Delphi process involving experts recruited and selected through purposive sampling. The Delphi technique was chosen because it is widely used to develop consensus in emerging or conceptually complex areas where standardized guidance is still evolving (Crompton et al., 2026; Diamond et al., 2014; Hsu & Sandford, 2007; Xiao et al., 2025). A multidisciplinary panel was assembled to ensure that the framework would be evaluated from complementary perspectives relevant to AI-assisted research and scholarly communication.

Participants were selected based on at least one of the following criteria: (a) a demonstrated publication record in peer-reviewed journals related to artificial intelligence, research methodology, or scholarly communication; (b) editorial experience as a journal editor, reviewer, or editorial board member; or (c) recognized expertise in research ethics, governance, or policy development. Potential participants were identified through database searches, editorial board listings, and professional networks and were invited via email and social media websites. The panel was intentionally composed to include expertise in research methodology and reporting standards, AI and data science, academic publishing and editorial practice, and research ethics and policy. This composition was intended to ensure that the PETMALU-AI checklist would be reviewed from multiple perspectives relevant to AI-assisted research writing, including methodological rigor, editorial feasibility, technical transparency, and ethical accountability.

The final panel size was consistent with recommended Delphi ranges, which commonly involve 10 to 30 participants depending on the scope of the study and the expertise required (Diamond et al., 2014; Hsu & Sandford, 2007). A panel of 21 was considered sufficient to capture diversity of opinion while maintaining feasibility across repeated rounds of

evaluation. As summarized in Table 1, the panel represented varied geographic locations, disciplinary and professional affiliations, and levels of relevant professional experience. The distribution of expertise also reflected the intended multidisciplinary structure of the validation process, with representation from research methodology and reporting standards, AI and data science, academic publishing and editorial practice, and research ethics and policy. One panelist who completed Round 1 did not complete Rounds 2 and 3. This noncompletion was treated as attrition rather than disagreement with the framework, and no imputation was performed. Final quantitative analyses were therefore based only on the 20 experts who completed the final validation rounds. Experts participated anonymously during the Delphi rounds to reduce the influence of disciplinary hierarchy, institutional affiliation, or dominant personalities on the consensus process.

Table 1. Demographic and Professional Characteristics of the Delphi Panel at Round 1 (n = 21).

Characteristic	Category	*n* (%)
Geographic location	Philippines	9 (42.9%)
	Other Asian countries	5 (23.8%)
	North America	4 (19.0%)
	Europe	3 (14.3%)
Primary expertise area	Research methodology and reporting standards	6 (28.6%)
	Artificial intelligence and data science	5 (23.8%)
	Academic publishing and editorial practice	5 (23.8%)
	Research ethics and policy	5 (23.8%)
Institutional or professional affiliation	Higher education institution	13 (61.9%)
	Journal/editorial or publishing-related role	4 (19.0%)
	Research ethics, governance, or policy organization	2 (9.5%)
	AI, data science, or technical research unit	2 (9.5%)
Years of relevant professional experience	5–9 years	5 (23.8%)
	10–14 years	8 (38.1%)
	15 years and above	8 (38.1%)

Note. Demographic characteristics are reported for the 21 experts who completed Round 1. One panelist did not complete Rounds 2 and 3; therefore, quantitative consensus and reliability analyses were based on the 20 experts who completed the final validation rounds.

2.6.2. Procedure

The validation process was conducted between January and March 2026 across three iterative rounds administered electronically through a structured survey platform. Online administration ensured consistency in data collection and preserved respondent anonymity. This approach was intended to reduce the influence of dominant individuals and allow participants to reconsider their judgments independently across rounds, which is a defining strength of Delphi methodology (Diamond et al., 2014; Hsu & Sandford, 2007). After each round, aggregated quantitative results and anonymized qualitative feedback were reviewed and used to refine item wording, clarify conceptual boundaries, and improve the operational distinctiveness of overlapping items before redistribution in the succeeding round. The survey instrument presented each candidate checklist item with its proposed domain, item label, operational wording, and a brief explanatory note. Across the rounds, experts evaluated items individually, while domain-level coherence was assessed through open-ended comments rather than through a separate quantitative scale because the primary validation target was the item-level checklist structure.

In Round 1, participants were presented with the initial checklist and asked to assess each item in terms of clarity, relevance, and completeness. They were also invited to provide open-ended comments regarding item wording, omissions, and structural organization.

The qualitative feedback obtained in this round was assessed using thematic analysis to identify recurrent concerns and suggestions for refinement (Braun & Clarke, 2006). Based on this analysis, items were revised for clarity, merged where conceptual overlap was observed, and expanded where experts identified missing reporting elements.

In Round 2, the revised checklist was redistributed for quantitative appraisal. Participants rated the relevance of each item using a 4-point Likert-type scale ranging from 1 (not relevant) to 4 (highly relevant), consistent with established approaches to content validity assessment (Polit & Beck, 2006). Experts were again allowed to provide comments to contextualize their ratings. I-CVI scores were computed as the proportion of experts assigning a rating of either 3 or 4 to each item. Items that did not meet the predetermined content validity threshold were flagged for further revision or possible removal.

In Round 3, participants received the refined checklist along with anonymized summary feedback from the previous round, including I-CVI values and aggregated comments. This controlled feedback process allowed participants to re-evaluate their prior judgments in light of group-level patterns while preserving anonymity and independent judgment (Diamond et al., 2014; Hsu & Sandford, 2007). Final ratings were then collected to establish item retention and determine overall consensus.

Throughout the process, item revision was guided by both quantitative indices and qualitative feedback. This dual-criterion approach was used to ensure that retained items were not only statistically supported but also conceptually meaningful and practically interpretable. Quantitative consensus indicators helped determine whether items met predefined standards for relevance and agreement, while qualitative comments provided contextual insight into wording clarity, redundancy, and domain fit. Integrating both forms of evidence strengthened the refinement process by ensuring that item retention was based on measurable expert agreement as well as substantive interpretive feedback.

2.6.3. Consensus Criteria

Items were retained if they satisfied both of the following criteria: an I-CVI of at least 0.80 and at least 80% agreement among panelists in Round 3. The I-CVI threshold was selected based on recommendations for content validity assessment, particularly in expert-reviewed instrument development (Polit & Beck, 2006; Yusoff, 2019). The 80% agreement threshold was adopted to reflect strong expert consensus, consistent with methodological guidance for Delphi studies (Diamond et al., 2014). Items failing to meet one or both criteria were either revised for subsequent rounds or removed if they were judged to be conceptually redundant, unclear, or insufficiently essential (Gattrell et al., 2024). No item that entered the final Delphi rating stage was removed entirely during validation. Instead, items that initially received lower clarity or relevance feedback were revised, reworded, or conceptually consolidated before the final round. Item reduction occurred primarily during the pre-Delphi internal refinement phase, where overlapping candidate formulations were merged to avoid redundancy and reduce reporting burden.

2.6.4. Reliability Analysis

Inter-rater reliability was assessed to evaluate the stability and consistency of expert judgments. Fleiss' (1971) kappa was calculated using the dichotomized relevance ratings employed in the content validity analysis, where ratings of 3 or 4 were classified as relevant and ratings of 1 or 2 as not relevant (Polit & Beck, 2006). The unit of analysis consisted of all expert-item relevance judgments across the 33 checklist items completed by the final-round panelists. The intraclass correlation coefficient (ICC) was used to assess consistency in the original four-point relevance ratings submitted by the final-round panelists (McHugh, 2012). For the ICC analysis, a two-way random-effects model with absolute agreement

for average measures, ICC(2,k), was used because the checklist was rated by multiple experts sampled from a broader population of relevant specialists, and the interest was in the consistency of mean ratings across raters (Koo & Li, 2016). Interpretation of reliability coefficients followed established conventions. Values above 0.75 were interpreted as indicating good reliability, while values approaching or exceeding 0.90 were considered indicative of excellent consistency (Koo & Li, 2016).

2.7. Ethical Considerations

Participation in the Delphi process was voluntary. Experts were informed of the purpose of the study, the iterative nature of the Delphi procedure, and the use of anonymized responses for checklist refinement. Completion of each survey round was treated as informed consent to participate. To preserve confidentiality and minimize conformity pressure, responses were anonymized during analysis and reporting. Because the study focused on expert evaluation of a reporting framework and did not involve sensitive personal data or vulnerable populations, the study was considered minimal risk in nature.

3. Results

The results are presented in five parts to show how PETMALU-AI moved from conceptual grounding to validated reporting structure. First, the modified Delphi panel retention, consensus outcomes, and reliability indices are reported. Second, the item revision process across Delphi rounds is summarized. Third, the final domain structure of PETMALU-AI is presented. Fourth, the three conceptual dimensions of authorship, logging, and use are mapped to the reporting domains. Finally, the validated 33-item PETMALU-AI checklist is presented with corresponding rationales and examples.

3.1. Delphi Panel Retention, Consensus, and Reliability Outcomes

The modified Delphi process supported the content validity, clarity, and structural coherence of the PETMALU-AI checklist. Across the three rounds, expert participation remained high, with 21 panelists completing Round 1 and 20 completing both Rounds 2 and 3, indicating strong retention throughout the validation process. Final consensus was achieved for all retained items, with Round 3 agreement rates ranging from 85% to 100%. Item-level content validity was likewise strong, as reflected in a mean I-CVI of 0.91 and an observed range of 0.85 to 1.00. The scale-level content validity index, computed as the average of the item-level CVIs (S-CVI/Ave), was 0.91, indicating excellent overall content validity of the checklist. These findings suggest that the final checklist items were consistently judged by the expert panel to be appropriate for inclusion. Inter-rater reliability further supported the robustness of the framework. Fleiss' kappa for dichotomized relevance ratings was 0.84, indicating strong agreement among panelists, while the intraclass correlation coefficient for average measures, ICC(2,k), was 0.89, indicating high consistency in item ratings across experts. These indices provide converging evidence that the final version of PETMALU-AI achieved both strong expert consensus and stable evaluative agreement. A summary of the Delphi outcomes is presented in Table 2, while a domain-level summary of the final Delphi validation outcomes is provided in Appendix C.

Table 2. Delphi Validation Results for PETMALU-AI Checklist.

Metric	Value
Number of experts invited	21
Number of experts completing Round 1	21
Number of experts completing Round 2	20

Table 2. *Cont.*

Metric	Value
Number of experts completing Round 3	20
Final number of items retained	33
Number of domains	9
Mean I-CVI	0.91
Range of I-CVI	0.85–1.00
S-CVI/Ave	0.91
Final agreement rate (Round 3)	85–100%
Fleiss' kappa (κ)	0.84
ICC, average measures [ICC(2,k)]	0.89

3.2. Item Revision and Refinement Across Delphi Rounds

The Delphi process resulted in targeted refinements to the wording, conceptual boundaries, and domain placement of several checklist items. Although the expert panel broadly supported the initial structure of PETMALU-AI, qualitative feedback from Round 1 identified three recurring areas for improvement. First, panelists recommended a clearer distinction between general disclosure of AI use and human accountability for AI-assisted outputs. Second, they emphasized that prompting should be represented as a methodological process rather than a purely technical detail. Third, they noted areas of conceptual overlap among items related to validation, limitations, and reproducibility. These comments informed revisions prior to Round 2, with most changes involving clarification, consolidation, or improved wording rather than major restructuring of the framework.

In Round 2, quantitative ratings confirmed the relevance of the revised items, while additional comments were used to improve the precision of items that approached the lower end of the consensus range. These refinements focused on making each item more operationally distinct. For example, items related to human verification, validation strategy, and robustness checks were clarified to distinguish routine author review from more formal procedures for assessing the reliability or consistency of AI-assisted outputs. Items related to reproducibility constraints and system evolution were also refined to distinguish limitations arising from AI variability from broader open-practice expectations. By Round 3, remaining comments focused primarily on wording precision rather than changes to the overall structure, indicating that the checklist had reached a stable form.

No checklist item was removed entirely during the Delphi validation rounds. Rather, the validation process resulted in targeted revisions to item wording, domain placement, and conceptual boundaries. The final 33-item structure reflects a process of refinement rather than deletion during expert validation. Earlier reductions occurred before the Delphi rounds, when preliminary candidate items with substantial overlap were merged into broader reporting items. For example, separate provisional items addressing minor language editing, drafting assistance, and substantive AI contribution were consolidated under materiality of AI use, while overlapping items on checking AI outputs, verifying generated content, and identifying hallucinations were refined into the broader validation and error/bias items. This approach preserved the conceptual coverage of the original item pool while improving usability and avoiding unnecessary duplication. Table 3 summarizes the numerical progression of checklist development, while Appendix B documents the corresponding item-level refinement process, including whether items were retained, reworded, merged during pre-Delphi refinement, or repositioned across domains.

Table 3. Progression of PETMALU-AI Checklist Development.

Development Stage	Number of Items	Main Activity
Initial candidate pool	42	Initial item generation from the literature and policy synthesis
Pre-Delphi internal refinement	33	Consolidation of 9 overlapping or overly narrow items
Round 1	33	Expert qualitative review of item clarity, relevance, and domain fit
Round 2	33	Quantitative relevance rating and targeted wording revisions
Round 3	33	Final consensus rating and validation
Final Checklist	33	Validated PETMALU-AI reporting framework

3.3. Final Domain Structure of the PETMALU-AI Framework

The finalized PETMALU-AI framework consists of 33 reporting items organized into nine domains spanning the major stages and methodological considerations of AI-assisted research writing. These domains address disclosure and scope of AI use, human accountability and contribution, AI system transparency, prompting and interaction processes, data and AI-generated content, validation and quality assurance, ethics and compliance, limitations and reflexivity, and reproducibility and open practices. Collectively, domains provide a structured approach for documenting not only whether AI was used, but also how it was used, who remained responsible for AI-assisted outputs, how generated content was reviewed, and what reproducibility constraints should be acknowledged.

The domain structure was designed to correspond to the typical reporting needs of AI-assisted scholarly work. Rather than concentrating AI disclosure in a single statement, PETMALU-AI distributes reporting expectations across the broader research-writing workflow. This structure allows authors to document AI involvement at different levels of specificity, including general disclosure, author responsibility, tool identification, prompting practices, data-related use, validation procedures, ethical safeguards, reflexive limitations, and reproducibility information. In this way, the framework treats AI reporting as a multidimensional process rather than a single compliance requirement.

The final domains were also traceable to the recurring principles identified during the narrative synthesis and expert validation process. Publisher and organizational policies primarily informed items related to disclosure, authorship accountability, policy compliance, and human responsibility (e.g., Elsevier, 2025; International Committee of Medical Journal Editors, 2026; Springer Nature, 2026; Wiley, 2026). Scholarly literature on AI-assisted writing, research integrity, methodological transparency, and reproducibility informed items related to prompting, validation, data provenance, limitations, and open practices (e.g., Dwivedi et al., 2023; Ganjavi et al., 2024; Khalifa & Albadawy, 2024; Resnik & Hosseini, 2026b). This synthesis-to-domain alignment ensured that the final structure was informed by expert judgment during the Delphi process while remaining grounded in existing policy expectations and emerging scholarly concerns.

To make this derivation explicit, Table 4 maps each PETMALU-AI domain and corresponding item numbers to the main evidence base, recurring reporting concern, and resulting checklist focus. This mapping demonstrates how broad concerns in the literature and policy environment were translated into specific reporting domains. It also clarifies that the domains are not isolated categories, but interrelated components of a unified framework for transparent, accountable, and reproducible AI-assisted research writing. A more detailed item-level refinement trail is provided in Appendix B.

Table 4. Evidence-to-Domain and Item Mapping for PETMALU-AI.

Domain (Items Covered)	Main Evidence Base	Recurring Reporting Concern	Resulting Checklist Focus
Disclosure and Scope of AI Use (P1–P3)	Publisher AI policies, journal disclosure requirements, research transparency literature, and emerging scholarship on AI-assisted writing	AI use is often disclosed only in broad terms, leaving readers uncertain about where, why, and how materially AI contributed to the manuscript or research process.	Requires authors to declare AI use, specify the scope of involvement, and distinguish minor editorial assistance from more substantive AI contribution.
Human Accountability and Contribution (P4–P7)	Authorship guidance, research integrity literature, publisher policies prohibiting AI authorship, and discussions of human responsibility in AI-assisted scholarship	AI systems cannot assume responsibility for scholarly claims, methodological decisions, errors, interpretations, or ethical obligations.	Clarifies human–AI contribution boundaries, confirms authorship compliance, assigns accountability, and identifies responsibility for prompt design.
AI System Transparency (P8–P11)	AI documentation literature, reproducibility scholarship, publisher disclosure expectations, and technical transparency discussions	Readers and reviewers may not know which AI system was used, whether model/version details mattered, how the system was accessed, or whether configuration choices influenced outputs.	Requires identification of AI tools, model or version information where available, access context, and relevant configuration details.
Prompting and Interaction Process (P12–P15)	Human–AI interaction literature, prompt-based workflow discussions, methodological transparency literature, and reproducibility scholarship	AI outputs are shaped by prompts, iterative exchanges, and author decisions, yet these interaction processes are usually invisible in published manuscripts.	Documents prompting approach, iterative use, output curation, and the availability of interaction records where feasible and appropriate.
Data and AI-Generated Content (P16–P19)	Data provenance literature, synthetic data discussions, AI-assisted data processing literature, and research ethics guidance	AI may generate, transform, augment, summarize, or classify data in ways that affect provenance, interpretation, reuse, and trustworthiness.	Requires transparency about data origin, justification for AI-generated or synthetic data, description of data-generation processes, and data or protocol availability.
Validation and Quality Assurance (P20–P23)	Literature on hallucination, fabricated citations, bias, error detection, verification, and scholarly integrity	AI-generated outputs may appear fluent and authoritative while containing inaccuracies, omissions, hallucinated references, biased framing, or unstable results.	Requires human verification, validation procedures, error and bias consideration, and robustness checks where relevant.
Ethics and Compliance (P24–P27)	Institutional policies, publisher guidance, professional standards, AI ethics literature, and privacy/confidentiality guidance	AI use may raise ethical concerns related to fairness, privacy, confidentiality, compliance, bias, and responsible research conduct.	Requires ethical consideration, policy compliance, bias mitigation, and privacy safeguards.
Limitations and Reflexivity (P28–P30)	Literature on epistemic risk, AI-mediated writing, interpretive framing, reflexivity, and AI reproducibility constraints	AI may shape framing, emphasis, organization, interpretation, and reproducibility even when human authors retain final control.	Encourages authors to report AI-related limitations, reflect on how AI may have shaped the manuscript, and acknowledge reproducibility constraints.
Reproducibility and Open Practices (P31–P33)	Reporting guideline traditions, open science literature, computational reproducibility scholarship, and AI workflow documentation literature	AI-assisted workflows may be difficult to reproduce because prompts, models, interfaces, access conditions, and platform behavior can change over time.	Encourages reporting of workflow details, availability of supporting materials, and disclosure of system evolution as a constraint on future replication.

Note. This table maps the nine domains and all 33 checklist items to the main bodies of evidence and recurring reporting concerns that informed the framework. The mapping is not intended to imply that each domain emerged from a single source category. Rather, it shows how recurring concerns across publisher policies, AI ethics literature, research integrity scholarship, reproducibility discourse, and reporting guideline traditions were translated into the final checklist structure.

3.4. Conceptual Dimensions and Domain Mapping

Building on the evidence-to-domain mapping presented in Table 4, the final framework is organized around three interrelated conceptual dimensions: authorship (who contributes), logging (how interactions are documented), and use (what functions AI performs). These dimensions were derived from the recurring concerns identified during the narrative synthesis and refined through expert feedback during the Delphi process. The au-

thorship dimension addresses questions of responsibility, contribution, and accountability. The logging dimension concerns the documentation of AI systems, prompts, interactions, workflows, and reproducibility-related materials. The use dimension focuses on the substantive functions performed by AI, including its role in data-related processes, output validation, and the interpretation or presentation of research.

Each of the nine PETMALU-AI domains aligns primarily with one of these three dimensions while still contributing to the overall purpose of transparent and accountable AI-use reporting. Domains related to disclosure, human accountability, and ethics were grouped under authorship because they clarify the boundaries between human responsibility and AI assistance. Domains related to AI system transparency, prompting and interaction processes, and reproducibility were grouped under logging because they document how AI-assisted processes were conducted and whether they can be traced or approximated. Domains related to data and AI-generated content, validation and quality assurance, and limitations and reflexivity were grouped under use because they address how AI contributed to research materials, outputs, verification, and interpretation.

Table 5 presents this mapping between the three conceptual dimensions and the nine reporting domains. The table is intended to clarify the internal organization of PETMALU-AI and show how the checklist items are connected to broader reporting concerns. This mapping also helps demonstrate that the framework is not simply a list of disclosure items. Rather, it is structured around the related but distinct problems of identifying who remains accountable, documenting how AI-assisted work was produced, and clarifying what functions AI performed in the research-writing process. By organizing the domains in this way, the framework helps prevent AI disclosure from being reduced to a generic compliance statement and instead positions it as an account of scholarly responsibility, process transparency, and methodological influence. The mapping also supports practical use by helping authors, reviewers, and editors locate relevant reporting expectations.

Table 5. Mapping of PETMALU-AI Dimensions to Reporting Domains.

Dimension	Reporting Domain	Rationale
Authorship	Section 1: Disclosure and Scope of AI Use	Establishes transparency regarding the extent of AI contribution to the research process
	Section 2: Human Accountability and Contribution	Ensures clear delineation of responsibility and intellectual ownership between human authors and AI systems
	Section 7: Ethics and Compliance	Reinforces accountability through adherence to authorship and ethical guidelines
Logging	Section 3: AI System Transparency	Documents system-level details that influence outputs and interactions
	Section 4: Prompting and Interaction Process	Captures how human–AI interactions are structured, iterated, and refined
	Section 9: Reproducibility and Open Practices	Enables traceability and replication through documentation of processes and materials
Use	Section 5: Data and AI-Generated Content	Defines how AI contributes to data creation, transformation, or augmentation
	Section 6: Validation and Quality Assurance	Ensures that AI-generated outputs are evaluated for accuracy and reliability
	Section 8: Limitations and Reflexivity	Acknowledges the impact of AI use on findings, interpretation, and generalizability

3.5. Final PETMALU-AI Checklist

The final PETMALU-AI checklist is presented in Table 6. The checklist includes explanatory rationale and illustrative examples for each reporting item to support consistent interpretation across disciplines and journal contexts. Rather than presenting isolated disclosure prompts, the checklist operationalizes the analytical dimensions of authorship, logging, and use into specific reporting expectations that can be applied during manuscript preparation, peer review, and editorial assessment. The inclusion of examples was intended to improve usability by translating abstract reporting expectations into recognizable manuscript practices. The Delphi results and the final structural configuration indicate that PETMALU-AI is not merely a list of disclosure prompts, but a multidimensional reporting framework capable of capturing the methodological, ethical, and reproducibility-related implications of AI-assisted research writing. The final checklist is intended to provide sufficient breadth for comprehensive reporting while remaining organized enough for practical use by authors, reviewers, and editors. To support consistent implementation, adjacent checklist items were worded to emphasize their distinct reporting functions, particularly where disclosure, validation, reproducibility, and accountability requirements could otherwise appear overlapping.

Table 6. PETMALU-AI Checklist with Rationale and Examples.

Item	Checklist	Rationale	Example
Section 1: Disclosure and Scope of AI Use			
P1	AI Use Disclosure Statement	Ensures transparency by requiring an explicit declaration that AI was used during the research or writing process.	*"AI tools were used during manuscript preparation."*
P2	Scope of AI Involvement	Clarifies the specific stages or tasks in which AI contributed to the research or writing workflow.	*"AI was used for literature summarization, language editing, and code generation."*
P3	Materiality of AI Use	Distinguishes minor language or formatting support from AI use that may have influenced study design, analysis, interpretation, or conclusions.	*"AI was used only for language editing and did not influence study design, analysis, interpretation, or conclusions."*
Section 2: Human Accountability and Contribution			
P4	Human–AI Contribution Delineation	Clarifies which parts of the manuscript or research process were produced, shaped, or finalized by human authors and which were assisted by AI.	*"All interpretations and conclusions were developed and finalized by the authors."*
P5	Authorship Compliance	Ensures that authorship is limited to accountable human contributors and that AI systems are not credited as authors.	*"No AI system is listed as an author in this manuscript."*
P6	Accountability Assignment	Identifies the human author or authors responsible for reviewing, verifying, and approving AI-assisted content or outputs.	*"The corresponding author assumes responsibility for all AI-assisted sections."*
P7	Prompt Design Responsibility	Specifies who designed, refined, and approved the prompts used to generate or revise AI-assisted outputs.	*"Prompts were designed, reviewed, and refined by the first author."*
Section 3: AI System Transparency			
P8	AI Tool Identification	Identifies the specific AI tool or platform used to support traceability and interpretation of AI-assisted outputs.	*"The study used ChatGPT for drafting assistance."*
P9	Model/Version Disclosure	Reports the AI model, version, or system release when such information is available and relevant to output interpretation.	*"GPT-4 architecture was used during manuscript preparation."*
P10	Access Context	Describes how the AI system was accessed or integrated into the research or writing workflow.	*"The tool was accessed through a web-based interface."*
P11	Configuration Disclosure	Reports relevant settings, parameters, or configuration choices that may have influenced AI-generated outputs.	*"Default settings were used without parameter modification."*

Table 6. *Cont.*

Item	Checklist	Rationale	Example
Section 4: Prompting and Interaction Process			
P12	Prompting Approach Description	Describes the general structure, purpose, or strategy of prompts used to guide AI-generated outputs.	*"Structured prompts were used to guide paragraph generation and revision."*
P13	Iterative Use Disclosure	Documents whether AI outputs were generated through repeated prompt-output cycles or progressive refinement.	*"Outputs were iteratively refined across three prompt cycles."*
P14	Output Curation Process	Explains how AI-generated outputs were selected, edited, accepted, rejected, or integrated into the manuscript.	*"Generated text was reviewed, selectively retained, and manually edited for accuracy and fit."*
P15	Interaction Records Availability	Indicates whether prompts, outputs, or workflow records are available to support auditability or verification when appropriate.	*"Prompt logs and major output revisions are available upon reasonable request."*
Section 5: Data and AI-Generated Content			
P16	Data Origin Transparency	Identifies whether data used or discussed in the study were human-generated, AI-generated, AI-augmented, or derived from existing sources.	*"All datasets were human-generated; no synthetic data were used."*
P17	Justification for AI Data	Explains the rationale for using AI-generated or synthetic data when such data are included in the study.	*"Synthetic data were used to simulate rare scenarios."*
P18	Data Generation Description	Describes how AI was used to generate, transform, augment, or modify data in the research process.	*"AI was used to augment missing data points."*
P19	Data Availability	Indicates whether datasets, AI-generated data, or data-generation protocols are accessible for verification or reuse.	*"The dataset is available in an open-access repository."*
Section 6: Validation and Quality Assurance			
P20	Human Verification Process	Documents author review of AI-generated outputs for accuracy, appropriateness, and alignment with the manuscript.	*"All AI-generated outputs were checked by the authors before inclusion."*
P21	Validation Strategy	Reports formal or external procedures used to evaluate the accuracy, reliability, or adequacy of AI-assisted outputs.	*"Generated summaries were compared with source documents and benchmark references."*
P22	Error and Bias Consideration	Acknowledges and addresses potential hallucinations, omissions, citation errors, biased framing, or other AI-related distortions.	*"Potential hallucinations, citation errors, and biased framing were evaluated during revision."*
P23	Robustness Checks	Assesses whether AI-generated outputs remained consistent across alternative prompts, repeated runs, or system variations.	*"Outputs were compared across multiple prompt variations to assess consistency."*
Section 7: Ethics and Compliance			
P24	Ethical Considerations	Addresses ethical issues raised by AI use in the study, including fairness, transparency, responsibility, and potential harm.	*"Ethical implications of AI-generated content were considered."*
P25	Policy Compliance Statement	Confirms adherence to relevant institutional, publisher, journal, or professional guidelines on AI use.	*"The study complies with journal AI use policies."*
P26	Bias Mitigation	Describes steps taken to reduce, detect, or manage bias introduced through AI-assisted processes or outputs.	*"Bias mitigation strategies included manual review and filtering."*
P27	Privacy Safeguards	Explains how sensitive, confidential, proprietary, or identifiable information was protected when using AI systems.	*"No personal or identifiable data were processed by AI systems."*
Section 8: Limitations and Reflexivity			
P28	AI-Related Limitations	Reports constraints introduced by the use of AI that may affect the manuscript, study process, or interpretation of outputs.	*"AI-generated outputs may lack contextual nuance and may reflect limitations of the training data."*

Table 6. *Cont.*

Item	Checklist	Rationale	Example
Section 8: Limitations and Reflexivity			
P29	Reflexive Consideration	Encourages author reflection on how AI assistance may have shaped framing, emphasis, organization, or presentation.	*"AI assistance may have influenced the organization and wording of selected sections, although all interpretations were finalized by the author."*
P30	Reproducibility Constraints	Identifies barriers to exact replication arising from variability, non-determinism, access conditions, or model changes.	*"Exact outputs may not be reproducible because generative AI systems can produce variable responses across sessions."*
Section 9: Reproducibility and Open Practices			
P31	Reproducibility Information	Provides procedural details needed to understand and approximate the AI-assisted workflow.	*"The methodology includes the AI tool, access context, prompting approach, and validation procedures."*
P32	Materials Availability	Specifies whether supporting materials such as prompts, workflows, logs, or templates are available and under what conditions.	*"Prompt templates and workflow notes are included in the supplementary materials."*
P33	System Evolution Disclaimer	Notes that AI models, interfaces, access conditions, or platform behavior may change over time and affect future outputs.	*"Future outputs may differ because the AI model and platform interface may be updated over time."*

4. Discussion

4.1. PETMALU-AI as a Framework for Structured AI-Use Reporting

The present study contributes to the growing literature on generative AI in scholarly writing by addressing a critical gap between the increasing normalization of AI-assisted research practices and the limited availability of structured mechanisms for documenting such practices. Recent studies have shown that generative AI tools are now used across multiple stages of academic work, including idea development, content structuring, literature synthesis, drafting, editing, data management, and publication-related support (Khalifa & Albadawy, 2024). However, the corresponding reporting practices have not developed at the same pace. Although journals and publishers increasingly require authors to disclose AI use, existing guidance remains uneven in terminology, level of detail, location of disclosure, and treatment of accountability (Ganjavi et al., 2024; Laffaye et al., 2025). PETMALU-AI responds to this inconsistency by offering a structured reporting framework that moves AI disclosure from a broad declarative statement to a more systematic account of how AI was used, what role it played, who remained accountable, how outputs were verified, and what limitations should be considered.

A first contribution of PETMALU-AI is that it operationalizes AI transparency as a reportable scholarly practice rather than a general ethical aspiration. Existing literature has repeatedly emphasized the need for transparency in the use of generative AI, particularly because undisclosed AI assistance may obscure bias, factual inaccuracy, fabricated references, and the extent of machine involvement in scholarly texts (Tang et al., 2024; Yousaf, 2025). However, transparency becomes difficult to implement consistently when authors are only instructed to "disclose AI use" without being guided on the scope, materiality, validation, or reproducibility implications of that use. PETMALU-AI addresses this limitation by translating broad transparency expectations into specific domains and checklist items. In doing so, it gives authors, reviewers, and editors a common vocabulary for distinguishing minor editorial assistance from more substantive AI involvement that may affect the development, presentation, or interpretation of research.

A second contribution concerns accountability. Current debates on generative AI in academic writing often focus on whether AI systems can be considered authors, whether AI-generated text challenges ownership, and how responsibility should be assigned when machine-generated content enters the scholarly record (Bozkurt, 2024; Dwivedi et al., 2023).

PETMALU-AI advances this discussion by treating accountability not merely as a statement that "authors remain responsible," but as a set of concrete reporting obligations. These include delineating human and AI contributions, identifying responsible authors, clarifying prompt design responsibility, verifying AI-assisted outputs, and documenting compliance with institutional or journal policies. This structure is important because AI systems cannot assume responsibility for accuracy, interpretation, ethical compliance, or correction of errors. Therefore, human accountability must be made visible not only as a principle, but as a documented part of the research-writing process.

A third contribution is the framework's attention to the materiality of AI use. Recent scholarship has argued that not all AI assistance carries the same implications for transparency and research integrity (Resnik & Hosseini, 2026a). For example, grammar correction, translation support, paragraph restructuring, literature summarization, synthetic data generation, and statistical interpretation do not carry equivalent risks or require the same depth of reporting. PETMALU-AI responds to this issue by distinguishing the presence of AI use from the significance of AI use. This proportional orientation prevents the checklist from becoming either too permissive or too burdensome. Minor editorial assistance may require concise disclosure, while more substantive AI involvement requires fuller documentation of prompts, outputs, validation, data provenance, limitations, and reproducibility constraints. In this way, the framework supports responsible transparency without assuming that all AI use should be treated as equally consequential.

A fourth contribution is that PETMALU-AI makes the often-invisible process of human–AI interaction more auditable. Generative AI outputs are shaped by prompts, iterative exchanges, system configurations, author selection, and post-generation editing (Xiao et al., 2025). Without documentation of these processes, readers and reviewers may be unable to determine how AI-assisted outputs were produced, filtered, modified, or integrated into the manuscript. This is especially important given persistent concerns that generative AI may produce fluent but inaccurate content, unsupported claims, biased framing, or fabricated references (Acut et al., 2025; Boretti, 2026; Calderon & Herrera, 2025; Resnik & Hosseini, 2026b). By including reporting items on prompting approach, iterative use, output curation, interaction records, validation procedures, and robustness checks, PETMALU-AI strengthens the auditability of AI-assisted writing. The checklist therefore asks authors to make the process sufficiently visible for scholarly evaluation.

Finally, PETMALU-AI contributes a governance-oriented reporting logic for AI-assisted scholarship. Its purpose is not to promote the use of generative AI in research writing, nor to suggest that AI assistance is inherently acceptable when disclosed. Rather, the framework raises the evidentiary burden for authors who choose to use AI tools by requiring documentation of transparency, accountability, validation, ethical safeguards, and reproducibility constraints. This distinction is essential because concerns about AI-assisted writing involve trust, interpretability, research integrity, and the protection of the scholarly record. PETMALU-AI consequently positions structured reporting as a mechanism of scrutiny. It makes AI involvement harder to obscure, easier to evaluate, and more accountable to readers, reviewers, editors, and the broader research community.

4.2. From Ethical Disclosure to Methodological Transparency

A central implication of this study is that AI-assisted research writing should not be understood solely as an ethical or disclosure issue (Mezzadri, 2025; Yousaf, 2025). While ethical concerns remain essential, the structure of PETMALU-AI reflects the position that generative AI may also function as a methodological influence within the research-writing process. This distinction is important because AI tools are no longer limited to grammar correction or surface-level language editing. Recent reviews indicate that AI is increasingly

used to support idea generation, content structuring, literature synthesis, data management, editing, and publication-related activities (Khalifa & Albadawy, 2024; Laffaye et al., 2025). When AI contributes to these stages, its role may affect not only how a manuscript is written but also how ideas are organized, how evidence is summarized, how arguments are framed, and how scholarly claims are presented. PETMALU-AI therefore treats AI use as a reportable part of the research workflow when it has the potential to shape the production, interpretation, or communication of knowledge.

This methodological framing distinguishes PETMALU-AI from approaches that reduce AI reporting to a brief acknowledgment or declaration. A disclosure statement can establish that AI was used, but it does not necessarily explain what the tool did, how its outputs were generated, whether those outputs were verified, or whether AI assistance influenced the presentation of findings. Such limitations are especially important because generative AI systems may produce fluent and convincing text while also introducing inaccuracies, fabricated references, plagiarism risks, biased framing, or unsupported claims (Cheng et al., 2025; Cohen & Moher, 2025). In this context, the central issue is not merely whether AI use is ethically permissible, but whether the resulting manuscript remains transparent, verifiable, and accountable. PETMALU-AI addresses this concern by requiring authors to document the scope of AI involvement, the materiality of its contribution, the prompting and interaction process, the validation procedures applied, and the reproducibility constraints introduced by the use of generative systems.

The inclusion of prompting, interaction logging, validation, and reproducibility domains is especially important because generative AI outputs are shaped by processes that are often invisible in published manuscripts. Unlike conventional writing tools, generative AI systems respond to user prompts, prior context, system settings, interface conditions, and iterative exchanges (Bozkurt et al., 2024). As a result, two authors using the same platform may obtain different outputs depending on how they formulate prompts, refine responses, select generated material, and edit the final text. These processes are not merely technical details. They can shape the substance, emphasis, organization, and evidentiary presentation of scholarly writing. By requiring authors to describe prompting approaches, iterative use, output curation, and the availability of interaction records where appropriate, PETMALU-AI makes visible the human decisions that mediate AI-assisted outputs. This is consistent with recent guidance emphasizing that authors must remain responsible for the accuracy, originality, integrity, and final approval of AI-assisted manuscripts (Cheng et al., 2025; Cleland et al., 2026; Taneja et al., 2025).

At the same time, PETMALU-AI does not expand into a general framework for AI ethics. Its ethical dimension is deliberately tied to reporting and methodological transparency. The checklist asks authors to report ethical considerations only insofar as they affect authorship accountability, reproducibility, privacy safeguards, bias mitigation, policy compliance, or validation. This boundary is important because the purpose of PETMALU-AI is not to determine whether generative AI is morally acceptable in all research contexts. Rather, its purpose is to ensure that when AI is used, its role is documented with enough specificity for readers, reviewers, and editors to evaluate the integrity of the manuscript. In this sense, it converts ethical concern into reportable methodological information.

This perspective also helps address potential criticism that a reporting checklist for AI-assisted writing might be interpreted as an endorsement of AI use. PETMALU-AI should not be read as a permissive framework that normalizes unrestricted reliance on generative AI. On the contrary, it raises the burden of justification for authors who use AI tools by requiring them to disclose the nature, extent, validation, and limitations of that use. Recent scholarship has emphasized that disclosure should be especially important when AI use is intentional and substantial, particularly when it contributes to evidence,

analysis, discussion, interpretation, or manuscript content (Resnik & Hosseini, 2026a). PETMALU-AI aligns with this view by treating AI use as methodologically consequential when it influences the content or reasoning of a manuscript. The framework therefore does not promote AI-assisted research writing; it makes AI-assisted research writing more difficult to hide, more open to scrutiny, and more accountable when it occurs.

Framing AI-assisted writing as methodological rather than merely ethical also clarifies the broader value of the checklist. If AI use is treated only as an ethical declaration, then disclosure may become a minimal compliance exercise. However, if AI use is treated as part of the research process, then reporting must address how AI affected the workflow, what safeguards were applied, and what limitations remain. This shift is necessary because the credibility of AI-assisted scholarship depends not only on author honesty but also on the interpretability and auditability of the processes through which AI-supported text, summaries, analyses, or data-related outputs were produced. PETMALU-AI contributes to this shift by offering a structured mechanism for transforming AI disclosure into methodological transparency. In doing so, it strengthens the conditions under which AI-assisted scholarship can be evaluated, reproduced, and trusted.

4.3. Extending Reporting Guideline Traditions to AI-Assisted Scholarship

PETMALU-AI is situated within the broader tradition of reporting guidelines, but it extends that tradition to a new form of scholarly production shaped by generative AI. Established reporting frameworks such as PRISMA (Page et al., 2021) and CONSORT (Hopewell et al., 2025) were developed to improve the completeness, transparency, interpretability, and reproducibility of research reporting. Their value lies not only in requiring authors to disclose information, but in standardizing the kinds of procedural details that readers, reviewers, and editors need to evaluate the credibility of a study. PETMALU-AI adopts this same logic by treating AI-assisted research writing as a process that requires structured documentation. Rather than asking authors to provide a generic statement that AI was used, the checklist specifies the domains through which AI involvement can be made visible, including disclosure, human accountability, system transparency, prompting, data provenance, validation, ethics, limitations, and reproducibility.

At the same time, PETMALU-AI differs from conventional reporting guidelines because generative AI introduces reporting problems that were not anticipated by earlier frameworks. Traditional reporting standards generally assume that methods, instruments, procedures, and analytical decisions can be described with reasonable stability. Generative AI systems complicate this assumption because outputs may vary across prompts, model versions, access conditions, system interfaces, configuration settings, and interaction histories (Bozkurt et al., 2024; Xiao et al., 2025). This variability means that AI-assisted writing cannot be reported only by naming the tool used. Authors may also need to describe how the tool was accessed, what kinds of prompts were used, how outputs were selected or rejected, whether generated material was verified, and what reproducibility constraints remain. These reporting needs are especially important because generative AI can produce fluent but inaccurate or unverifiable content, making procedural transparency central to the evaluation of AI-assisted manuscripts.

The relevance of this extension is supported by recent evidence showing that journal and publisher guidance on generative AI remains inconsistent. Ganjavi et al. (2024) found substantial variation in publisher and journal instructions to authors, including differences in whether guidance existed, what forms of AI use were permitted, where disclosure should appear, and what information authors were expected to provide. Cleland et al. (2026) similarly noted that journals differ in whether they allow AI for language polishing, prohibit AI-generated text, require disclosure in the Methods section, require disclosure in

the Acknowledgments, or impose other journal-specific expectations. This inconsistency creates practical uncertainty for authors and reviewers (Alduais et al., 2025; Garcia, 2023). PETMALU-AI responds to this problem by offering an operational structure that can help translate broad policy expectations into reportable manuscript elements. In this respect, the framework does not replace journal policies. Instead, it can serve as a bridge between policy-level principles and concrete reporting practice.

PETMALU-AI also aligns with the emergence of more specialized AI-reporting guidance. For example, the GAMER Statement proposed a checklist for reporting the use of generative AI tools in medical research, with items covering general declaration, AI tool specifications, prompting techniques, the role of AI in the study, AI-assisted manuscript sections, content verification, data privacy, and the potential impact of AI on conclusions (Luo et al., 2025). Similarly, CHART provides reporting recommendations for studies evaluating generative AI-driven chatbots when summarizing clinical evidence or providing health advice, while ICMJE guidance emphasizes disclosure, human accountability, and the exclusion of AI systems from authorship (Huo et al., 2025). PETMALU-AI is consistent with this checklist-based direction but differs in scope and conceptual organization. While GAMER is explicitly oriented toward medical research and CHART toward chatbot health advice studies, PETMALU-AI is designed for broader AI-assisted research writing across disciplines. It also organizes AI reporting around the interrelated dimensions of authorship, logging, and use. This organization clarifies the relationship between human responsibility, documentation of interaction, and the functions performed by AI. This broader structure allows PETMALU-AI to capture whether AI was used and how human authors governed, verified, delimited, and contextualized that use.

A further distinction is PETMALU-AI's emphasis on proportional reporting. Existing reporting guidelines are often applied to specific research designs or methodological traditions, whereas AI assistance may vary widely in depth and consequence. Some AI use may involve minor grammar correction, while other use may involve literature summarization, coding assistance, data transformation, synthetic data generation, or interpretation support. These forms of AI use do not carry the same implications for transparency or reproducibility. PETMALU-AI therefore extends reporting guideline logic by incorporating materiality into disclosure expectations. This approach is consistent with arguments that AI disclosure should be mandatory when use is intentional and substantial, particularly when it directly affects the content of the research or produces evidence, analysis, or discussion that supports a study's conclusions (Martínez-Requejo et al., 2025; Resnik & Hosseini, 2026a). By distinguishing different levels of AI involvement, the PETMALU-AI checklist helps prevent both underreporting of consequential AI use and overburdening of authors whose AI use was limited to low-materiality editorial support.

This comparison with existing reporting traditions also clarifies what PETMALU-AI does not claim to do. Like PRISMA does not guarantee the quality of a systematic review and CONSORT does not guarantee the methodological strength of a clinical trial, PETMALU-AI does not guarantee that AI-assisted writing is accurate, ethical, or acceptable. A checklist can improve reporting completeness, but it cannot substitute for sound judgment, rigorous verification, responsible authorship, or editorial oversight. Its function is procedural and evaluative because it makes AI involvement more visible and allows claims, processes, and safeguards to be assessed. This distinction is important because a framework for reporting AI use could be misread as a framework for legitimizing AI use. PETMALU-AI should instead be understood as a mechanism for scrutiny. It does not lower the threshold for acceptable AI-assisted scholarship. Instead, it raises the standard for what authors must disclose, justify, and contextualize when AI tools are used.

By extending reporting guideline traditions to AI-assisted scholarship, PETMALU-AI contributes to a more mature form of AI governance in research writing. It recognizes that the challenge is no longer limited to deciding whether AI use should be disclosed. The more difficult task is determining what information is needed for AI-assisted scholarship to remain interpretable, reviewable, and reproducible. In addressing this task, PETMALU-AI provides a structured framework that connects long-standing principles of transparent research reporting with the emerging realities of generative AI. This contribution extends the logic of reporting guidelines to a research environment in which authorship, process documentation, validation, and reproducibility must now account for machine-assisted forms of scholarly production (Moher et al., 2010).

4.4. Expert Consensus and Validation of the PETMALU-AI Framework

The Delphi findings provide preliminary empirical support for the content validity, clarity, and conceptual coherence of PETMALU-AI. This is important because the strength of a reporting framework depends on its conceptual relevance as well as on whether expert communities can understand, use, and accept it in practice. The high item-level content validity indices, strong final agreement rates, and favorable reliability estimates indicate that the final checklist items were consistently judged by the expert panel as relevant and appropriate for documenting AI-assisted research writing. These results suggest that the checklist was theoretically defensible and recognizable to experts as a practical structure for organizing disclosure, accountability, validation, and reproducibility concerns.

The use of a modified Delphi process is especially appropriate in this context because AI-assisted research writing remains an emerging and unsettled area of scholarly practice. The Delphi technique is commonly used to build expert consensus on complex issues characterized by uncertainty, incomplete standardization, and the need for multidisciplinary judgment (Crompton et al., 2026; Hasson et al., 2025). These conditions closely describe the current state of AI-assisted academic writing, where publisher policies, disciplinary expectations, and reporting practices continue to evolve. In this sense, the modified Delphi process was a validation procedure and a methodological response to the instability of the field. It allowed the framework to be refined through structured expert judgment rather than through the author's interpretation of the literature alone.

The significance of the findings is also reflected in the nature of the refinements made across rounds. The expert panel did not simply endorse the preliminary framework without qualification. Instead, the panel helped sharpen the distinction between disclosure and accountability, clarify the role of prompting as a methodological process, and reduce conceptual overlap among validation, limitations, and reproducibility items. These refinements strengthened the framework because they addressed precisely the areas where AI-assisted writing creates ambiguity. For instance, a manuscript may disclose AI use but still fail to identify who verified the outputs, how prompts shaped the response, whether generated content was checked against source materials, or why exact reproducibility may be difficult. Consequently, the feedback improved the checklist by making its domains more operationally distinct and more closely aligned with the actual reporting problems created by generative AI systems. This positive consequence is consistent with Delphi-based refinement practices that use expert feedback to improve clarity, relevance, and practical applicability (Diamond et al., 2014; Gattrell et al., 2024).

These findings are consistent with recent developments in consensus-based reporting guideline construction. The ACCORD reporting guideline, for example, was developed through a modified Delphi process and illustrates how expert consensus can be used to refine reporting items for complex methodological practices (Gattrell et al., 2024). Similarly, DELPHISTAR was developed to improve the reporting of Delphi studies in health and

social sciences, underscoring the broader movement toward greater transparency and methodological rigor in consensus-based research (Niederberger et al., 2024). In the AI domain, the CHART statement was also developed through systematic review, Delphi consensus, panel meetings, and pilot testing to support transparent reporting of studies involving generative AI-driven chatbots (Huo et al., 2025). These examples show that consensus methods remain central to the creation of reporting standards in areas where methodological practices are heterogeneous, rapidly evolving, or insufficiently standardized.

The PETMALU-AI Delphi findings should nevertheless be interpreted with appropriate caution. Expert agreement provides evidence that the checklist items are relevant, clear, and conceptually acceptable, but it does not by itself establish that the checklist will improve disclosure quality, reviewer confidence, or editorial consistency in actual publication workflows. This distinction is important because reporting guideline development is typically an iterative process. A framework may first be validated for content and structure, then subsequently evaluated for usability, implementation fidelity, inter-rater consistency, and impact on reporting quality. PETMALU-AI should therefore be understood as a validated reporting framework in its initial development stage, not as a fully tested editorial instrument. Future implementation studies are needed to examine how authors apply the checklist, how reviewers interpret checklist-based disclosures, and whether editors find the framework useful in evaluating AI-assisted manuscripts.

At the same time, the convergence of expert ratings and qualitative refinements provides a strong foundation for such future work. The interdisciplinary composition of the panel was particularly valuable because PETMALU-AI sits at the intersection of research methodology, AI systems, scholarly publishing, and research ethics. A checklist developed from only one of these perspectives would risk overemphasizing one dimension while neglecting others. For example, a technically oriented framework might focus heavily on tool identification and model details, while an ethics-oriented framework might emphasize authorship and accountability without sufficient attention to prompts, validation, or reproducibility. The Delphi process helped balance these concerns by requiring the framework to be intelligible across multiple expert communities.

Overall, the Delphi findings support the argument that PETMALU-AI offers a coherent and expert-validated structure for reporting AI-assisted research writing. Its value lies not only in the final set of retained items but also in the refinement process that clarified how those items should function as a unified framework. The strong consensus outcomes indicate that the framework captures concerns that experts recognize as relevant to transparent and accountable AI use. However, its future value will depend on continued testing in real-world scholarly workflows. In this respect, the Delphi results should be viewed as an important validation milestone rather than the endpoint of framework development.

4.5. Implications for Research, Peer Review, and Editorial Practice

The framework has several practical implications. For researchers, PETMALU-AI provides a structured way to disclose AI use without reducing reporting to vague declarations that offer little interpretive value. By prompting authors to specify whether AI was used, what role it played, how outputs were validated, and what constraints affect reproducibility, the checklist may improve the interpretability of AI-assisted manuscripts and reduce ambiguity for readers. For peer reviewers, the framework may function as an evaluative aid by making AI involvement more visible and methodologically legible. One of the difficulties in reviewing AI-assisted work is that AI influence may remain hidden behind polished prose, creating uncertainty about which portions of a manuscript reflect human interpretation, machine assistance, or a hybrid process. Consequently, A

structured reporting instrument such as PETMALU-AI may support more informed review by clarifying the extent and nature of AI contribution.

This point is especially important because recent evidence suggests that transparency alone does not automatically produce positive evaluative outcomes. Schilke and Reimann (2025), for instance, found that disclosure of AI use can reduce trust in the user, indicating that transparency may sometimes be interpreted as a signal of diminished legitimacy when it is presented without sufficient contextual detail. This creates a governance dilemma for AI-assisted scholarship. If disclosure is implemented only as a minimal declaration, readers, reviewers, or editors may treat AI use as a proxy for diminished rigor even when authors exercised careful judgment and verification. Conversely, if disclosure is accompanied by information about the purpose of AI use, the materiality of its contribution, the validation procedures applied, and the human authors responsible for the final content, transparency becomes more than an admission of tool use. It becomes an accountability mechanism. PETMALU-AI is therefore designed to reduce the interpretive ambiguity surrounding AI disclosure by shifting attention from whether AI was used to how it was used, governed, and checked. This distinction is important because responsible AI governance should not discourage transparent reporting through reputational penalty but should create reporting conditions under which appropriate AI use can be evaluated fairly and in relation to the safeguards implemented by human researchers.

For journals and publishers, the framework offers a possible bridge between broad policy principles and actual reporting practice. Current publisher guidance tends to define responsibilities at a high level, but implementation remains inconsistent, as illustrated by prior analyses of editorial AI policies (Laffaye et al., 2025). Alduais et al. (2025) similarly showed that AI-related guidance commonly emphasizes disclosure, academic integrity, transparency, and human accountability, but these principles often require clearer operational mechanisms for consistent implementation. PETMALU-AI may help address this gap by providing a concrete framework that journals can adapt as a submission checklist, supplementary disclosure form, or reviewer prompt. To support such implementation, Appendix D provides an author disclosure template, while Appendix E provides a reviewer and editor evaluation form. Nevertheless, such implementation should be understood as a mechanism for scrutiny rather than endorsement. A structured checklist does not make AI-assisted writing automatically acceptable from an ethical, methodological, or editorial standpoint. Instead, it provides a basis for determining whether AI use was minor or material, adequately verified or insufficiently checked, ethically appropriate or potentially problematic, and reproducible or difficult to audit.

A further implementation consideration concerns proportional reporting. Although PETMALU-AI is comprehensive, it is not intended to impose the same reporting burden on all manuscripts regardless of the extent of AI involvement. Minor uses of AI, such as language polishing or formatting support, may require only concise disclosure, while more substantive uses involving drafting, data transformation, prompt-based analysis, or interpretive support require more detailed documentation. Similarly, items related to prompt logs, interaction records, configuration settings, and materials availability should be applied in ways that are feasible, ethical, and context-sensitive. Full disclosure may not always be possible when platforms restrict access to settings, when interaction logs contain confidential information, or when AI use is limited to low-materiality editorial assistance. In such cases, authors should report the available details, explain any constraints, and preserve the central principles of transparency, accountability, and human verification. To operationalize this approach, Table 7 provides a practical guide for determining the expected depth of reporting based on the materiality of AI use. The guide is intended to

help authors, reviewers, and editors distinguish between minor editorial assistance and more substantive forms of AI involvement that may require fuller documentation.

Table 7. Proportional Reporting Guide for Applying PETMALU-AI.

Level of AI Use	Typical Examples	Recommended Reporting Depth	Items Most Likely to Apply
Minimal editorial assistance	Grammar checking, spelling correction, formatting suggestions, sentence-level polishing	Brief disclosure may be sufficient, especially if required by the journal. Authors should still confirm that the AI tool did not affect the study design, analysis, interpretation, or conclusions.	P1, P2, P3, P8, P20, P25
Moderate writing assistance	Reorganizing paragraphs, improving flow, summarizing author-provided notes, helping draft non-analytical sections	More detailed disclosure is recommended because AI may have shaped presentation, emphasis, or structure. Authors should describe the scope of use, output curation, and human verification.	P1–P4, P6–P8, P12–P15, P20, P22, P28–P30
Substantive research-process assistance	Assistance with coding, literature synthesis, data cleaning, qualitative categorization, statistical interpretation, or analytical planning	Detailed reporting is expected because AI may have influenced methodological or interpretive decisions. Authors should document prompts, validation, accountability, and limitations.	P1–P15, P20–P23, P24–P30, P31–P33
Data-generative or data-transformative assistance	Synthetic data generation, data augmentation, AI-generated examples, simulated cases, AI-assisted coding of empirical material	Full reporting is recommended because AI may affect the evidence base itself. Authors should document data origin, justification, generation process, safeguards, validation, and availability.	P1–P33, with particular emphasis on P16–P23, P26–P27, and P31–P33
No AI use	No generative AI tool was used in the research or manuscript preparation process	A statement may be included only if required by the journal or useful for transparency. PETMALU-AI is otherwise not applicable.	Not applicable, except a brief non-use statement if required

Note. This guide is intended to support proportional application of PETMALU-AI. Not all items will apply to every manuscript. Authors should report what is relevant, explain unavailable information when necessary, and avoid including sensitive materials in the name of transparency.

As shown in Table 7, proportional reporting does not mean selective transparency. Rather, it means that the depth of reporting should correspond to the degree to which AI may have influenced the manuscript, research process, data, analysis, interpretation, or reproducibility. This method helps prevent unnecessary reporting burden for low-materiality use while preserving stronger documentation expectations for AI involvement that may affect scholarly judgment or the evidentiary basis of the study.

This proportional approach is also important because recent concerns about hallucinated citations and unverifiable AI-generated claims demonstrate that the problem is not whether AI was used, but whether its outputs were critically checked before entering the scholarly record. Reports of hallucinated references and related integrity risks reinforce the need for structured verification rather than blanket acceptance or blanket rejection of

AI-assisted writing. Consequently, PETMALU-AI positions transparency as part of research quality assurance by helping journals evaluate AI use proportionately rather than treating all disclosure as either harmless compliance or evidence of misconduct.

4.6. Limitations and Future Directions

Several limitations warrant consideration, each of which also identifies a direction for future development. First, although the expert panel was deliberately multidisciplinary, it cannot be assumed to represent the full range of disciplinary contexts in which AI-assisted research writing is now emerging. The salience, interpretive value, and practical applicability of specific reporting items may differ across fields with distinct methodological cultures, publication conventions, and levels of AI integration. Further validation is needed across diverse disciplines, journal types, and research traditions to determine whether PETMALU-AI requires contextual adaptation, discipline-specific examples, or field-specific implementation guidance. Moreover, although the Delphi panel included experts from multiple countries, participants from the Philippines constituted the largest subgroup in the initial panel. Further validation involving more geographically balanced expert panels would strengthen the framework's international generalizability.

Second, the present validation was expert-based rather than implementation-based. Accordingly, while the Delphi process provides evidence for the content validity, clarity, and conceptual coherence of the framework, it does not yet establish how PETMALU-AI will function in routine manuscript preparation, peer review, or editorial decision-making. Future studies should examine the checklist in real-world scholarly workflows, including author use during manuscript preparation, reviewer interpretation during peer review, and editorial application during submission screening. Such studies would help determine whether PETMALU-AI improves disclosure quality, reviewer confidence, editorial consistency, and accountability in actual publication contexts.

Third, the present study did not systematically evaluate how existing AI disclosure policies are implemented, enforced, or experienced in journal workflows. This was outside the scope of the checklist development phase, but it remains an important limitation because the existence of a policy does not necessarily ensure consistent disclosure, interpretation, or enforcement. Future research should examine the gap between AI policy design and editorial practice, including whether structured tools such as PETMALU-AI can help translate broad policy principles into more consistent reporting expectations, reviewer prompts, and editorial decisions (Ganjavi et al., 2024; Laffaye et al., 2025).

Fourth, some reporting domains, particularly those related to prompts, interaction logs, configuration settings, and materials sharing, may be affected by feasibility constraints, confidentiality concerns, proprietary platform restrictions, or journal-specific disclosure norms. These constraints mean that complete documentation may not always be possible or appropriate in every research context. Future work should refine guidance on proportional reporting, including when full documentation is necessary, when partial reporting is sufficient, and how authors should explain unavailable or restricted AI-related materials without compromising privacy, confidentiality, or intellectual property.

Finally, although Appendix A provides a provisional scoring rubric, additional validation is needed before the rubric can be recommended as a formal evaluative instrument. The present study validated the checklist structure, not the scoring rubric as an independent assessment tool. Future studies should therefore examine inter-rater reliability when different authors, reviewers, or editors apply the rubric to the same manuscripts, as well as the extent to which rubric scores correspond to perceived reporting completeness, reviewer confidence, and editorial usefulness. These directions position PETMALU-AI not as a

fixed endpoint, but as an evolving reporting infrastructure capable of maturing alongside changes in generative AI systems, editorial workflows, and governance expectations.

5. Conclusions

As AI becomes increasingly embedded in the production of scholarly knowledge, the future of research will depend not only on what can be generated but also on what can be justified, traced, and trusted. PETMALU-AI was developed in response to this emerging imperative. It provides a structured and Delphi-validated framework for reporting AI-assisted research writing, addressing a critical gap between the rapid normalization of generative AI and the continued absence of standardized mechanisms for documenting its role in scholarly work. The principal outcome of this study is a 33-item checklist organized across nine reporting domains: disclosure and scope of AI use, human accountability and contribution, AI system transparency, prompting and interaction processes, data and AI-generated content, validation and quality assurance, ethics and compliance, limitations and reflexivity, and reproducibility and open practices. The modified Delphi validation supported the content validity, conceptual coherence, and practical interpretability of the framework, as reflected in strong expert agreement and favorable reliability indices.

The significance of PETMALU-AI is its recognition that AI use in research is no longer a peripheral or technical matter but a methodological, epistemic, and ethical concern. When AI systems help shape drafts, synthesize ideas, refine language, generate data-related outputs, or influence interpretive framing, their role must be rendered visible in ways that preserve human accountability and protect the integrity of the research record. PETMALU-AI translates broad principles of disclosure, accountability, validation, and reproducibility into specific reporting items that authors, reviewers, and editors can use in manuscript preparation and evaluation. Importantly, the framework does not promote uncritical AI use in research writing. Instead, it raises the evidentiary burden when AI is used by requiring proportional reporting, human verification, accountability assignment, and attention to reproducibility constraints. If the future of scholarship is to be co-authored by intelligent systems, then the future of research integrity must be designed with equal intelligence and even greater care. PETMALU-AI represents a step toward that future by making AI involvement visible, assessable, and open to scrutiny without compromising the clarity, responsibility, and rigor on which science and scholarship depend.

Funding: This research received no external funding.

Institutional Review Board Statement: Formal ethics approval was not required because the study involved expert consultation for checklist development and validation and did not involve vulnerable populations, intervention, or collection of sensitive personal data. The study adhered to widely accepted ethical research standards, including voluntary participation, informed consent, anonymity, confidentiality of responses, and the right to withdraw at any time.

Informed Consent Statement: Informed consent was obtained from all study participants.

Data Availability Statement: Data are available upon reasonable request.

Acknowledgments: The author gratefully acknowledges the expert panelists who participated in the Delphi validation process and contributed their time, expertise, and professional judgment to the refinement of the PETMALU-AI checklist. During the revision of this manuscript, the author used ChatGPT (v5.1) by OpenAI to assist with language refinement, organization of reviewer responses, and editorial polishing. Reflecting the reporting principles proposed in the framework, AI assistance was limited to manuscript refinement and did not determine the study design, data interpretation, checklist validation outcomes, or scholarly conclusions. The author reviewed, verified, and revised all AI-assisted outputs and takes full responsibility for the content of this publication.

Conflicts of Interest: The author declares no conflicts of interest.

Appendix A. PETMALU-AI Checklist with Scoring Rubric

This appendix provides a provisional scoring rubric intended to support formative use during manuscript preparation and checklist familiarization. The rubric uses a three-point scoring structure in which a score of 0 indicates that the item is not reported, 1 indicates that the item is partially reported or implicitly addressed, and 2 indicates that the item is fully and explicitly reported. However, the present study validated the content and structure of the checklist rather than the scoring rubric as an independent evaluative instrument. Accordingly, the rubric should be interpreted as an implementation aid rather than a validated scoring tool for formal editorial decision-making. Further studies are needed to examine inter-rater consistency, usability, and decision validity when the rubric is applied by authors, reviewers, or editors to actual manuscripts.

Table A1. PETMALU-AI Provisional Scoring Rubric for Formative Checklist Use.

Item	Checklist Item	Description	Reported (Yes/No)	Score (0–2)	Page
Section 1: Disclosure and Scope of AI Use					
P1	AI Use Disclosure Statement	Provide a clear statement describing the use of AI in the study			
P2	Scope of AI Involvement	Specify the functional areas where AI was used in the research process			
P3	Materiality of AI Use	Indicate whether AI use was minor support or a substantial contribution			
Section 2: Human Accountability and Contribution					
P4	Human–AI Contribution Delineation	Distinguish between human contributions and AI-assisted outputs			
P5	Authorship Compliance	Confirm that AI systems are not listed as authors in the manuscript			
P6	Accountability Assignment	Identify the author(s) responsible for AI-assisted content and outputs			
P7	Prompt Design Responsibility	Specify who designed, refined, and validated the prompts used			
Section 3: AI System Transparency					
P8	AI Tool Identification	Identify all AI tools used in the research and writing process			
P9	Model/Version Disclosure	Specify the AI model or version used where relevant to the study			
P10	Access Context	Describe how the AI system was accessed or integrated into the workflow			
P11	Configuration Disclosure	Report relevant system settings if they influence generated outputs			
Section 4: Prompting and Interaction Process					
P12	Prompting Approach Description	Describe the general prompting strategy used to generate outputs			
P13	Iterative Use Disclosure	Indicate whether outputs were generated through iterative refinement			
P14	Output Curation Process	Explain how AI-generated outputs were selected, edited, or rejected			
P15	Interaction Records Availability	Indicate whether interaction logs are available when necessary			

Table A1. *Cont.*

Item	Checklist Item	Description	Reported (Yes/No)	Score (0–2)	Page
Section 5: Data and AI-Generated Content					
P16	Data Origin Transparency	Classify data sources as human-generated, AI-generated, or AI-augmented			
P17	Justification for AI Data	Provide justification for the use of AI-generated or synthetic data			
P18	Data Generation Description	Describe how AI was used to generate or modify data in the study			
P19	Data Availability	Indicate whether datasets or data generation protocols are accessible			
Section 6: Validation and Quality Assurance					
P20	Human Verification Process	Describe how AI-generated outputs were reviewed and verified by humans			
P21	Validation Strategy	Report any formal validation procedures applied to AI-assisted outputs			
P22	Error and Bias Consideration	Acknowledge potential errors, biases, or hallucinations in AI outputs			
P23	Robustness Checks	Report consistency or sensitivity checks across prompts or system runs			
Section 7: Ethics and Compliance					
P24	Ethical Considerations	Address relevant ethical implications of using AI in the study			
P25	Policy Compliance Statement	Confirm compliance with institutional or publisher AI guidelines			
P26	Bias Mitigation	Describe any steps taken to mitigate bias in AI-assisted processes			
P27	Privacy Safeguards	Ensure that no sensitive or identifiable data were improperly used			
Section 8: Limitations and Reflexivity					
P28	AI-Related Limitations	Discuss limitations introduced by the use of AI in the research			
P29	Reflexive Consideration	Reflect on how AI use may have influenced study outcomes or interpretations			
P30	Reproducibility Constraints	Address challenges related to reproducibility due to AI variability			
Section 9: Reproducibility and Open Practices					
P31	Reproducibility Information	Provide sufficient detail to allow understanding of AI-assisted processes			
P32	Materials Availability	Share prompts, workflows, or materials where feasible and appropriate			
P33	System Evolution Disclaimer	Note that outputs may vary due to changes in AI systems over time			

Appendix B. PETMALU-AI Item Development and Refinement Trail

This appendix summarizes how the PETMALU-AI checklist moved from the initial evidence-informed item pool to the final 33-item structure. The development process was not treated as a simple accumulation of disclosure items. Instead, candidate items were

reviewed for conceptual overlap, reporting usefulness, feasibility, and alignment with the intended scope of the framework. Items that were too narrow, repetitive, or difficult to apply across disciplines were merged or reworded. Items that addressed emerging but underreported issues, such as prompt design, interaction records, system evolution, and AI-generated data, were retained when they were judged to contribute meaningfully to transparency, accountability, or reproducibility.

Table A2. Summary of Item Refinement from Candidate Concerns to Final PETMALU-AI Items.

Item	Initial Reporting Concern	Refinement Decision
P1	Authors should declare whether AI was used.	Retained as a core disclosure item.
P2	Authors should state which parts of the manuscript involve AI assistance.	Retained and broadened to include research workflow stages.
P3	Minor language editing should be distinguished from substantive AI contribution.	Consolidated into a materiality item to support proportional reporting.
P4	Human and AI contributions should not be blurred.	Retained as a separate accountability item.
P5	AI systems should not be listed as authors.	Retained because of strong alignment with publisher and authorship guidance.
P6	Someone should be responsible for checking AI-assisted content.	Retained and framed as accountability assignment.
P7	Prompt authorship or responsibility should be documented.	Retained because prompt design can influence outputs and interpretations.
P8	The AI tool should be named.	Retained as a basic traceability requirement.
P9	Model or version should be reported when available.	Retained but worded flexibly because not all platforms provide stable version information.
P10	Authors should describe how the AI system was accessed.	Retained to distinguish web-based, API-based, embedded, or institutional access.
P11	AI settings and parameters should be reported.	Retained but limited to relevant configurations to avoid unnecessary reporting burden.
P12	Prompts should be disclosed.	Refined into a broader prompting approach item rather than requiring full prompt disclosure in all cases.
P13	Repeated prompt-output cycles should be documented.	Retained as iterative use disclosure.
P14	Authors should explain how AI outputs were selected or rejected.	Retained as output curation process.
P15	Prompt logs should be shared.	Reworded as interaction records availability because full sharing may not always be feasible or ethical.
P16	Data sources should be clear when AI is involved.	Retained as data origin transparency.
P17	Synthetic or AI-generated data should be justified.	Retained as justification for AI data.
P18	AI-generated or modified data should be described.	Retained as data generation description.
P19	AI-generated datasets should be accessible where possible.	Retained as data availability, with flexibility for privacy and proprietary constraints.
P20	AI outputs should be checked by humans.	Retained as human verification process.
P21	AI-assisted outputs should be validated.	Retained as validation strategy.
P22	Hallucinations, omissions, citation errors, and bias should be addressed.	Consolidated into error and bias consideration.
P23	Outputs should be tested across prompts or runs.	Retained as robustness checks where relevant.
P24	Ethical issues should be reported.	Retained as ethical considerations.
P25	Authors should comply with institutional and publisher policies.	Retained as policy compliance statement.
P26	Bias introduced by AI should be mitigated.	Retained as bias mitigation.
P27	Sensitive or identifiable information should be protected.	Retained as privacy safeguards.
P28	AI-related limitations should be discussed.	Retained as AI-related limitations.
P29	Authors should reflect on how AI shaped the manuscript.	Retained as reflexive consideration.
P30	Exact reproduction may be difficult because AI systems vary.	Retained as reproducibility constraints.
P31	Authors should provide enough workflow information for readers to understand AI use.	Retained as reproducibility information.
P32	Supporting materials should be shared where possible.	Retained as materials availability.
P33	AI systems may change over time.	Retained as system evolution disclaimer.

Note. The table summarizes the logic of refinement rather than a complete audit trail of every wording change. Its purpose is to make the checklist development process more transparent while avoiding the impression that the final items emerged fully formed.

Appendix C. Additional Delphi Validation Summary

This appendix provides a domain-level summary of the Delphi validation outcomes. This summary complements the aggregate validation results reported in Table 2 by showing how consensus was distributed across the nine PETMALU-AI domains.

Table A3. Domain-Level Delphi Validation Summary for PETMALU-AI.

Domain	Number of Items	Round 3 I-CVI Range	Round 3 Agreement Range	Interpretation
Disclosure and Scope of AI Use	3	0.90–1.00	90–100%	Strong consensus
Human Accountability and Contribution	4	0.85–1.00	85–100%	Strong consensus
AI System Transparency	4	0.85–0.95	85–95%	Strong consensus
Prompting and Interaction Process	4	0.85–0.95	85–95%	Strong consensus
Data and AI-Generated Content	4	0.85–0.95	85–95%	Strong consensus
Validation and Quality Assurance	4	0.90–1.00	90–100%	Strong consensus
Ethics and Compliance	4	0.90–1.00	90–100%	Strong consensus
Limitations and Reflexivity	3	0.85–0.95	85–95%	Strong consensus
Reproducibility and Open Practices	3	0.85–0.95	85–95%	Strong consensus

Note. I-CVI refers to the item-level content validity index, calculated as the proportion of experts rating an item as relevant or highly relevant. Agreement refers to the percentage of experts endorsing retention or adequacy of the item in the final Delphi round. The values are reported at the domain level to provide transparency without overextending the precision of the validation results.

Appendix D. PETMALU-AI Author Disclosure Template

This appendix is intended to help authors prepare a structured AI-use disclosure. It may be submitted as a manuscript statement, supplementary file, or internal author checklist, depending on journal requirements. Authors do not need to complete every field when an item is not applicable. However, omitted or unavailable information should be briefly explained when the AI use was substantive.

Table A4. Author Disclosure Template for AI-Assisted Research Writing.

Disclosure Component	Author Response
Was generative AI used in the preparation of this manuscript or in any part of the research process?	
AI tool(s) used	
Model, version, or system release, if known	
Date or period of AI use	
Access context, such as web interface, API, institutional platform, or integrated software	
Purpose of AI use	
Manuscript sections or research stages affected by AI use	
Whether AI use was minor, moderate, or substantive	

Table A4. *Cont.*

Disclosure Component	Author Response
Whether AI influenced study design, data collection, analysis, interpretation, or conclusions	
Person(s) responsible for prompt design and refinement	
General prompting approach used	
Whether iterative prompt-output cycles were used	
How AI-generated outputs were selected, revised, rejected, or integrated	
Human verification process	
Validation or fact-checking procedures used	
Whether AI was used to generate, transform, augment, summarize, or analyze data	
Data privacy or confidentiality safeguards	
Bias, hallucination, or error-checking procedures	
Availability of prompts, logs, workflows, or templates	
AI-related limitations	
Human author(s) responsible for the final content	
Compliance with journal, publisher, institutional, or professional AI policies	

Appendix E. PETMALU-AI Reviewer and Editor Evaluation Form

This form is intended to support peer reviewers and editors in evaluating the completeness and interpretability of AI-use reporting. It should not be used as an automatic acceptance or rejection tool. Rather, it provides a structured way to identify whether AI use has been reported clearly enough for the manuscript to be evaluated fairly.

Table A5. Reviewer and Editor Evaluation Form for PETMALU-AI Reporting.

Evaluation Area	Guiding Question	Adequately Reported?	Reviewer or Editor Notes
Disclosure	Does the manuscript clearly state whether AI was used?	Yes/Partly/No/Not applicable	
Scope	Does the manuscript identify where and for what purpose AI was used?	Yes/Partly/No/Not applicable	
Materiality	Does the manuscript distinguish minor editorial assistance from substantive AI involvement?	Yes/Partly/No/Not applicable	
Human accountability	Are responsible human authors clearly identified?	Yes/Partly/No/Not applicable	
Authorship compliance	Is it clear that no AI system is treated as an author?	Yes/Partly/No/Not applicable	
Tool transparency	Are the AI tool, model, version, or access context reported where relevant?	Yes/Partly/No/Not applicable	
Prompting process	Is the prompting approach described sufficiently for readers to understand how outputs were produced?	Yes/Partly/No/Not applicable	
Output curation	Does the manuscript explain how AI outputs were reviewed, selected, edited, or rejected?	Yes/Partly/No/Not applicable	
Data provenance	If AI was used with data, is the origin and treatment of the data clear?	Yes/Partly/No/Not applicable	
Validation	Are verification, fact-checking, or validation procedures reported?	Yes/Partly/No/Not applicable	

Table A5. *Cont.*

Evaluation Area	Guiding Question	Adequately Reported?	Reviewer or Editor Notes
Bias and error management	Does the manuscript address possible hallucinations, bias, omissions, or citation errors?	Yes/Partly/No/Not applicable	
Ethics and privacy	Are privacy, confidentiality, fairness, and policy compliance addressed?	Yes/Partly/No/Not applicable	
Limitations	Does the manuscript discuss limitations introduced by AI use?	Yes/Partly/No/Not applicable	
Reproducibility	Are prompts, workflows, logs, or relevant materials available or explained where appropriate?	Yes/Partly/No/Not applicable	
Overall judgment	Is the AI-use reporting sufficiently transparent for review or editorial assessment?	Yes/Partly/No	

Note. When AI reporting is incomplete, but AI use appears minor, a concise disclosure revision may be sufficient. When AI reporting is incomplete and AI use appears substantive, a fuller PETMALU-AI disclosure may be requested before methodological or editorial evaluation proceeds.

References

Acut, D. P., Malabago, N. K., Malicoban, E. V., Galamiton, N. S., & Garcia, M. B. (2025). "ChatGPT 4.0 ghosted us while conducting literature search:" Modeling the chatbot's generated non-existent references using regression analysis. *Internet Reference Services Quarterly*, *29*(1), 27–54. [CrossRef]

Alduais, A., Qadhi, S., Chaaban, Y., & Khraisheh, M. (2025). Utilizing generative AI responsibly and ethically for research purposes in higher education: A policy analysis. *Serials Review*, *51*(3–4), 120–170. [CrossRef]

Arksey, H., & O'Malley, L. (2005). Scoping studies: Towards a methodological framework. *International Journal of Social Research Methodology*, *8*(1), 19–32. [CrossRef]

Boretti, A. (2026). Hallucinations in generative AI: A threat to scholarly integrity and the urgent need for publisher-led academically supervised verification. *Energy Research & Social Science*, *136*, 104720. [CrossRef]

Bozkurt, A. (2024). GenAI et al.: Cocreation, authorship, ownership, academic ethics and integrity in a time of generative AI. *Open Praxis*, *16*(1), 1–10. [CrossRef]

Bozkurt, A., Xiao, J., Farrow, R., Bai, J. Y. H., Nerantzi, C., Moore, S., Dron, J., Stracke, C. M., Singh, L., Crompton, H., Koutropoulos, A., Terentev, E., Pazurek, A., Nichols, M., Sidorkin, A. M., Costello, E., Watson, S., Mulligan, D., Honeychurch, S., … Asino, T. I. (2024). The manifesto for teaching and learning in a time of generative AI: A critical collective stance to better navigate the future. *Open Praxis*, *16*(4), 487–513. [CrossRef]

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101. [CrossRef]

Calderon, R., & Herrera, F. (2025). And Plato met ChatGPT: An ethical reflection on the use of chatbots in scientific research writing, with a particular focus on the social sciences. *Humanities and Social Sciences Communications*, *12*(1), 713. [CrossRef]

Cheng, A., Calhoun, A., & Reedy, G. (2025). Artificial intelligence-assisted academic writing: Recommendations for ethical use. *Advances in Simulation*, *10*(1), 22. [CrossRef] [PubMed]

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. [CrossRef] [PubMed]

Cohen, J. F., & Moher, D. (2025). Generative artificial intelligence and academic writing: Friend or foe? *Journal of Clinical Epidemiology*, *179*, 111646. [CrossRef] [PubMed]

Crompton, H., Burke, D., Nickel, C., Bozkurt, A., Miao, F., Sharples, M., Greene, J. A., Parsons, D., Gill-Simmen, L., Edmett, A., Pegrum, M., de Waard, I., Bonk, C. J., Garcia, M. B., Curry, J. H., Lindsey, L., Yang, M., Marshall, S., Bali, M., … Yu, S. (2026). Governing generative AI in higher education: A global Delphi study on policy and practice. *International Journal of Educational Technology in Higher Education*, *23*(1), 21. [CrossRef]

Diamond, I. R., Grant, R. C., Feldman, B. M., Pencharz, P. B., Ling, S. C., Moore, A. M., & Wales, P. W. (2014). Defining consensus: A systematic review recommends methodologic criteria for reporting of Delphi studies. *Journal of Clinical Epidemiology*, *67*(4), 401–409. [CrossRef] [PubMed]

Dorta-González, P., López-Puig, A. J., Dorta-González, M. I., & González-Betancor, S. M. (2024). Generative artificial intelligence usage by researchers at work: Effects of gender, career stage, type of workplace, and perceived barriers. *Telematics and Informatics*, *94*, 102187. [CrossRef]

Dwivedi, Y. K., Kshetri, N., Hughes, L., Slade, E. L., Jeyaraj, A., Kar, A. K., Baabdullah, A. M., Koohang, A., Raghavan, V., Ahuja, M., Albanna, H., Albashrawi, M. A., Al-Busaidi, A. S., Balakrishnan, J., Barlette, Y., Basu, S., Bose, I., Brooks, L., Buhalis, D., . . . Wright, R. (2023). Opinion paper: "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*, 102642. [CrossRef]

Elsevier. (2025). *Generative AI policies for journals*. Elsevier. Available online: https://www.elsevier.com/about/policies-and-standards/generative-ai-policies-for-journals (accessed on 15 June 2026).

Fleiss, J. L. (1971). Measuring nominal scale agreement among many raters. *Psychological Bulletin*, *76*(5), 378–382. [CrossRef]

Ganjavi, C., Eppler, M. B., Pekcan, A., Biedermann, B., Abreu, A., Collins, G. S., Gill, I. S., & E Cacciamani, G. (2024). Publishers' and journals' instructions to authors on use of generative artificial intelligence in academic and scientific publishing: Bibliometric analysis. *BMJ*, *384*, e077192. [CrossRef] [PubMed]

Garcia, M. B. (2023). Using AI tools in writing peer review reports: Should academic journals embrace the use of ChatGPT? *Annals of Biomedical Engineering*, *52*(2), 139–140. [CrossRef] [PubMed]

Garcia, M. B. (2025). ChatGPT as an academic writing tool: Factors influencing researchers' intention to write manuscripts using generative artificial intelligence. *International Journal of Human–Computer Interaction*, *41*(24), 15514–15528. [CrossRef]

Gattrell, W. T., Logullo, P., van Zuuren, E. J., Price, A., Hughes, E. L., Blazey, P., Winchester, C. C., Tovey, D., Goldman, K., Hungin, A. P., & Harrison, N. (2024). ACCORD (ACcurate COnsensus Reporting Document): A reporting guideline for consensus methods in biomedicine developed via a modified Delphi. *PLoS Medicine*, *21*(1), e1004326. [CrossRef] [PubMed]

Hasson, F., Keeney, S., & McKenna, H. (2025). Revisiting the Delphi technique—Research thinking and practice: A discussion paper. *International Journal of Nursing Studies*, *168*, 105119. [CrossRef] [PubMed]

Hopewell, S., Chan, A., Collins, G. S., Hróbjartsson, A., Moher, D., Schulz, K. F., Tunn, R., Aggarwal, R., Berkwits, M., A Berlin, J., Bhandari, N., Butcher, N. J., Campbell, M. K., Chidebe, R. C. W., Elbourne, D., Farmer, A., A Fergusson, D., Golub, R. M., Goodman, S. N., . . . Boutron, I. (2025). CONSORT 2025 statement: Updated guideline for reporting randomised trials. *BMJ*, *389*, e081123. [CrossRef] [PubMed]

Hsu, C., & Sandford, B. A. (2007). The Delphi technique: Making sense of consensus. *Practical Assessment, Research, and Evaluation*, *12*(1), 10. [CrossRef]

Huo, B., Collins, G., Chartash, D., Thirunavukarasu, A., Flanagin, A., Iorio, A., Cacciamani, G., Chen, X., Liu, N., Mathur, P., Chan, A., Laine, C., Pacella, D., Berkwits, M., Antoniou, S. A., Camaradou, J. C., Canfield, C., Mittelman, M., Feeney, T., . . . Guyatt, G. (2025). Reporting guideline for chatbot health advice studies: The CHART statement. *Artificial Intelligence in Medicine*, *168*, 103222. [CrossRef] [PubMed]

International Committee of Medical Journal Editors. (2026). *Recommendations for the conduct, reporting, editing, and publication of scholarly work in medical journals*. Available online: https://www.icmje.org/icmje-recommendations.pdf (accessed on 15 June 2026).

Khalifa, M., & Albadawy, M. (2024). Using artificial intelligence in academic writing and research: An essential productivity tool. *Computer Methods and Programs in Biomedicine Update*, *5*, 100145. [CrossRef]

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, *15*(2), 155–163. [CrossRef] [PubMed]

Laffaye, T. A., Carlson, B. H., Freeman, W. K., & Ayoub, C. (2025). Artificial intelligence for manuscript writing: Policies and implementation in cardiovascular journals. *Cardiovascular Diagnosis and Therapy*, *15*(5), 1107–1112. [CrossRef] [PubMed]

Levac, D., Colquhoun, H., & O'BRien, K. K. (2010). Scoping studies: Advancing the methodology. *Implementation Science*, *5*(1), 69. [CrossRef] [PubMed]

Luo, X., Tham, Y. C., Giuffrè, M., Ranisch, R., Daher, M., Lam, K., Eriksen, A. V., Hsu, C., Ozaki, A., de Moraes, F. Y., Khanna, S., Su, K., Begagić, E., Bian, Z., Chen, Y., & Estill, J. (2025). Reporting guideline for the use of generative artificial intelligence tools in MEdical research: The GAMER statement. *BMJ Evidence-Based Medicine*, *30*(6), 390–400. [CrossRef] [PubMed]

Martínez-Requejo, S., Redondo-Duarte, S., Jiménez-García, E., & Ruiz-Lázaro, J. (2025). Technoethics and the use of artificial intelligence in educational contexts: Reflections on integrity, transparency, and equity. In *Pitfalls of AI integration in education: Skill obsolescence, misuse, and bias*. IGI Global. [CrossRef]

McHugh, M. L. (2012). Interrater Reliability: The Kappa Statistic. *Biochemia Medica*, *22*(3), 276–282. [CrossRef]

Mezzadri, D. (2025). The paradox of ethical AI-assisted research. *Journal of Academic Ethics*, *23*(4), 2653–2667. [CrossRef]

Moher, D., Schulz, K. F., Simera, I., & Altman, D. G. (2010). Guidance for developers of health research reporting guidelines. *PLoS Medicine*, *7*(2), e1000217. [CrossRef] [PubMed]

Niederberger, M., Schifano, J., Deckert, S., Hirt, J., Homberg, A., Köberich, S., Kuhn, R., Rommel, A., Sonnberger, M., & Network, T. D. (2024). Delphi studies in social and health sciences—Recommendations for an interdisciplinary standardized reporting (DELPHISTAR). Results of a Delphi study. *PLoS ONE*, *19*(8), e0304651. [CrossRef] [PubMed]

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., A Akl, E., E Brennan, S., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, *372*, n71. [CrossRef] [PubMed]

Park, J. (2025). Towards a transparent and reproducible AI-assisted research paper writing. *Genomics & Informatics*, *23*(1), 26. [CrossRef] [PubMed]

Peters, M. D. J., Godfrey, C. M., Khalil, H., McInerney, P., Parker, D., & Soares, C. B. (2015). Guidance for conducting systematic scoping reviews. *JBI Evidence Implementation*, *13*(3), 141–146. [CrossRef] [PubMed]

Polit, D. F., & Beck, C. T. (2006). The content validity index: Are you sure you know what's being reported? Critique and recommendations. *Research in Nursing & Health*, *29*(5), 489–497. [CrossRef] [PubMed]

Resnik, D. B., & Hosseini, M. (2026a). Disclosing artificial intelligence use in scientific research and publication: When should disclosure be mandatory, optional, or unnecessary? *Accountability in Research*, *33*(2), 2481949. [CrossRef] [PubMed]

Resnik, D. B., & Hosseini, M. (2026b). Hallucinated citations produced by generative artificial intelligence may constitute research misconduct when citations function as data in scholarly papers. *Accountability in Research*, 2645390. [CrossRef] [PubMed]

Schilke, O., & Reimann, M. (2025). The transparency dilemma: How AI disclosure erodes trust. *Organizational Behavior and Human Decision Processes*, *188*, 104405. [CrossRef]

Shimray, S. R., & Subaveerapandiyan, A. (2025). Artificial intelligence in academic writing and research: Adoption and effectiveness. *Open Information Science*, *9*(1), 20250026. [CrossRef]

Springer Nature. (2026). *Editorial policies.* Springer Nature. Available online: https://www.springernature.com/gp/policies/editorial-policies (accessed on 15 June 2026).

Taneja, D., Prabagaren, H., & Thomas, M. R. (2025). AI in academia: Balancing integrity, ethics, and learning amid evolving norms of authorship and scholarship. In *Pitfalls of AI integration in education: Skill obsolescence, misuse, and bias*. IGI Global. [CrossRef]

Tang, A., Li, K.-K., Kwok, K. O., Cao, L., Luong, S., & Tam, W. (2024). The importance of transparency: Declaring the use of generative artificial intelligence (AI) in academic writing. *Journal of Nursing Scholarship*, *56*(2), 314–318. [CrossRef] [PubMed]

Wiley. (2026). *Using AI tools in your research.* Wiley. Available online: https://www.wiley.com/en-nl/publish/article/ai-guidelines/ (accessed on 15 June 2026).

Xiao, J., Bozkurt, A., Nichols, M., Pazurek, A., Stracke, C. M., Bai, J. Y. H., Farrow, R., Mulligan, D., Nerantzi, C., Sharma, R. C., Singh, L., Frumin, I., Swindell, A., Honeychurch, S., Bond, M., Dron, J., Moore, S., Leng, J., van Tryon, P. J. S., … Themeli, C. (2025). Venturing into the Unknown: Critical insights into grey areas and pioneering future directions in educational generative AI research. *TechTrends*, *69*(3), 582–597. [CrossRef]

Yousaf, M. N. (2025). Practical considerations and ethical implications of using artificial intelligence in writing scientific manuscripts. *ACG Case Reports Journal*, *12*(2), e01629. [CrossRef] [PubMed]

Yusoff, M. S. B. (2019). ABC of content validation and content validity index calculation. *Education in Medicine Journal*, *11*(2), 49–54. [CrossRef]