ICRRD JOURNAL 2026, VOL 7, ISSUE 2, 181-189
ISSN: 2773-5958
https://doi.org/10.53272/icrrd
https://www.icrrd.com

INTERNATIONAL CENTER FOR RESEARCH AND RESOURCE DEVELOPMENT
ICRRD QUALITY INDEX RESEARCH JOURNAL
ISSN: 2773-5958, https://doi.org/10.53272/icrrd

Ethical Use of Artificial Intelligence in Formal Research Investigation: A Qualitative Case Study

Muhammad Reaz Hawladar[1*], Wan Sarah Binti Hossen[2]

[12] *Faculty of Sharia and Law, Universiti Sains Islam Malaysia (USIM), Negeri Sembilan, Malaysia*
**Corresponding Author Email: reazmh@gmail.com*

Received: 19 May 2026 • Accepted: 11 October 2026 • Published: 09 November 2026

Abstract

Generative artificial intelligence (AI) tools such as large language models are now used across the research cycle, from literature searching and coding qualitative data to drafting manuscripts. Their use promises efficiency but also raises problems of fabrication, bias, confidentiality, authorship and accountability. This article examines how ethical principles are, and are not, upheld when AI is used in formal research investigation, meaning the systematic, methodologically governed inquiry that produces knowledge claims for scholarly scrutiny. A qualitative multiple-case study design was used. Four publicly documented cases of AI use in research and scholarly publishing were analysed through document analysis and reflexive thematic analysis, and were compared against international ethics and integrity guidance. An illustrative composite case of a university research team is added to show how the identified themes operate in everyday practice. Five themes were developed: (1) verification as the central ethical duty; (2) transparency and disclosure gaps; (3) accountability that cannot be delegated; (4) confidentiality and data protection; and (5) bias, equity and the erosion of researcher judgement. Failures arose less from the technology itself than from weak human oversight and unclear institutional rules. Ethical use of AI in research depends on a human-in-the-loop model supported by disclosure, verification protocols, data governance and training. A practical framework (the V-T-A-C-J model) is proposed.

Keywords: *Generative artificial intelligence (AI, Ethics of AI, Framework of AI in research*

Introduction

Research is a trust-based enterprise. Society, funders and fellow scholars rely on the assumption that reported methods were actually followed, that data are genuine and that named authors stand behind their claims. The rapid spread of generative AI since late 2022 has strained that assumption. A tool that can produce fluent text, summarise papers, suggest codes for interview transcripts and generate images in seconds is attractive to overloaded researchers. Yet the same tool can invent references, reproduce hidden biases, expose confidential data to third

parties and produce text that looks authoritative without being grounded in evidence (Bender et al., 2021; Ji et al., 2023).

Scholarly bodies responded quickly. Major publishers and editorial associations stated that AI systems cannot be authors because they cannot take responsibility for work, and that use of AI should be disclosed (Thorp, 2023; Stokel-Walker, 2023; COPE, 2023; ICMJE, 2023). Broader frameworks such as the UNESCO Recommendation on the Ethics of Artificial Intelligence (2021), the ALLEA European Code of Conduct for Research Integrity (2023 revision) and the European Union AI Act (2024) added further.

expectations. In Malaysia, the National Guidelines on AI Governance and Ethics issued by the government in 2024 set principles that research institutions are expected to translate into local practice.

Despite this guidance, there is limited qualitative understanding of *how* ethical breakdowns occur in real research settings and what practices prevent them. Much of the literature is conceptual or based on surveys of attitudes. This article contributes an in-depth, case-based account of ethical use of AI in formal research investigation.

Problem statement

Researchers are adopting AI tools faster than institutions are developing enforceable norms. As a result, individual researchers must make ethical judgements at the point of use, often without clear standards, training or oversight.

Research questions

1. How have ethical failures and ethical safeguards been manifested in documented cases of AI use in research and scholarly publishing?
2. Which ethical principles are most frequently at stake?
3. What practices can support ethical use of AI in formal research investigation?

Scope and definitions

Formal research investigation is used here to mean inquiry that follows an explicit design, is subject to ethical oversight (for example, institutional review), generates or analyses evidence and is reported for scrutiny, including peer review. *AI* refers mainly to generative AI and large language model tools, together with machine-learning tools used for analysis.

Conceptual Background

Ethical principles guiding research

Traditional research ethics rests on respect for persons, beneficence and justice (the Belmont tradition), complemented by integrity norms such as honesty, accountability, objectivity and care (ALLEA, 2023). Reviews of AI ethics guidelines show convergence on transparency, justice and fairness, non-maleficence, responsibility and privacy (Jobin et al., 2019), and Floridi and Cowls (2019) propose a unified set of five principles: beneficence, non-maleficence, autonomy, justice and explicability. Resnik and Hosseini (2024) argue that using AI in science needs new, specific guidance because existing norms do not fully address issues such as opacity, hallucination and responsibility.

Synthesis for this study

From these sources, five sensitising concepts guided the analysis:

- **Verification:** the duty to check AI outputs against primary evidence.
- **Transparency:** honest disclosure of what AI was used for, and how.
- **Accountability:** human responsibility for all content and decisions.

- **Confidentiality:** protection of participants, unpublished work and sensitive data.
- **Justice and judgement:** fairness, bias awareness and preservation of the researcher's interpretive role.

These concepts were treated as starting points, not fixed categories, so that themes could emerge inductively.

Methodology

Research design

A qualitative, interpretivist, multiple-case study design was adopted. Case study is suited to contemporary phenomena where boundaries between phenomenon and context are blurred and where "how" and "why" questions are posed (Yin, 2018; Stake, 1995). A multiple-case design allows cross-case comparison and strengthens analytic generalisation, meaning that findings are generalised to theory and guidance, not to populations (Hossen et al., 2026).

Case selection

Purposive (information-oriented) sampling was used. Cases had to be (a) publicly documented by credible sources such as journals, courts, editorial statements or professional bodies; (b) clearly connected to AI use in research or scholarly publication; and (c) varied in type of ethical issue. Four cases met these criteria.

Table 1: Case setting with core issue

Case	Setting	Core issue
A. AI-generated figure in a peer-reviewed biomedical paper (2024)	Journal publishing	Fabricated or non-scientific imagery passing peer review; later retraction
B. Text artefacts of chatbot interfaces in a published physics paper (2023)	Journal publishing	Undisclosed AI text generation; failed editorial checks; retraction
C. Fabricated legal citations from a chatbot (*Mata v. Avianca*, 2023)	Legal investigation and filing	Unverified AI outputs presented as authority; judicial sanctions
D. Policy responses on AI authorship (COPE, ICMJE, major publishers, 2023)	Governance	How institutions define acceptable use and authorship

Case C comes from the legal profession, not academic research. It was included deliberately because it is a formal investigation context with a well-documented record of verification failure, and its lessons transfer directly to research. Case D is a governance case that acts as a counterpoint to the failure cases.

Data sources

Data consisted of public documents: retraction notices, editorial statements, court opinions, news and scholarly commentary, and policy texts. Only information available in the public record was used, and no personal data beyond what is in those records was collected. Because the study analysed public documents, formal review by a human-participant ethics committee may not be required, but researchers should confirm this with their own institution.

Illustrative composite case

To show how the themes play out in ordinary practice, a composite scenario (Case E) is presented in the Discussion. **It is a constructed illustration, not empirical data.** It is informed by patterns across Cases A to D and by common research workflows. If you intend to present this as empirical work, replace or supplement Case E with real data, for example semi-structured interviews with researchers and research-ethics committee members, collected with

ethical approval and informed consent.

Data analysis

Reflexive thematic analysis (Braun and Clarke, 2006) was applied in six phases: familiarisation with documents; generating initial codes; searching for themes; reviewing themes against coded extracts and the full dataset; defining and naming themes; and producing the report. Coding was done both within each case and across cases. A cross-case matrix compared how each ethical principle was upheld or breached.

Trustworthiness and reflexivity

Quality was supported through Lincoln and Guba's (1985) criteria. *Credibility* was addressed by triangulating several document types per case; *transferability* by thick description of each case's context; *dependability* by keeping an audit trail of coding decisions; and *confirmability* through reflexive notes. The researcher acknowledges a pro-integrity stance and an interest in AI-supported learning, and sought disconfirming evidence by also analysing a governance case in which AI use is permitted under conditions.

Limitations of the design

Document-based cases are biased toward *detected and publicised* failures; many ethical uses and many undetected failures are invisible. The cases come from a fast-moving field, and policies may have changed since the sources were published. Findings are interpretive and analytically, not statistically, generalisable.

Case Descriptions

Case A: An AI-generated figure in a peer-reviewed paper

In early 2024 a paper on spermatogonial stem cell signalling in a biomedical journal drew public attention for including an AI-generated illustration of a rat with anatomically impossible features and nonsense labels. The paper had passed peer review and editorial processing before online criticism prompted a retraction, and the publisher stated that the figure did not meet its standards of accuracy. The case shows how visual output, which readers tend to trust as data-like evidence, can enter the scholarly record when reviewers and editors do not scrutinise it. It also raises disclosure issues, as the use of generative image tools was visible only in the figure's credit.

Case B: Chatbot artefacts left in a published paper

In 2023 an article in a physics journal was flagged by research-integrity sleuths because it contained phrases characteristic of a chatbot interface, such as a button label for regenerating a response, left in the text. The paper was subsequently retracted because the authors had not declared the use of AI text generation and the manuscript had not been adequately checked. The case illustrates that undisclosed AI use is often detected not through sophisticated detection software but through simple carelessness, and that peer review and editorial screening failed twice (Hossen et al., 2026).

Case C: Fabricated authorities in a formal legal investigation

In *Mata v. Avianca* (S.D.N.Y., 2023), lawyers submitted a brief citing court decisions that a chatbot had invented, including plausible case names, quotations and citations. When the court and opposing counsel could not locate the cases, the lawyers asked the chatbot itself to confirm they were real, and it did. The court sanctioned the lawyers and emphasised that technology does not displace the duty to verify. Although set in law, the case is a clear analogue for research: a literature review or reference list generated by an AI tool and not checked against actual sources has the same structure of failure.

Case D: Governance responses on authorship and disclosure

In 2023 the Committee on Publication Ethics (COPE) stated that AI tools cannot meet authorship requirements because they cannot take responsibility or manage conflicts of interest, and that authors must be transparent about AI use. The International Committee of Medical Journal Editors (ICMJE) issued consistent recommendations, and leading publishers such as Springer Nature and Science journals published policies that disallow AI authorship and expect disclosure of AI assistance beyond basic copy-editing (Thorp, 2023; Stokel-Walker, 2023). These responses show an emerging consensus but also leave grey areas, for example what counts as "basic" assistance, how detailed disclosure should be, and how rules apply to AI used in data analysis rather than writing.

Findings

Five themes were developed through cross-case analysis.

Theme 1: Verification is the central ethical duty

Cases A, B and C all share a failure of verification. In Case C, the lawyers treated the chatbot as a trustworthy source and even as a verifier of itself. In Case A, a nonsensical image survived review. In Case B, obvious artefacts were not noticed. The common pattern is *automation complacency*: fluent, confident output lowers the vigilance of human checkers. Hallucination is a known property of language models, which generate statistically plausible text and not verified facts (Bender et al., 2021; Ji et al., 2023). Ethically, this moves the burden to the user. Using AI in research implies a duty to confirm every citation, figure, quotation and statistic against primary sources before it informs a knowledge claim.

> *Interpretive memo: Verification is not a final proofreading step. It is a design requirement of any AI-assisted workflow, built in at the planning stage.*

Theme 2: Transparency and disclosure gaps

Case B's retraction rested on non-disclosure, and Case A turned on how the image was produced and checked. Case D shows institutions converging on disclosure, but also variation in thresholds. Researchers face a *disclosure dilemma*: some fear that declaring AI use will invite stigma or rejection, creating an incentive to conceal. Where policies are vague, concealment can be rationalised as "just editing." Transparency is ethically important because readers and reviewers need to judge the reliability of methods. A methods section that omits an AI step that shaped coding, translation or analysis is materially incomplete.

Theme 3: Accountability cannot be delegated

Across cases, institutions placed responsibility on identifiable humans: the court sanctioned the lawyers, not the software; the journals retracted the papers; COPE and ICMJE reject AI authorship specifically because accountability cannot be assigned to a tool. In practice, however, responsibility becomes diffuse when many actors are involved (authors, reviewers, editors, publishers, tool vendors). The cases show a *responsibility gap* in which each actor assumed another was checking. Ethical practice requires explicitly assigning who verifies what.

Theme 4: Confidentiality and data protection

This theme is less visible in the retraction cases but is central to the governance documents and to research ethics generally. Entering interview transcripts, participant identifiers, unpublished manuscripts or peer-review material into public AI services may breach consent, confidentiality and data-protection obligations, because inputs can be stored, processed in other jurisdictions or used for model improvement depending on the service terms. Research integrity codes and AI governance frameworks (ALLEA, 2023; UNESCO, 2021) stress privacy and data

governance. For qualitative researchers in particular, whose data are rich and potentially identifying, this is a first-order concern. Participants' consent forms rarely anticipate AI processing, which means consent may need to be revisited.

Theme 5: Bias, equity and the erosion of researcher judgement

Language models reflect patterns in their training data and may under-represent or stereotype groups, languages and contexts, notably from the Global South (Bender et al., 2021). In qualitative work, delegating interpretation to AI risks flattening meaning, missing context and replacing reflexive engagement with data by pattern summaries. There is also an equity dimension: AI can help non-native English-speaking researchers write more clearly, which is a justice-enhancing use, while uneven access to paid tools and detection-related suspicion can disadvantage the same researchers. Ethical use therefore requires weighing benefits and harms across groups, and retaining the researcher's interpretive authority.

Table 2: Cross-case summary

Principle	Case A	Case B	Case C	Case D
Verification	Breached	Breached	Breached	Required by policy
Transparency	Weak	Breached	Weak	Required by policy
Accountability	Diffuse	Diffuse	Enforced by sanction	Human-only authorship
Confidentiality	Not salient	Not salient	Not salient	Addressed by wider frameworks
Justice and judgement	Not salient	Not salient	Not salient	Partly addressed

Illustrative Composite Case (Case E)

This case is constructed for illustration and is not empirical data.

A four-person team at a university is conducting a qualitative study of social support among older adults. The team has 30 interview transcripts. A junior researcher, under deadline pressure, uploads the transcripts to a free public chatbot to "suggest themes," asks it to draft the literature review with references, and uses it to translate quotations from a local language into English.

Applying the five themes to this scenario shows where ethical risks arise:

- **Verification:** The chatbot-produced literature review contains two references that do not exist and one misattributed finding. Without a verification step, these would enter the manuscript.
- **Transparency:** The methods section says only that "thematic analysis was conducted." It does not state that an AI tool generated candidate themes or helped with translation.
- **Accountability:** The supervisor assumes the junior researcher checked everything; the junior researcher assumes the supervisor will review it. No one owns verification.
- **Confidentiality:** Participants consented to data being used by the research team only. Uploading transcripts to a public service may breach that consent and data-protection rules, even if names were removed, because stories and context can identify older adults in small communities.
- **Justice and judgement:** The AI-suggested themes reflect common Western ageing narratives and miss locally specific meanings of family obligation. Translation flattens culturally loaded terms.

How the same project could be conducted ethically

The team could (a) use an institution-approved or locally hosted tool, or avoid AI for raw transcripts altogether; (b) obtain explicit consent for any AI processing; (c) treat AI-suggested themes as one input and carry out independent human coding with a documented comparison; (d) verify every reference against databases such as Scopus or Web of Science; (e) have a bilingual team member validate translations; (f) record prompts, tool versions and decisions in an audit log; and (g) disclose all AI use in the methods and acknowledgements.

Discussion

What the cases reveal

The documented failures are rarely about AI doing something exotic. They are about ordinary lapses of professional care that AI makes easier and more consequential. Fluency creates false confidence; speed encourages skipping checks; and diffuse workflows hide responsibility. This supports Resnik and Hosseini's (2024) argument that guidance must specify concrete practices and not merely restate general integrity values.

Why disclosure norms are not enough

Case D shows progress in policy, but disclosure alone would not have prevented Cases A to C. Disclosure informs readers; it does not guarantee accuracy. Ethical use needs a combination of disclosure, verification and accountability structures, plus attention to data governance that disclosure policies often overlook.

The V-T-A-C-J framework

Synthesising the themes, this article proposes a practical framework for researchers, supervisors and ethics committees:

1. **V: Verify.** Check every AI-generated fact, citation, figure and translation against primary sources. Never use AI to verify AI.
2. **T: Transparently disclose.** Report the tool, version, purpose, date of use and extent of human revision in the methods or acknowledgements.
3. **A: Assign accountability.** Name the person responsible for each AI-assisted task. AI is never an author.
4. **C: Protect confidentiality.** Do not enter identifiable, sensitive or unpublished material into tools that lack appropriate data-protection terms; update consent and ethics applications where AI processing is planned.
5. **J: Keep human judgement and justice.** Reserve interpretation and key analytic decisions for researchers, and assess bias and equity impacts, including for under-represented languages and communities.

Implications

For researchers: Treat AI like a very fast but unreliable research assistant who must be supervised. Keep an AI-use log.

For supervisors and institutions: Provide clear, enforceable policies that distinguish low-risk uses (for example, grammar checking) from high-risk ones (for example, data analysis or generating citations); provide approved tools; train students and staff; and include AI in ethics-application forms.

For journals and reviewers: Standardise disclosure statements, check references and images, and clarify who is responsible for screening.

For policymakers: Align national guidance, such as Malaysia's AI governance principles, with research-integrity codes, and support equitable access to tools and training.

Limitations and future research

The study relies on publicly documented cases, which favour detected failures. Future research should gather primary qualitative data through interviews and focus groups with researchers, supervisors, editors and ethics committee members; examine discipline-specific practices (for example, in clinical, social science and engineering research); and evaluate whether training and checklists based on the V-T-A-C-J framework actually improve practice.

Conclusion

AI can support formal research investigation, but its ethical use depends on people, not on the tools. The cases analysed show that the main risks are unverified outputs, undisclosed use, diffuse responsibility, breaches of confidentiality and loss of interpretive judgement. Researchers and institutions can address these through a human-in-the-loop approach that combines verification, transparency, accountability, data protection and critical judgement. Because AI capabilities and rules change quickly, ethical practice should be treated as an ongoing process of reflection and updating, not as a one-time compliance exercise.

Funding: The research did not receive financial assistance from any funding entity.

Conflicts of Interest: The author has no conflicts of interest to disclose concerning this study.

Declarations: This manuscript has not been published to any other journal or online sources.

Data Availability: The author has all the data employed in this research and is open to sharing it upon reasonable request.

REFERENCES

Please verify all details (volume, pages, DOIs) against the original sources before submission.

ALLEA (2023). *The European Code of Conduct for Research Integrity* (Revised edition). ALLEA – All European Academies.

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 610–623.

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

Committee on Publication Ethics (COPE) (2023). *Authorship and AI tools: COPE position statement*.

Floridi, L., & Cowls, J. (2019). A unified framework of five principles for AI in society. *Harvard Data Science Review, 1*(1).

International Committee of Medical Journal Editors (ICMJE) (2023). *Recommendations for the Conduct, Reporting, Editing, and Publication of Scholarly Work in Medical Journals* (updated guidance on AI-assisted technology).

Ji, Z., Lee, N., Frieske, R., et al. (2023). Survey of hallucination in natural language generation. *ACM Computing Surveys, 55*(12), 1–38.

Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine*

Intelligence, 1, 389–399.
Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic Inquiry*. Sage.
Hossen, M. S., Pauzi, H. M., Islam, M. S., & Salleh, S. F. (2026). ELDERLY LIFE SATISFACTION THROUGH SOCIAL INTERACTION AND FORMAL CARE CENTER MANAGEMENT. *Asian People Journal (APJ)*, *9*(1), 1–15.
Hossen, M. S., Hoque, M. N., Riyadh, M., & Ullah, N. (2026). Determinants of Wellbeing in Later Life: The Mediating Role of Social Support and Community Participation Among Seniors. Journal of Content Validation, 2(2), 166-178.
Ministry of Science, Technology and Innovation, Malaysia (2024). *National Guidelines on AI Governance and Ethics*.
Mata v. Avianca, Inc., No. 22-cv-1461 (S.D.N.Y. 2023).
Resnik, D. B., & Hosseini, M. (2024). The ethics of using artificial intelligence in scientific research: New guidance needed for a new tool. *AI and Ethics*.
Stake, R. E. (1995). *The Art of Case Study Research*. Sage.
Stokel-Walker, C. (2023). ChatGPT listed as author on research papers: Many scientists disapprove. *Nature, 613*, 620–621.
Thorp, H. H. (2023). ChatGPT is fun, but not an author. *Science, 379*(6630), 313.
UNESCO (2021). *Recommendation on the Ethics of Artificial Intelligence*. UNESCO.
Yin, R. K. (2018). *Case Study Research and Applications: Design and Methods* (6th ed.). Sage.

This is an Open Access article distributed under the terms of the Creative Commons Attribution 4.0 International License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium upon the work for non-commercial, provided the original work is properly cited.