

ETHICAL VULNERABILITIES IN AI-GENERATED RESEARCH: A STUDY ON LLM-BASED TOOLS

Ankit Agarwal

Department of Computer Science, NCERT, New Delhi
Research Scholar, Malwanchal University, Indore

Dr. Rajesh D.

Associate Professor, DICT, CIET, NCERT, New Delhi

Abstract

ChatGPT, GPT-4, Claude, and LLaMA are examples of Large Language Models (LLM) and have changed the landscape of academia and research in a very short time. Such models help in the drafting, summarisation, translation and analysis of the material, but also raise serious ethical issues regarding the authorship, plagiarism, hallucination, misinformation, data privacy, bias and academic integrity. Using the real-life literature of scholars and the most recent guidelines of major publishing houses and ethical organizations, this paper analyzes the ethical considerations that can be considered as the use of LLMs in the research. The interpretation of the main themes formed in the narrative literature review and thematic analysis was used as the qualitative method applied to explain the results of the research conducted by peer-reviewed sources and institutional recommendations. The results indicate that, despite the fact that LLMs promote the productivity of research, they increase the ethical issues of the past and introduce new ethical dilemmas related to accountability, transparency, and justice. The research paper suggests that scholars should use LLMs as an aid, publicly share their use, thoroughly authenticate the generated information, and protect sensitive data. The paper concludes that to achieve responsible integration of LLMs, institutional policies should be developed, users should practice responsibly and good human control should be exercised to ensure that the integrity of scholarly work is preserved.

Keywords: Artificial Intelligence; Large Language Models; LLMs; Academic Ethics; Academic Integrity.

1. INTRODUCTION

The past few years have seen the rapid development of Generative Artificial Intelligence, and large language models are now ubiquitous in academic writing, research processes, and tertiary education. ChatGPT and GPT-4 are LLM that have astounding skills at creating structured text and summarising scholarly work, refining language and even suggesting an argument. As Khalifa (2024) found, AI tools are increasingly used by researchers to conduct a literature review, develop and edit a paper, as it is more efficient and is more language-sophisticated. On the same note, Ocampo (2023) observed that, in reference to scientific writing, LLMs lead to better elucidation and organization of the text, yet at the same time, it prompts the issue of hallucination, misinterpretation and over-reliance.

Within the ethical framework of research, the inquiries about the authorship, originality, accountability, and reliability of the AI-generated content become difficult to resolve through the use of LLCM. As Weidinger et al. (2021) emphasize, misinformation, amplifying and eroding epistemic trust, as well as other social risks of LLM have direct implications on scholarly communication. COPE (2023) additionally noted that AI tools were unable to comply with the authorship standards, which led to new dilemmas in researchers.

This research seeks to discuss these ethical problems in a traditional scholarly manner. The paper examines the way the LLMs are changing the standard research practices, the ethical risks they cause and how these are being addressed by the academic community by examining real scholarly materials and developing institutional guidelines.

2. OBJECTIVES OF THE STUDY

The research aims at achieving the following goals:

- i. To examine the roles that Large Language Models play at different phases of the research lifecycle.
- ii. To determine significant ethical issues of using LLM in research and academic writing.
- iii. To create emergent recommendations in ethical organizations, publishers, and universities on the use of AI.
- iv. To make ethical issues meaningful based on the academic integrity and responsible research practices framework.
- v. To provide viable suggestions to researchers, educators, institutions and policy-makers.

3. REVIEW OF LITERATURE

3.1 LLMs and Scientific Writing

The rising power of LLMs in academic writing is a subject that has attracted the attention of scholars. According to Ocampo (2023), ChatGPT is more effective at enhancing coherence, grammar and structure, but it tends to produce false or fabricated information that has to be checked manually. Almarie (2023) also emphasized the fact that ChatGPT is often used to come with references that are hallucinations in the form of citations that seem real but are not, which can be dangerous to the reliability of science.

The research on AI-assisted scientific writing indicates that scholars rely on LLMs to summarize the literature and extract their main ideas and refine their expressions; however, they should be alert to the fact that subtle meaning distortions can occur. Khalifa (2024) states that AI-based technologies facilitate academic writing and at the same time offload the responsibility of verifying to human writers who must make sure that they do not include any fabricated information.

3.2 LLM Ethical and Social risk

Fundamental taxonomy of ethical risks related to LLMs has been presented by Weidinger et al. (2021), who reported that such issues as misinformation, discrimination, malicious use, and erosion of trust were the most common. Stahl (2024) also brought the discussion further by using the already developed ethical lenses, autonomy, justice, responsibility and beneficence, to examine the predicament of ChatGPT in moral reasoning.

These concerns have since been extended to research contexts in domains by subsequent studies. Zhui et al. (2024) demonstrated that the content produced by LLM in the medical research may present misleading statements that could have detrimental effects. Equally, Almufarreh (2025) reported that discriminative results with the help of LLMs encourage the exclusion of minority voices and support the current inequities in the field of educational research.

3.3 Academic Dishonesty, Ghostwriting and Plagiarism

A major literature is in favor of the fact that LLMs threaten the norms of academic integrity. As pointed out by Guleria et al. (2023), AI tools are often used by students to create essays, paraphrases and make complex arguments, which can be considered ghostwriting. Kendall (2024) cautioned that even academic publishing and peer review can be used to mass-produce fraudulent research articles using LLMs.

Pudasaini (2024) showed that current plagiarism detection systems have a hard time distinguishing between AI-generated and human-generated text, so the definitions of plagiarism should be expanded to incorporate the notions of ideas-level and AI-assisted generation.

3.4 Ethics on authorship, responsibility and publication

COPE (2023) made it clear that AI tools should not appear among the authors, because this title presupposes responsibility, accountability, and decision-making, which are all the qualities that AI systems lack. According to a thematic analysis conducted by Ethical Guidelines for Generative AI (2025), publishers are settling on three norms, namely, AI cannot be listed as an author, the usage of it should not be hidden, but rather showcased in a transparent manner, and all the content must be checked by human authors.

IEEE (2023) and AIMS Press (2024) guidelines do not allow the application of AI tools to falsify data or manipulate images and remind researchers that they should be in charge of all the features of the scholarly record.

3.5 Higher Education Institutional Responses

Academic institutions across the world have started reforming their academic integrity models. The School of International Studies of JNU (2025) made all theses subject to

plagiarism checks and controlled AI assisted writing emphasizing open disclosure. The AI committee (2025) at IIT Delhi suggested that students and faculty be AI-verified when they generate any content and that they should be critically monitored by humans.

All these aspects suggest that research ethics is changing to reflect the changing nature of academic practice, which is driven by LLM.

4. RESEARCH METHODOLOGY

The research design used in this study is a qualitative, exploratory research design that is based on the use of a literature review based on a narrative analysis and thematic interpretation.

4.1 Data Sources

Data were collected from:

- Peer-reviewed journal articles.
- Research editorials
- Publisher guidelines
- COPE, IEEE and educational institution policy documents.

Systematic reviews and analyses of ethics.

4.2 Inclusion Criteria

The sources were incorporated as they:

- Large Language Models or generative AI;
- Ethical, methodological or academic integrity concerns have been analysed;
- Were published between 2021 and 2025;
- Provided the conceptual or empirical knowledge of the research ethics.

4.3 Analytical Procedure

This literature was coded based on themes such as authorship, plagiarism, hallucination, bias, privacy, disclosure, and accountability. The codes were coded up into higher theme groupings as they appear under the Data Interpretation section.

The appropriateness of this methodological approach can be explained by the fact that the given sphere is developing at a rather rapid pace, and qualitative synthesis will provide the opportunity to deepen knowledge of new ethical standards.

5. THEMATIC ANALYSIS (THEMATIC INPLGation of Data)

5.1 Authorship and Accountability

The second item is authorship and accountability, which is observed promptly once the research topic is recognized. The general opinion among the sources on whether AI can meet the requirements of authorship is clearly in the negative. COPE (2023) claimed that authorship involves responsibility and the capability to reply to peer review which AI systems simply do not have. This brings about ethical uncertainties as LLMs come up with large sections of manuscripts. According to Ocampo (2023), AI can help organize the arguments but the intellectual input should not be humanized.

5.2 Ghost-writing, Integrity Problems and Plagiarism

The LLMs generate novel forms of plagiarism by producing novel-seeming text based on phenomena in their training data. Guleria et al. (2023) noted that AI tools were widely misused by students who wanted to write complete assignments. Pudasaini (2024) also demonstrated that there is a possibility of AI-generated text overlooking plagiarism-detection programs, which puts pressure on institutions to revise their academic integrity models.

Kendall (2024) cautioned about AI-driven paper mills, which partially automate the process of producing fraudulent pieces of research and puts the reputation of academic publishing at risk.

Misinformation and hallucination are considered to be typical signs of the disease.

5.3 Hallucination and Misinformation

Various authors underline the threat of illusory sources or fake statements. According to Almarie (2023), there were instances when ChatGPT generated credible but fabricated references. Zhui et al. (2024) demonstrated that clinical research can be dangerous to have its harmful misinterpretation because hallucinating medical content.

These errors are significant dangers to evidence-based scholarship, which destroys the reliability and replicability of scientific work.

5.4 Bias and Epistemic Injustice

Weidinger et al. (2021) and Almufarreh (2025) concluded that LLMs enhance biases inherent in their training data, making it difficult to produce knowledge in an unbiased manner. Underrepresented groups can be disproportionately harmed by biased outputs, literature reviews can be biased, and research agendas can be distorted.

5.5 Privacy and Confidentiality

Publisher policies prohibit posting confidential or property information into third party LLM interfaces. According to IEEE (2023), researchers can break the confidentiality with trainees inadvertently when they provide datasets to AI platforms that archive model training prompts.

5.6 Responsible Disclosure and Governance

Thematic analysis shows that there is a high level of consensus that users need:

- Disclose AI usage;
- Check every AI generated content;
- Do not manipulate AI-generated data;
- Maintain complete accountability on research integrity.

Responsible governance implies coordination of researchers, institutions and publishers, as Ethical Guidelines for Generative AI (2025) recommends.

6. FINDINGS

The main findings are:

- 6.1 LLM is actively being incorporated in research writing, but their application continues to be inconsistently revealed (Khalifa, 2024).
- 6.2 The practice of authorship has never been put under such pressure, and new norms and guidelines have to be offered (COPE, 2023).
- 6.3 With current systems, AI-generated plagiarism is impossible to detect (Pudasaini, 2024).
- 6.4 One of the significant risks to academic reliability is hallucination of facts, references and summaries (Ocampo, 2023).
- 6.5 LLMs replicate systemic biases, which marginalise some systems of knowledge (Weidinger et al., 2021).
- 6.6 The risk of data privacy is still little known, particularly to students and young scholars (IEEE, 2023).
- 6.7 Policies of AI are slowly being implemented in institutions and practices are broad in nature (JNU, 2025; IIT Delhi, 2025).

7. SUGGESTIONS

7.1 For Researchers

- Use LLMs as support and not as a substitute; intellectual and ethical accountability.
- Check all the facts, references and interpretations produced by AI.

- Do not input personal information in open AI platforms.
- Report AI usage in recognition or methodology.
- Be trained in AI literacy and responsible use.

7.2 of Universities and Institutions.

- Revise the policies governing academic integrity to accommodate AI-generated work.
- Conduct AI ethics and digital literacy workshops.
- Secure data privacy with institution-approved AI tools.
- Insert AI-use questions on ethics clearance forms.
- Fund the study of AI effects on academic integrity.

7.3 For Publishers and Editors

- Make authors state about any AI use at the time of manuscript submission.
- Keep banning AI as an author.
- Engage superior vetting procedures of references and information integrity.
- Formulate discipline-related guidelines to use AI.
- Promote AI ethics and governance.

8. CONCLUSION

LLM is a paradigm shift in scholarly writing and research. Their capability to produce fluent, structured text has an evident advantage in terms of productivity and accessibility. Nevertheless, as this paper has demonstrated, they also pose serious ethical dilemmas in terms of authorship, plagiarism, hallucination, fairness, privacy and accountability. The literature shows that the responsible integration of LLMs can only be achieved by integrating technology knowledge with good moral conducts.

Academic community starts to react by providing revised guidelines, yet constant revising and adjusting will be required due to the development of AI technologies. Finally, human intellectual, critical thinking and scholarly integrity should be improved by LLMs rather than substituted. Ethical standards will ensure that research can be trusted only when human researchers are in full control of the originality of their work, the accuracy and the ethical standards in all of their aspects.

References:

1. Almarie, B. (2023). *The use of large language models in science*. Saudi Journal of Anaesthesia.
2. Almufarreh, A. (2025). *Ethical implications of ChatGPT and other large language models in education*. *Frontiers in Artificial Intelligence*, 5, 1615761.
3. AIMS Press. (2024). *Guidelines for the use of AI tools in writing and research*.
4. Committee on Publication Ethics (COPE). (2023). *Authorship and AI tools*.

5. Ethical guidelines for the use of generative artificial intelligence in scholarly publishing. (2025). *[Journal article]*.
6. Guleria, A., et al. (2023). *ChatGPT: Ethical concerns and challenges in academics and research*. Journal of Infection in Developing Countries.
7. IEEE Robotics and Automation Society. (2023). *Guidelines for generative AI usage in publications*.
8. IIT Delhi AI Committee. (2025). *AI but verify: Navigating the future of learning*.
9. Jawaharlal Nehru University. (2025). *Revised research rulebook on AI use and plagiarism checks*.
10. Kendall, G. (2024). *Risks of abuse of large language models in scholarly publishing*. Learned Publishing.
11. Khalifa, M. (2024). *Using artificial intelligence in academic writing and research: A systematic review*. Discover Artificial Intelligence.
12. Nie, B., et al. (2023). *ChatGPT and AI-written research papers: Ethical considerations for scholarly publishing*. Information Matters, 3(3).
13. Ocampo, T. S. C. (2023). *ChatGPT and scientific writing: Ethical boundaries*. International Journal of Surgery Open.
14. Pudasaini, S. (2024). *A survey on plagiarism detection in large language models*. arXiv:2407.13105.
15. Stahl, B. C. (2024). *The ethics of ChatGPT: Exploring the ethical issues*. Information & Management.
16. Weidinger, L., et al. (2021). *Ethical and social risks of harm from language models*. arXiv:2112.04359.
17. Zhui, L., et al. (2024). *Ethical considerations of LLMs in medicine*. Journal of Medical Internet R