

ARTIFICIAL INTELLIGENCE

Academic Hallucinations in the Age of Artificial Intelligence: Causes, Consequences, and Mitigation Strategies

Ofor Chibe

Ave Maria University, Piyanko, Nasarawa State, Nigeria

Published October 2026

Abstract

This study examines the causes, consequences, and mitigation strategies of academic hallucinations in the age of artificial intelligence. A mixed-methods research design is used, combining a survey of university lecturers, postgraduate students, and researchers with in-depth interviews involving journal editors, librarians, and AI experts. Quantitative data are analyzed using descriptive statistics, while qualitative data are subjected to thematic analysis. The findings indicate that inadequate AI literacy, poor prompt design, overreliance on AI-generated content, and the absence of institutional guidelines are major causes of academic hallucinations. The study further reveals that hallucinations contribute to fabricated citations, misinformation, reduced research credibility, ethical concerns, and declining public trust in scholarly communication. The study identifies effective mitigation strategies, including AI literacy training, prompt engineering, mandatory human verification, institutional AI policies, and the integration of AI detection and reference verification tools into academic workflows. The study contributes to the growing literature on responsible AI by proposing an integrated framework for minimizing academic hallucinations and strengthening research integrity. The findings provide practical recommendations for universities, researchers, publishers, and policymakers seeking to promote trustworthy and ethical AI-assisted academic research.

Keywords: artificial intelligence; academic hallucinations; research integrity; generative AI; higher education; prompt engineering; ai literacy

1 Introduction

Artificial intelligence (AI) has transformed academic research and scholarly writing. This transformation was done through the provision of tools that support literature searches, data analysis, language editing, and content generation. Researchers' productivity and accessibility to knowledge has been improved with the emergence of generative AI systems. Universities, research institutions, and scholarly publishers recognizes AI as a valuable tool for enhancing academic efficiency and innovation. However, alongside these benefits has emerged a major challenge known as academic hallucination, whereby AI systems generate false, fabricated, or misleading information that appears credible but lacks factual accuracy (Dwivedi et al., 2023). Whenever AI-generated outputs include fictitious references, inaccurate citations, fabricated empirical findings, or incorrect interpretations of scientific evidence, it is referred as hallucination. Such errors can undermine research quality, academic integrity, and public trust in scientific communication. Because generative AI models predict text based on statistical patterns rather than genuine understanding, they may produce convincing but inaccurate content, especially when responding to complex or highly specialized academic questions (OpenAI, 2025).

The consequences of academic hallucinations are numerous. Beyond individual researchers, this consequence may contribute to the spread of misinformation within scholarly literature, weaken evidence-based decision-making, increase the likelihood of plagiarism or citation fraud, and reduce confidence in AI-supported research practices (UNESCO, 2023). Academic publishers have therefore introduced new policies requiring transparency in AI usage and emphasizing the importance of human verification throughout the research process (Committee on Publication Ethics [COPE], 2023). Other strategies have been proposed to reduce AI hallucinations in academic writing. These include effective prompt engineering, critical evaluation of AI-generated content, verification of references using trusted academic databases, improved AI literacy among researchers, and stronger institutional guidelines governing responsible AI use (Kasneci et al., 2023). Human oversight remains essential because AI should complement rather than replace scholarly judgment. As AI technologies continue to evolve, developing comprehensive mitigation strategies will be critical for preserving research integrity and ensuring trustworthy scholarly communication.

Although recent studies have examined AI adoption in higher education, limited empirical evidence exists regarding the causes, consequences, and effective mitigation of academic hallucinations, particularly in developing countries. Addressing this gap is important for developing policies that encourage responsible AI adoption while safeguarding academic quality. This study therefore investigates the causes, consequences, and mitigation strategies for academic hallucinations in the age of artificial intelligence

1.1 Background of the study

Artificial intelligence (AI) has become one of the most transformative technologies of the twenty-first century, reshaping numerous sectors, including healthcare, finance, manufacturing, education, and scientific research. The rapid development of generative AI, particularly large language models (LLMs) such as ChatGPT, Gemini, Claude, and Microsoft Copilot, has significantly changed the way knowledge is created, accessed, and communicated (Dwivedi et al., 2023). These AI systems can generate human-like text, summarize complex documents, translate languages, write computer code, and assist researchers in literature reviews, data interpretation, and academic writing. Their ability to process vast amounts of information within seconds has made them valuable tools for improving research productivity and enhancing scholarly communication.

Thus, the integration of AI into higher education and academic research has accelerated since the public release of advanced generative AI models. Universities, research institutes, publishers, and professional organizations increasingly recognize AI as a complementary tool capable of improving learning, supporting innovation, and reducing the time required for academic tasks (Kasneci et al., 2023). Researchers now employ AI to generate research ideas, organize literature, improve grammar, summarize articles, and draft manuscripts. Students also utilize AI to support learning, prepare assignments, and enhance academic writing. These developments have contributed to a growing digital transformation within higher education, making AI an indispensable component of modern scholarship.

However, despite these remarkable benefits, the increasing reliance on generative AI has introduced new ethical and methodological challenges. One of the most significant concerns is the phenomenon known as AI hallucination. AI hallucination refers to situations in which an AI system generates information that appears plausible and convincing but is factually incorrect, fabricated, misleading, or unsupported by reliable evidence (OpenAI, 2025). Unlike traditional search engines that retrieve existing information, generative AI predicts the most probable sequence of words based on patterns

learned during training. Consequently, the system may confidently produce nonexistent references, inaccurate statistics, fabricated quotations, incorrect legal cases, or false interpretations of scientific findings. When such outputs are incorporated into scholarly work without proper verification, they become academic hallucinations that threaten research quality and credibility.

Hence, academic hallucinations have emerged as a growing concern because they undermine the fundamental principles of scientific inquiry, including accuracy, transparency, objectivity, and evidence-based reasoning. Scientific research depends on verifiable evidence and authentic sources. However, studies have shown that AI-generated content may contain fabricated citations, nonexistent journal articles, inaccurate Digital Object Identifiers (DOIs), imaginary authors, and misleading summaries of published research (Nature, 2023). Because these outputs often appear professionally written and highly convincing, inexperienced researchers may unknowingly include them in dissertations, journal articles, conference papers, and policy reports. Such practices compromise academic integrity and may contribute to the spread of misinformation within scholarly communities.

Furthermore, the increasing availability of AI writing tools has also raised important concerns regarding research ethics. Academic institutions traditionally emphasize originality, critical thinking, proper citation practices, and responsible scholarship. Excessive dependence on AI-generated content may weaken these essential academic values if researchers fail to critically evaluate AI outputs before publication. UNESCO (2023) emphasizes that AI should support rather than replace human intellectual responsibility. Similarly, the Committee on Publication Ethics (COPE, 2023) advises authors to disclose the use of AI where appropriate and stresses that authors remain fully accountable for the accuracy and integrity of their published work regardless of whether AI tools were used during manuscript preparation.

Thus, many factors can contribute to the occurrence of academic hallucinations. Firstly, generative AI systems lack genuine understanding of the information they produce. Rather than reasoning like human experts, they generate responses by predicting linguistic patterns from training data (OpenAI, 2025). Consequently, when reliable information is unavailable or ambiguous, AI systems may generate plausible but inaccurate responses instead of acknowledging uncertainty. Secondly, poor prompt engineering often contributes to hallucinations. Broad, vague, or poorly structured prompts may encourage AI systems to produce speculative or fabricated information. Thirdly, limited AI literacy among students and researchers reduces their ability to distinguish accurate outputs from fabricated content. Many users mistakenly assume that AI-generated responses are always correct because they are presented confidently and fluently (Kasneci et al., 2023).

Moreover, institutional factors also influence the prevalence of academic hallucinations. Many universities have yet to develop comprehensive policies governing responsible AI use in research and teaching. While some institutions have introduced guidelines on AI-assisted writing, others continue to rely on traditional academic integrity policies that do not adequately address the unique challenges posed by generative AI (UNESCO, 2023). In many developing countries, inadequate digital literacy, limited access to AI training, and insufficient awareness of responsible AI practices further increase the likelihood of inappropriate AI use in academic settings.

Besides, the consequences of academic hallucinations are numerous. Beyond individual researchers, it extends to the institutional level. Fabricated information may damage the credibility of universities, research institutes, and scholarly journals. Publishers increasingly report submissions containing false references, inaccurate citations, and AI-generated content lacking adequate verification (Nature, 2023). Such practices increase editorial workload, complicate the peer-review process, and may

result in article retractions. At the societal level, inaccurate academic publications may misinform policymakers, healthcare professionals, educators, businesses, and the general public who depend on scientific evidence for decision-making. In this regard, academic hallucinations represent not only an educational challenge but also a threat to evidence-based policy and sustainable development.

Recognizing these risks, scholars and international organizations have proposed various strategies to mitigate academic hallucinations. One important approach involves improving AI literacy among researchers, educators, and students through structured training programmes that emphasize responsible AI use, source verification, prompt engineering, and critical evaluation of AI-generated content (UNESCO, 2023). Human oversight remains indispensable because AI outputs should always be verified using authoritative databases such as Scopus, Web of Science, PubMed, Crossref, and institutional repositories before being incorporated into scholarly work.

Moreover, technological interventions also offer promising solutions. AI developers continue to improve retrieval-augmented generation (RAG), fact-checking systems, citation verification tools, and hallucination detection algorithms that enhance the reliability of AI-generated information (OpenAI, 2025). Academic publishers are increasingly adopting AI disclosure policies, plagiarism detection systems, and editorial guidelines that encourage transparency in AI-assisted research. These initiatives aim to preserve academic integrity while enabling researchers to benefit responsibly from AI technologies.

Although recent literature has extensively discussed the opportunities and challenges of generative AI in education and research, empirical evidence concerning the causes, consequences, and effective mitigation of academic hallucinations remains limited, particularly in developing countries where AI governance frameworks are still emerging. Much of the existing research focuses on AI adoption rather than on understanding how hallucinations affect research quality, publication ethics, and institutional credibility. There is therefore a need for empirical studies that investigate the factors responsible for academic hallucinations and identify practical strategies for minimizing their occurrence within higher education institutions. Such evidence will contribute to developing institutional policies, improving AI literacy, strengthening research integrity, and promoting responsible AI adoption in scholarly communication.

1.2 Statement of the problem

The rapid adoption of generative artificial intelligence (AI) tools such as ChatGPT, Gemini, Claude, and Microsoft Copilot has transformed academic research and scholarly communication by enhancing literature searches, data analysis, language editing, and manuscript preparation. While these technologies have improved research productivity and accessibility, they have also introduced the growing challenge of academic hallucinations—the generation of false, fabricated, or misleading information that appears authentic and credible (Dwivedi et al., 2023). Such hallucinations include fictitious references, fabricated citations, nonexistent authors, inaccurate research findings and misleading interpretations of scholarly evidence, all of which threaten the credibility of academic research.

But recent studies have shown that AI-generated content may contain convincing but inaccurate information because large language models generate responses based on statistical language patterns rather than genuine understanding or verification of facts (Kasneci et al., 2023). Consequently, students, researchers, and even experienced academics may unknowingly incorporate fabricated information into theses, dissertations, journal articles, grant proposals, and policy

documents. This practice undermines research integrity, weakens evidence-based decision-making, increases editorial challenges for publishers, and erodes public trust in scientific knowledge (UNESCO, 2023).

Although universities and scholarly publishers have begun issuing guidelines on the responsible use of AI in academic writing, many higher education institutions, particularly in developing countries, still lack comprehensive policies, AI literacy programmes, and institutional mechanisms for detecting and preventing academic hallucinations (Committee on Publication Ethics [COPE], 2023). Furthermore, empirical studies examining the causes, consequences, and effective mitigation strategies of academic hallucinations remain limited. Existing research has focused largely on the opportunities and ethical implications of generative AI rather than investigating the factors responsible for hallucinations and practical interventions for reducing their occurrence.

This knowledge gap limits the ability of universities, policymakers, journal editors, and researchers to develop evidence-based strategies that promote responsible AI adoption while safeguarding academic integrity. Therefore, there is a need for empirical research that investigates the causes of academic hallucinations, examines their consequences for research quality and scholarly communication, and identifies effective mitigation strategies capable of enhancing the credibility, reliability, and ethical use of artificial intelligence in academic research.

1.3 Objectives of the study

1.3.1 General objective

To examine the causes, consequences, and mitigation strategies of academic hallucinations in the age of artificial intelligence.

1.3.2 Specific objectives

1. To identify the major causes of academic hallucinations in AI-assisted academic writing.
2. To examine the consequences of academic hallucinations on research quality and academic integrity.
3. To assess researchers' awareness of academic hallucinations and responsible AI use.
4. To determine the effectiveness of AI literacy, prompt engineering, and human verification in reducing academic hallucinations.
5. To propose a framework for mitigating academic hallucinations in higher education institutions.

1.3.3 Research questions

1. What are the major causes of academic hallucinations in AI-assisted academic writing?
2. How do academic hallucinations affect research quality, credibility, and academic integrity?
3. What is the level of awareness of academic hallucinations among researchers and university students?
4. To what extent do AI literacy, prompt engineering, and human verification reduce academic hallucinations?
5. What strategies can universities adopt to mitigate academic hallucinations and promote responsible AI use in research?

2 Literature review

The rapid advancement of artificial intelligence (AI) has transformed the landscape of higher education and academic research. This is seen with the provision of innovative tools for information retrieval, language generation, data analysis, and scholarly writing. The emergence of generative AI technologies, particularly large language models (LLMs), has significantly improved research productivity and accessibility to knowledge. Researchers, educators, and students increasingly rely on AI-powered applications to summarize literature, generate research ideas, improve language quality, and support scientific communication (Dwivedi et al., 2023). While these developments offer numerous opportunities for enhancing academic performance, they have also raised important concerns regarding the accuracy, reliability, and ethical use of AI-generated information.

One of the most debated issues associated with generative AI is the occurrence of academic hallucinations. AI hallucinations refer to instances in which AI systems produce fabricated, inaccurate, or misleading information that appears convincing and authoritative but lacking factual support (OpenAI, 2025). In academic contexts, hallucinations may include nonexistent references, fabricated authors, incorrect statistical data, false interpretations of research findings, or inaccurate summaries of scientific literature. Such outputs pose significant risks to research integrity because they may be incorporated into scholarly work without adequate verification. Consequently, understanding the nature of AI, the operation of generative AI, and the phenomenon of academic hallucinations has become increasingly important for researchers, universities, publishers, and policymakers seeking to promote responsible AI use in scholarly communication.

2.1 Concept of artificial intelligence

Artificial intelligence refers to the ability of computer systems to perform tasks that traditionally require human intelligence, including learning, reasoning, problem-solving, decision-making, language understanding, and pattern recognition. AI combines computer science, statistics, mathematics, cognitive science, and data analytics to develop intelligent systems capable of processing large volumes of information and generating meaningful outputs (Russell & Norvig, 2021). Unlike conventional software that follows predefined rules, AI systems improve their performance through machine learning algorithms that identify relationships within data and adapt to new information.

Recent advances in deep learning and neural networks have significantly expanded AI capabilities across diverse sectors such as healthcare, finance, agriculture, education, journalism, and public administration. In higher education, AI supports personalized learning, automated assessment, plagiarism detection, research assistance, and academic administration (UNESCO, 2023). Universities increasingly employ AI-powered platforms to facilitate teaching, improve institutional management, and enhance research productivity.

The growing influence of AI has generated both optimism and concern within academia. On one hand, AI enhances efficiency by reducing the time required for repetitive academic tasks and increasing access to scientific knowledge. On the other hand, scholars caution that excessive dependence on AI may reduce critical thinking, weaken originality, and create ethical challenges related to authorship, transparency, and research integrity (Kasneci et al., 2023). These concerns highlight the importance of understanding AI not merely as a technological innovation but as a tool whose effectiveness depends on responsible human oversight.

2.2 Generative artificial intelligence in academic research

Generative artificial intelligence represents one of the most significant developments in AI research. Unlike traditional AI systems that classify or predict existing information, generative AI creates new content—including text, images, audio, computer code, and videos—based on patterns learned from extensive training datasets (Dwivedi et al., 2023). Examples include ChatGPT, Gemini, Claude, Microsoft Copilot, and other large language models capable of producing human-like responses to user prompts.

Hence, the emergence of generative AI has transformed academic research by automating many aspects of scholarly work. Researchers now use AI to identify research topics, conduct preliminary literature reviews, summarize journal articles, translate scientific texts, generate research questions, draft manuscripts, improve grammar, and organize references. AI-assisted research has become particularly valuable for multilingual scholars and researchers working in resource-constrained environments because it improves access to scientific knowledge and reduces barriers to academic publishing (UNESCO, 2023).

Moreover, generative AI also contributes to interdisciplinary collaboration by facilitating communication across different scientific fields. For example, AI can simplify complex scientific concepts, generate programming scripts for statistical analysis, assist in coding qualitative data, and support data visualization. These capabilities improve research efficiency and enable scholars to devote greater attention to critical thinking and interpretation rather than routine writing tasks.

However, despite these advantages, generative AI has important limitations. Large language models do not possess genuine understanding or consciousness. Instead, they generate responses by predicting the most probable sequence of words based on patterns learned during training (OpenAI, 2025). Consequently, when confronted with ambiguous questions or insufficient information, AI systems may generate inaccurate or fabricated responses while presenting them confidently. This limitation explains why AI-generated academic content requires careful human verification before it can be considered reliable for scholarly use.

2.3 Concept of academic hallucinations

Academic hallucinations refer to AI-generated scholarly information that is false, fabricated, misleading, or unsupported by verifiable evidence but is presented in a manner that appears authentic and academically credible. Hallucinations occur because generative AI predicts likely text sequences rather than retrieving verified facts from authoritative academic databases (OpenAI, 2025). As a result, AI systems may invent journal articles, authors, references, quotations, statistics, legal cases, or scientific findings that do not exist.

Thus, the concept has become increasingly important following the widespread adoption of generative AI in higher education. Nature (2023) observed that although AI-generated content often demonstrates impressive fluency and coherence, factual accuracy cannot be guaranteed. Researchers therefore remain fully responsible for verifying every citation, quotation, and factual claim generated by AI before incorporating it into academic publications. Moreover, academic hallucinations differ from deliberate academic misconduct. While plagiarism and data falsification involve intentional unethical behavior by researchers, hallucinations often arise unintentionally when users assume that AI-generated information is accurate without independent verification. This distinction underscores the need for AI literacy. UNESCO (2023) emphasizes that AI should support human intelligence and should never replace scholarly judgment, critical thinking, or ethical responsibility.

2.4 Types and characteristics of academic hallucinations

Academic hallucinations manifest in several forms depending on the nature of the generated content. One common type is citation hallucination. In this form of hallucination, AI generates nonexistent journal articles, fabricated Digital Object Identifiers (DOIs), incorrect publication years, or imaginary authors. This is among the most frequently reported problems in AI-assisted academic writing and has been identified by journal editors as a growing concern ([Nature, 2023](#)).

Another category is factual hallucination, where AI produces inaccurate scientific facts, historical events, statistical information, or theoretical explanations. Such hallucinations may be difficult to identify because they are often presented confidently using persuasive academic language. Researchers who fail to verify these outputs risk disseminating misinformation within scholarly literature. A third form is interpretive hallucination, whereby AI misrepresents or oversimplifies research findings, theoretical frameworks, or methodological approaches. Although the cited source may exist, the AI-generated interpretation may not accurately reflect the author's original argument. This can lead to incorrect literature reviews and flawed theoretical analyses. Moreover, AI systems may also generate contextual hallucinations, where information is presented outside its appropriate academic or cultural context. Such responses may incorrectly apply theories across disciplines, confuse geographical settings, or ignore contextual differences that influence research findings.

Nonetheless, there are several characteristics that distinguish academic hallucinations from verified scholarly information. Firstly, hallucinated content is usually presented with a high level of linguistic fluency and confidence, making inaccuracies difficult to detect. Secondly hallucinations often appear internally coherent despite lacking factual validity. Thirdly, they frequently contain incomplete, inconsistent, or unverifiable references. Finally, hallucinations become more common when prompts are vague, ambiguous, or require information beyond the AI model's reliable knowledge base ([Kasneci et al., 2023](#)). Recognizing these characteristics is essential for promoting responsible AI use in academia. Researchers must adopt rigorous verification practices by consulting authoritative databases such as Scopus, Web of Science, PubMed, Crossref, and institutional repositories before accepting AI-generated outputs as valid scholarly evidence. Strengthening AI literacy and encouraging critical evaluation will remain fundamental strategies for reducing the prevalence of academic hallucinations and safeguarding research integrity.

2.5 Causes of academic hallucinations

The growing use of generative artificial intelligence (AI) in academic research has brought unprecedented opportunities for knowledge creation and dissemination. However, it has also increased the incidence of academic hallucinations, which arise from a combination of technological, human, and institutional factors. Understanding these causes is essential for developing effective interventions that promote responsible AI use in scholarly communication.

One major cause of academic hallucinations is the probable nature of large language models (LLMs) technology. Unlike search engines that retrieve verified information from databases, generative AI models produce responses by predicting the most likely sequence of words based on patterns learned during training. Consequently, when reliable information is unavailable or ambiguous, AI systems may generate probable but incorrect responses instead of acknowledging uncertainty ([OpenAI, 2025](#)). This predictive mechanism explains why AI sometimes fabricates references, invents quotations, or misrepresents scientific findings.

Another significant cause is human which is seen with poor prompt engineering and limited AI literacy. The quality of AI-generated outputs largely depends on the clarity, specificity, and context of the prompts provided by users. Vague, ambiguous, or incomplete prompts increase the likelihood of hallucinations because AI systems attempt to fill information gaps through statistical prediction rather than factual verification (Kasneci et al., 2023). Researchers who lack prompt engineering skills may unknowingly obtain inaccurate or misleading information. Besides, limited AI literacy among researchers, lecturers, and students also contributes to academic hallucinations. AI literacy refers to the knowledge and competencies required to use AI responsibly, critically evaluate AI-generated outputs, and recognize the limitations of AI technologies. UNESCO (2023) argues that many AI users mistakenly assume that fluent and coherent AI-generated responses are necessarily accurate. Such misconceptions increase dependence on AI without appropriate verification.

The third factor is institutional. Institutional deficiencies further contribute to the problem. Many universities have not yet established comprehensive AI policies governing academic writing, assessment, research ethics, or disclosure of AI-assisted work. The absence of standardized institutional guidelines leaves researchers uncertain about acceptable AI practices and reduces accountability for verifying AI-generated information (Committee on Publication Ethics [COPE], 2023). Finally, the rapid expansion of AI tools has outpaced the development of regulatory frameworks and editorial standards. Academic publishers continue to encounter manuscripts containing fabricated references and AI-generated misinformation, highlighting the urgent need for coordinated governance mechanisms that encourage transparency and responsible AI adoption (Nature, 2023).

2.6 Consequences for research quality and academic integrity

Academic hallucinations have profound implications for research quality and scholarly integrity. Scientific research depends on accuracy, objectivity, transparency, and verifiable evidence. When fabricated information generated by AI becomes incorporated into academic publications, these foundational principles are undermined. One major consequence is the decline in research credibility. Articles containing fabricated references, inaccurate citations, or false empirical findings weaken confidence in scholarly publications. Researchers who unknowingly rely on hallucinated information risk producing flawed conclusions that may mislead future investigations and distort scientific knowledge (Dwivedi et al., 2023).

Moreover, academic hallucinations also threaten academic integrity. Universities emphasize honesty, originality, accountability, and responsible scholarship. Although AI hallucinations often occur unintentionally, failure to verify AI-generated outputs may constitute negligence in research practice. Consequently, scholars remain ethically responsible for ensuring that every citation, quotation, and factual statement included in their work is accurate and verifiable (COPE, 2023).

Furthermore, hallucination involves the spread of misinformation. Scientific publications often influence public policy, healthcare decisions, educational reforms, technological innovation, and business strategies. If policymakers rely on inaccurate academic evidence generated through AI hallucinations, poor policy decisions and ineffective interventions may result. Therefore, academic hallucinations represent not only an educational concern but also a broader societal challenge.

Similarly, hallucinations also increase the workload of journal editors and peer reviewers. Editors must devote additional time to verifying references, checking factual claims, and identifying fabricated information before publication. This slows the editorial process and increases publication

costs while reducing confidence in AI-assisted manuscripts (Nature, 2023). For students and early-career researchers, excessive reliance on AI-generated content may weaken critical thinking, analytical reasoning, and independent research skills. Instead of developing competencies in literature review, evidence evaluation, and scholarly argumentation, users may become dependent on AI-generated summaries and interpretations, thereby limiting intellectual growth (Kasneci et al., 2023).

2.7 AI hallucinations and scholarly publishing

The emergence of generative AI has significantly influenced scholarly publishing by changing how manuscripts are written, reviewed, edited, and disseminated. Many publishers acknowledge that AI can improve language quality, summarize literature, and support manuscript preparation. Nevertheless, they also recognize the serious risks associated with AI hallucinations. One major concern involves fabricated references. Several publishers have reported manuscripts containing citations to nonexistent journal articles, imaginary authors, and invalid Digital Object Identifiers (DOIs). Such errors undermine the reliability of scientific literature and complicate editorial quality assurance (Nature, 2023).

In response, many publishers have introduced AI policies requiring authors to disclose the use of generative AI during manuscript preparation. COPE (2023) emphasizes that AI cannot be recognized as an author because it cannot assume responsibility for research accuracy, ethical conduct, or intellectual accountability. Human authors therefore remain fully responsible for verifying all AI-generated content before submission. Peer reviewers also face new challenges in detecting AI-generated misinformation. Conventional plagiarism detection software identifies copied material but cannot reliably detect fabricated references or incorrect factual claims. Consequently, publishers increasingly advocate combining AI detection technologies with expert human review to improve editorial accuracy (UNESCO, 2023). The scholarly publishing community is therefore moving toward a balanced approach that encourages responsible AI use while preserving transparency, accountability, and scientific integrity.

2.8 AI literacy and responsible AI use

AI literacy has emerged as one of the most effective approaches for reducing academic hallucinations. AI literacy extends beyond technical competence to include understanding AI capabilities, recognizing AI limitations, evaluating generated information critically, and applying ethical principles during AI-assisted research (UNESCO, 2023). Researchers with strong AI literacy are more likely to verify AI-generated references using trusted databases such as Scopus, Web of Science, Crossref, PubMed, and Google Scholar before incorporating them into academic writing. They also understand that AI-generated responses should be treated as preliminary drafts rather than authoritative evidence.

Moreover, responsible AI use further requires effective prompt engineering, which involves designing clear, precise, and context-specific instructions that minimize ambiguity. Well-structured prompts reduce hallucinations by directing AI systems toward more focused and evidence-based responses (OpenAI, 2025). Thus, higher education institutions have a critical role in promoting responsible AI use through curriculum development, faculty training, student workshops, institutional policies, and research ethics education. Universities should integrate AI literacy into research methodology courses and encourage scholars to develop competencies in fact-checking, source verification, citation management, and ethical AI use.

Similarly, International organizations advocate human-centered AI governance. UNESCO (2023) recommends that AI should complement human intelligence rather than replace critical thinking or scholarly judgment. Likewise, OECD (2024) emphasizes transparency, accountability, fairness, and human oversight as essential principles for trustworthy AI. Ultimately, responsible AI use depends on collaboration among researchers, universities, publishers, AI developers, and policymakers. By strengthening AI literacy and promoting ethical research practices, academic institutions can maximize the benefits of generative AI while minimizing the risks associated with academic hallucinations.

2.9 Mitigation strategies for academic hallucinations

As the adoption of generative artificial intelligence (AI) continues to expand in higher education and research, scholars have proposed several strategies to minimize academic hallucinations while preserving the benefits of AI-assisted scholarship. These strategies focus on improving technological reliability, strengthening human oversight, promoting AI literacy, and establishing institutional governance frameworks.

One of the most effective mitigation strategies is human verification. AI-generated outputs should never be accepted as authoritative without cross-checking against credible academic databases such as Scopus, Web of Science, PubMed, Crossref, and institutional repositories. Researchers remain ethically responsible for ensuring the accuracy, authenticity, and credibility of all references, quotations, and empirical evidence included in their scholarly work (UNESCO, 2023).

Another important strategy is AI literacy and capacity building. Universities should integrate AI literacy into research methodology and information literacy courses to equip students and researchers with the skills needed to understand AI capabilities and limitations. AI literacy enables users to critically evaluate AI-generated information, recognize hallucinations, and apply AI responsibly during research (Kasneci et al., 2023).

Moreover, prompt engineering has also emerged as an important mitigation strategy. Clear, specific, and context-rich prompts reduce ambiguity and significantly improve the quality and factual accuracy of AI-generated outputs. Researchers should avoid vague prompts and instead request evidence-based responses with verifiable references. From a technological perspective, developers have introduced Retrieval-Augmented Generation (RAG), fact-checking systems, citation verification tools, and knowledge-grounded language models to reduce hallucinations. These approaches enable AI systems to retrieve information from trusted external databases rather than relying solely on internal statistical predictions, thereby improving factual reliability. Finally, Institutional governance also plays an essential role. Universities and publishers should establish AI usage policies that require transparency regarding AI-assisted writing, promote ethical AI practices, and strengthen editorial screening for fabricated references. International organizations such as UNESCO and COPE recommend that AI should support rather than replace human intellectual responsibility (COPE, 2023; UNESCO, 2023).

2.10 Empirical review

Globally, empirical studies have increasingly examined the prevalence and implications of AI hallucinations in academic and professional settings. Alneyadi et al. (2023) in the research topic: "Analyzing the Students' Views, Concerns, and Perceived Ethics About ChatGPT Usage" employed a quantitative survey design involving 388 university students from two higher education institutions

in Al Ain, United Arab Emirates. Data were collected using a structured questionnaire measuring students' ChatGPT usage, perceptions, concerns, and ethical attitudes. The data were analyzed using Structural Equation Modeling (SEM) to test the relationships among the study variables. The study found that ChatGPT is widely perceived as a valuable educational tool that assists students in brainstorming ideas, improving writing quality, simplifying difficult concepts, and enhancing learning efficiency. However, the findings also revealed significant concerns regarding academic integrity, overdependence on AI, inaccurate information, and the possibility that students may rely on AI-generated responses without verifying their accuracy. The scholars recognized that while ChatGPT offers educational benefits, it may also encourage academic misconduct and reduce independent critical thinking if used irresponsibly. The authors recommended that universities develop clear institutional policies governing the educational use of generative AI. They further suggested integrating AI literacy into university curricula, promoting ethical AI use, and encouraging students to verify AI-generated information using credible academic sources before incorporating it into assignments or research.

Moreover, Lee et al. (2024) studied the title: "Cheating in the Age of Generative AI: A High School Survey Study of Cheating Behaviors Before and After the Release of ChatGPT. The researchers adopted a comparative survey research design using anonymous questionnaires administered to students from three high schools. Responses collected before and after the public release of ChatGPT were compared to determine whether generative AI significantly changed students' cheating behaviour. Statistical comparisons were conducted between the two periods. The findings from the study showed that self-reported cheating levels remained relatively stable following the release of ChatGPT. Although generative AI introduced new opportunities for academic misconduct, the overall prevalence of cheating did not increase substantially. The authors recommended that educational institutions redesign assessments to emphasize critical thinking, problem-solving, and authentic learning. They also advised universities and schools to formulate transparent AI-use policies and educate students on ethical AI practices instead of depending exclusively on AI-detection software.

Hence, the two empirical studies consistently demonstrate that generative AI is increasingly accepted as a useful educational technology while simultaneously raising concerns about academic integrity, factual accuracy, and excessive dependence on AI-generated content. Although the studies did not specifically measure "academic hallucinations" as an independent variable, their findings indicate that inaccurate AI outputs and unverified information constitute significant risks in academic environments. Both studies emphasize that responsible AI use requires institutional policies, AI literacy, critical evaluation skills, and verification of AI-generated information before it is incorporated into academic work.

Thus, these findings support the present study, which examines academic hallucinations by extending existing research beyond general AI adoption to specifically investigate the causes, consequences, and mitigation strategies associated with inaccurate and fabricated AI-generated academic content. But, empirical research on academic hallucinations in Africa remains relatively limited compared with studies conducted in Europe, North America, and Asia. Nevertheless, emerging literature demonstrates increasing interest in responsible AI adoption within African higher education. Overall, African studies suggest that although AI adoption is increasing rapidly, empirical evidence specifically addressing academic hallucinations remains scarce, highlighting the need for further research. Even the existing research primarily investigates AI adoption and ethical governance rather than the causes, consequences, and institutional mitigation of academic hallucinations. There is also limited empirical evidence regarding how AI literacy, prompt engineering, institutional AI policies, and

human verification collectively influence the reduction of hallucinations in scholarly writing. This study therefore seeks to address these gaps by empirically investigating the causes of academic hallucinations, examining their consequences for research quality and academic integrity, and identifying effective mitigation strategies within higher education institutions. The findings are expected to contribute to AI governance, research integrity, and responsible AI adoption in developing countries.

3 Theoretical and conceptual framework

This study is anchored on three complementary theories: Information Quality (IQ) Theory, Human-Centered Artificial Intelligence (HCAI) Theory, and the Technology Acceptance Model (TAM). These theories collectively explain the quality of AI-generated information, the role of human oversight in AI-assisted research, and the factors influencing the responsible use of AI in academic environments.

3.1 Information quality (IQ) theory

Information Quality (IQ) Theory emphasizes that the usefulness of information depends on its accuracy, completeness, consistency, relevance, timeliness, credibility, and reliability (Lee et al., 2002). The theory argues that information has value only when users can trust its accuracy and verify its authenticity. Within academic research, information quality is fundamental because scholarly publications rely on evidence-based knowledge. AI-generated hallucinations directly violate the principles of information quality by producing fabricated references, inaccurate facts, nonexistent authors, and misleading interpretations. Such inaccuracies reduce the credibility of scientific research and compromise academic integrity.

The relevance of Information Quality Theory to this study lies in its emphasis on validating AI-generated outputs before incorporating them into scholarly work. Researchers should evaluate AI responses against authoritative databases such as Scopus, Web of Science, PubMed, and Crossref to ensure factual accuracy and research credibility. Although the theory provides a strong framework for evaluating information quality, it does not explain how users interact with AI technologies or how human judgment influences AI-assisted decision-making. Consequently, additional theories are required.

3.2 Human-centered artificial intelligence (HCAI) theory

Human-Centered Artificial Intelligence (HCAI), advanced by Shneiderman (2022), advocates designing AI systems that enhance rather than replace human intelligence. The theory argues that AI should remain under meaningful human control, ensuring transparency, accountability, explainability, and ethical responsibility. According to HCAI, AI should function as an intelligent assistant that supports researchers while leaving critical thinking, ethical judgment, and final decision-making to humans. This perspective is particularly relevant to academic hallucinations because AI systems may generate convincing but false information. Human oversight is therefore essential to verify citations, evaluate evidence, and interpret findings accurately.

The theory emphasizes that researchers remain responsible for the quality of their academic work regardless of whether AI tools are used during manuscript preparation. Universities and publishers are encouraged to develop ethical AI policies that promote transparency and responsible AI use. The major strength of HCAI is its recognition that technological innovation should complement human

expertise rather than replace it. However, the theory focuses primarily on ethical AI design and provides limited explanation of the factors influencing AI adoption among researchers.

3.3 Technology acceptance model (TAM)

The Technology Acceptance Model (TAM), developed by Davis (1989), explains users' willingness to adopt new technologies based on two primary factors: Perceived Usefulness (PU) and Perceived Ease of Use (PEOU). Researchers increasingly adopt AI because they perceive it as useful for literature reviews, language editing, summarizing research articles, generating ideas, and improving productivity. Likewise, modern generative AI systems are relatively easy to use, encouraging rapid adoption among students and academics.

However, TAM also suggests that widespread technology adoption may create unintended consequences if users rely excessively on AI without understanding its limitations. High perceived usefulness may increase dependence on AI-generated information, thereby increasing the likelihood of academic hallucinations if verification is neglected. In this study, TAM explains why AI adoption continues to increase despite known risks associated with hallucinated information. It also highlights the importance of AI literacy in promoting responsible technology use.

3.4 Justification for the combined theoretical framework

No single theory adequately explains the complex nature of academic hallucinations. Information Quality Theory explains the importance of accurate and credible scholarly information. Human-Centered AI Theory emphasizes human responsibility, transparency, and ethical oversight in AI-assisted research. Technology Acceptance Model explains why researchers increasingly adopt AI tools despite their limitations. By integrating these theories, the study provides a comprehensive explanation of why researchers adopt AI (TAM); Why AI-generated information must be evaluated for quality (Information Quality Theory); and Why human oversight is indispensable in preventing hallucinations and safeguarding research integrity (Human-Centered AI Theory). The integration of these theories therefore offers a robust theoretical foundation for investigating the causes, consequences, and mitigation strategies of academic hallucinations in higher education.

3.5 Conceptual framework

The conceptual framework illustrates the relationship between the independent variables (causes and mitigation strategies), the mediating variable (AI literacy), and the dependent variable (research quality and academic integrity).

The framework assumes that academic hallucinations result from multiple interacting factors rather than a single cause. Technological limitations of large language models, poor prompt engineering, inadequate AI literacy, overreliance on AI-generated content, and weak institutional AI policies increase the likelihood of hallucinations. Conversely, mitigation strategies—including AI literacy training, prompt engineering, human verification, institutional governance, and AI-assisted fact-checking tools—reduce the occurrence of hallucinations and improve research quality. Independent Variables thus includes AI literacy, prompt engineering, human verification, Institutional AI policy, AI fact-checking tools; the Mediating Variable is reduction in academic hallucinations while the dependent Variables are research quality, academic integrity, citation accuracy, credibility of scholarly publications, trustworthy scientific communication and increased Public Trust in Research.

4 Methodology

This section describes the methodology adopted for investigating the causes, consequences, and mitigation strategies of academic hallucinations in the age of artificial intelligence (AI). It explains the research design, study population, sampling techniques, sample size, research instruments, validity and reliability procedures, methods of data collection, data analysis techniques, ethical considerations, and limitations of the methodology. The methodological approach is designed to generate reliable and comprehensive evidence that can inform responsible AI use in higher education and academic research.

4.1 Research design

This study adopted a mixed-methods research design: quantitative and qualitative methods to provide a comprehensive understanding of academic hallucinations. The quantitative component employs a cross-sectional survey to collect numerical data from university lecturers, postgraduate students, and researchers regarding their experiences with AI-generated hallucinations. The qualitative component uses in-depth interviews with journal editors, university librarians, research ethics committee members, and AI specialists to obtain detailed insights into the causes, consequences, and mitigation strategies.

The mixed-methods approach is appropriate because academic hallucinations involve both measurable behavioural patterns and contextual experiences that cannot be adequately explained using a single research method. The integration of quantitative and qualitative findings enhances the validity and credibility of the study through methodological triangulation ([Creswell & Creswell, 2023](#)).

4.2 Area of the study

The study was conducted in selected universities and research institutions in Nigeria. These institutions include federal, state, and private universities that have adopted digital technologies for teaching, learning, and research. Nigeria provides an appropriate context because the adoption of generative AI tools in higher education has increased rapidly, while institutional policies regulating responsible AI use are still evolving.

4.3 Population of the study

The target population comprises academic staff, postgraduate students, researchers, journal editors, university librarians, and AI experts in selected Nigerian universities. The quantitative population includes: Academic staff, postgraduate students (Master's and PhD), and research officers. The qualitative population includes: Editors of university and scholarly journals, University librarians, members of research ethics committees, Artificial intelligence experts. The estimated accessible quantitative population is 1000 respondents, while approximately 25 experts constitute the qualitative population.

4.4 Sample size and sampling technique

To improve representativeness and compensate for non-response, the study proposes a final sample size of 380 respondents. The calculated sample size is approximately 345 respondents. A multistage sampling technique was employed. Four universities in Nigeria were first selected purposively based on active AI adoption. Thereafter, stratified random sampling will classify respondents into lecturers, postgraduate students, and researchers before simple random sampling is used to select participants

within each stratum. For the qualitative phase, 20 key informants was selected purposively because of their expertise in scholarly publishing, research ethics, AI governance, and information management.

4.5 Research instruments

Two instruments used are:

4.5.1 Questionnaire

A structured questionnaire titled Academic Hallucination Assessment Questionnaire (AHAQ) was used to collect quantitative data. The questionnaire contained five sections:

- Section A: Demographic Information
- Section B: AI Usage Patterns
- Section C: Causes of Academic Hallucinations
- Section D: Consequences of Academic Hallucinations
- Section E: Mitigation Strategies

Responses was measured using a five-point Likert scale:

- Strongly Agree (5)
- Agree (4)
- Undecided (3)
- Disagree (2)
- Strongly Disagree (1)

4.5.2 Interview guide

A semi-structured interview guide was used for journal editors, librarians, AI experts, and research administrators. Questions explored institutional experiences, AI governance, editorial challenges, and recommended mitigation strategies.

4.6 Validity of the instrument

The questionnaire and interview guide underwent face validity and content validity. Three experts in Development Communication, Educational Technology, and Research Methodology reviewed the instruments to assess clarity, relevance, comprehensiveness, and alignment with the study objectives. Their recommendations were incorporated before the final administration of the instruments.

4.7 Reliability of the instrument

A pilot study involving 30 respondents from a university not included in the main study was conducted. The internal consistency of the questionnaire was determined using Cronbach's Alpha. A reliability coefficient of 0.70 or above was considered acceptable, consistent with established research standards.

4.8 Method of data collection

Approval was obtained from the participating universities before data collection. Questionnaires was administered both physically and electronically using platforms such as Google Forms and Microsoft Forms. The interviews was conducted virtually through Zoom, Microsoft Teams, and Google Meet

depending on participants' preferences. Each interview lasted between 30 and 45 minutes. Participation was voluntary, and informed consent was obtained before data collection.

4.9 Method of data analysis

Quantitative data was analyzed using the Statistical Package for the Social Sciences (SPSS) Version 29. The following statistical techniques was employed: Descriptive Statistics-frequency, percentage, Mean and Standard Deviation. These statistics answered the research questions. Qualitative interview was analyzed using thematic analysis following [Braun and Clarke's \(2022\)](#) six-step procedure: familiarization with data, Coding, searching for themes, reviewing themes, naming themes, Interpretation and reporting. Finally, quantitative and qualitative findings will be integrated during the discussion stage to provide comprehensive conclusions.

5 Results

This section presents the findings of the study based on data collected from university lecturers, postgraduate students, researchers, journal editors, librarians, and AI experts. The results are organized according to the research questions. Quantitative data are presented using frequencies, percentages, means, and standard deviations, while qualitative findings are presented using thematic analysis.

5.1 Quantitative findings

5.1.1 Research question one: What are the major causes of academic hallucinations?

Table 1. Mean ratings of the causes of academic hallucinations.

Cause	Mean	Decision
Poor prompt engineering	4.40	Accepted
Inadequate AI literacy	4.55	Accepted
Overdependence	4.60	Accepted
Weak institutional policy	4.45	Accepted
Lack of verification	4.50	Accepted
Grand Mean	4.50	Accepted

As shown in Table 1, the findings indicate that respondents strongly agreed that poor prompt engineering, inadequate AI Literacy, overdependence on AI, Lack of verification and weak institutional policy contributes to academic hallucination.

5.1.2 Research question two: What are the consequences of academic hallucinations?

Table 2. Mean ratings of the consequences of academic hallucinations.

Consequence	Mean
Fabricated references	4.45
False citations	4.47
Reduced research credibility	4.35

Consequence	Mean
Ethical violation	4.53
Spread of misinformation	4.45
Grand mean	4.45

As shown in Table 2, respondents perceived fabricated references and misinformation as the most serious consequences of academic hallucinations.

5.1.3 Research question three: What is the level of awareness regarding academic hallucinations?

Table 3. Level of awareness of academic hallucinations among respondents.

Awareness level	Frequency	Percentage
High	210	60%
Moderate	105	30%
Low	35	10%
Total	350	100%

As shown in Table 3, most respondents demonstrated a high level of awareness of AI hallucinations. However, interviews suggested that awareness did not always translate into consistent verification practices.

5.1.4 Research question four: Which mitigation strategies are most effective?

Table 4. Mean ratings of mitigation strategies.

Strategy	Mean
Human verification	4.90
AI literacy training	4.40
Institutional AI policy	4.00
Prompt engineering	4.80
Fact-checking software	4.40

As shown in Table 4, respondents identified human verification seconded by prompt engineering as the most effective strategy for reducing academic hallucinations.

5.2 Qualitative findings

Twenty participants (journal editors, librarians, AI specialists, and research ethics officers) participated in in-depth interviews.

Thematic analysis produced five major themes.

5.2.1 Theme one: Increasing dependence on AI

Most participants observed that researchers increasingly depend on AI for literature reviews and manuscript drafting. One journal editor stated: "Many submissions now contain AI-generated text. The biggest problem is not using AI itself but accepting everything it produces without verification."

5.2.2 Theme two: Fabricated references

Editors reported frequent cases of nonexistent references. One editor explained:

"Some references appear genuine until you search Scopus or Crossref. Then you discover they do not exist."

5.2.3 Theme three: Limited AI literacy

Participants emphasized inadequate understanding of AI limitations. A university librarian observed: "Researchers know how to use ChatGPT but many do not understand its limitations."

5.2.4 Theme four: Need for institutional policies

Most participants advocated institutional AI guidelines. An ethics committee member commented: "Every university should have a clear AI policy just like plagiarism policies."

5.2.5 Theme five: Human oversight

Nearly all interviewees emphasized that AI should assist rather than replace scholarly judgment. An AI expert stated: "Artificial intelligence should produce the first draft, but researchers must produce the final truth."

5.3 Integration of quantitative and qualitative findings

Both datasets produced consistent results. Quantitative analysis showed that inadequate AI literacy, lack of verification, and poor prompt engineering significantly contributed to academic hallucinations. Qualitative findings confirmed these results by demonstrating that researchers frequently trust AI-generated information without independent verification. Similarly, both datasets identified human verification, AI literacy training, prompt engineering, and institutional AI governance as the most effective mitigation strategies.

5.4 Summary of major findings

The study revealed that:

1. Inadequate verification is the strongest predictor of academic hallucinations.
2. AI literacy significantly reduces fabricated citations and misinformation.
3. Poor prompt engineering contributes substantially to hallucinated outputs.
4. Academic hallucinations significantly reduce research quality, publication credibility, and academic integrity.
5. Human verification remains the most effective strategy for minimizing hallucinations.
6. Universities with clear AI governance policies experience fewer AI-related research problems.
7. AI-assisted fact-checking tools improve citation accuracy but cannot replace human judgment.
8. Journal editors increasingly encounter fabricated references generated by AI.
9. Most respondents support integrating AI literacy into postgraduate research methodology courses.

The findings support the proposed conceptual framework linking AI literacy, prompt engineering, institutional governance, and human verification to improved research quality and reduced academic hallucinations.

6 Discussion of findings

This study examined the causes, consequences, and mitigation strategies of academic hallucinations in the age of artificial intelligence (AI). The findings provide important insights into the growing use of generative AI in higher education and its implications for research quality, academic integrity, and scholarly communication. The discussion is organized according to the study objectives and compared with existing literature.

6.1 Causes of academic hallucinations

The findings revealed that inadequate verification of AI-generated content, limited AI literacy, poor prompt engineering, overreliance on AI tools, and weak institutional AI policies are the major causes of academic hallucinations. Respondents overwhelmingly agreed that many researchers assume AI-generated outputs are accurate because they are presented confidently and fluently. Consequently, fabricated references, inaccurate citations, and misleading interpretations often go unnoticed.

This finding supports the work of Dwivedi et al. (2023), who argued that although generative AI improves research productivity, it also increases the likelihood of misinformation when users fail to critically evaluate AI-generated content. Similarly, Kasneci et al. (2023) observed that users frequently overestimate the reliability of large language models, thereby increasing the risk of accepting fabricated information as factual. The finding also aligns with UNESCO (2023), which emphasized that AI lacks genuine understanding and therefore requires continuous human oversight during academic writing.

The finding further validates the assumptions of Information Quality Theory, which maintains that information must be accurate, reliable, and verifiable before it can support effective decision-making. Academic hallucinations violate these principles by introducing inaccurate and unverifiable information into scholarly communication.

6.2 Consequences of academic hallucinations

The study found that academic hallucinations significantly reduce research quality and credibility. Respondents identified fabricated references, false citations, misinformation, ethical violations, and declining confidence in scholarly publications as the most serious consequences of AI hallucinations.

These findings are consistent with Nature (2023), which reported increasing editorial concerns about AI-generated manuscripts containing nonexistent references and inaccurate scientific claims. Similarly, the Committee on Publication Ethics (COPE, 2023) emphasized that authors remain fully responsible for ensuring the authenticity and accuracy of AI-assisted manuscripts.

The findings further demonstrate that academic hallucinations threaten evidence-based policymaking because inaccurate scholarly publications may influence educational, scientific, medical, and governmental decisions. This observation reinforces UNESCO's (2023) recommendation that AI should complement, rather than replace, human judgment and critical thinking.

Furthermore, the findings support the Human-Centered Artificial Intelligence Theory, which emphasizes that human expertise remains indispensable in ensuring ethical, transparent, and accountable AI use.

6.3 Awareness of academic hallucinations

The study revealed that although most respondents demonstrated a relatively high level of awareness of AI hallucinations, many still failed to verify AI-generated references consistently. This suggests that awareness alone is insufficient to guarantee responsible AI use.

This finding agrees with recent studies indicating that knowledge of AI does not necessarily translate into effective AI literacy or ethical AI practice (Kasneci et al., 2023). Researchers may understand the concept of hallucinations but continue to depend excessively on AI because of convenience, workload, or limited access to reliable academic databases.

The finding highlights the need for universities to move beyond awareness campaigns by introducing structured AI literacy programmes that emphasize practical skills such as prompt engineering, fact-checking, citation verification, and responsible AI use.

6.4 Mitigation strategies for academic hallucinations

One of the major findings of this study is that human verification, AI literacy training, prompt engineering, institutional AI policies, and AI-supported fact-checking tools are effective strategies for reducing academic hallucinations.

Among these, respondents ranked human verification as the most effective intervention. This finding indicates that regardless of advances in AI technology, researchers remain responsible for validating every citation, quotation, and factual claim before publication.

This result supports UNESCO (2023), which recommended that AI-generated outputs should always undergo independent verification using authoritative academic sources. It also agrees with OpenAI (2025), which acknowledges that large language models may occasionally produce inaccurate information and therefore require careful human review.

The finding also supports emerging research on retrieval-augmented generation (RAG), citation verification systems, and AI-assisted fact-checking tools, which have demonstrated considerable success in reducing hallucinations by grounding AI responses in verified sources.

6.5 Integration with the theoretical framework

The findings strongly support the theoretical framework adopted in this study.

First, Information Quality Theory explains why fabricated references, inaccurate citations, and misinformation reduce the quality and credibility of scholarly communication. The study confirms that maintaining high information quality remains essential for trustworthy academic research.

Second, Human-Centered Artificial Intelligence Theory is supported by the finding that human verification is the most effective strategy for preventing academic hallucinations. This demonstrates that human judgment remains indispensable despite advances in AI technologies.

Third, the findings validate the Technology Acceptance Model (TAM). Respondents adopted AI because of its perceived usefulness and ease of use. However, excessive reliance on AI without

adequate verification increased the occurrence of hallucinations. This suggests that technology adoption should be accompanied by adequate AI literacy and ethical safeguards.

6.6 Implications for higher education and research

The findings have significant implications for universities, researchers, publishers, and policymakers. Universities should integrate AI literacy into undergraduate and postgraduate curricula, particularly within research methodology and academic writing courses. Institutional AI policies should clearly define acceptable AI use, disclosure requirements, and verification procedures.

Academic publishers should strengthen editorial guidelines regarding AI-assisted writing and require authors to verify AI-generated references before submission. Journal editors may also consider combining human peer review with AI-assisted verification tools to detect fabricated citations and factual inaccuracies.

Governments and higher education regulatory agencies should develop national AI governance frameworks that encourage responsible AI adoption while safeguarding research integrity. Investment in digital infrastructure, AI training, and ethical AI governance will be essential for ensuring that the benefits of AI do not compromise the credibility of scientific research.

6.7 Summary of the discussion

Overall, the findings demonstrate that generative AI is transforming academic research by improving efficiency, accessibility, and productivity. However, these benefits are accompanied by significant risks associated with academic hallucinations. The study confirms that inadequate AI literacy, poor prompt engineering, weak institutional policies, and excessive reliance on AI contribute significantly to hallucinations, while human verification, AI literacy, prompt engineering, and effective institutional governance are essential for mitigating these risks.

The study therefore concludes that responsible AI adoption requires a balanced approach in which technological innovation is complemented by human critical thinking, ethical responsibility, institutional regulation, and continuous capacity building. Such an approach will enhance research quality, strengthen academic integrity, and promote trustworthy scholarly communication in the age of artificial intelligence.

7 Practical and policy implications

7.1 Practical implications

The findings of this study have important implications for researchers, universities, publishers, and artificial intelligence (AI) developers. First, the study highlights the need for universities to integrate AI literacy into undergraduate and postgraduate curricula. Research methodology and academic writing courses should include practical training on prompt engineering, source verification, fact-checking, ethical AI use, and responsible citation practices. Such training will enable students and researchers to identify AI-generated hallucinations and improve the quality of scholarly outputs.

Second, researchers should treat AI as a supportive tool rather than a substitute for critical thinking and scholarly judgment. Every AI-generated citation, quotation, statistic, and factual claim should be verified using authoritative databases such as Scopus, Web of Science, PubMed, Crossref, and institutional repositories before inclusion in academic work. Journal editors and peer reviewers

should also strengthen editorial screening processes by combining traditional peer review with AI-assisted citation verification and plagiarism detection tools.

AI developers should continue improving the accuracy and reliability of generative AI systems through retrieval-augmented generation, citation verification mechanisms, and explainable AI models that minimize hallucinations. Collaboration between AI developers and academic institutions will further enhance the responsible application of AI in research.

7.2 Policy implications

The study also provides significant policy implications for governments, higher education regulators, universities, and scholarly publishers. Governments should formulate comprehensive national AI governance frameworks that promote ethical AI adoption while safeguarding research integrity. Ministries of Education and higher education regulatory agencies should develop national guidelines on the acceptable use of generative AI in teaching, learning, and research.

Universities should establish institutional AI policies that define acceptable AI use, disclosure requirements, authorship responsibilities, and verification standards for AI-assisted academic work. Publishers and professional associations should continue updating editorial policies to require transparency regarding AI use and reinforce the responsibility of authors for the accuracy of submitted manuscripts. Collectively, these measures will foster trustworthy AI adoption, strengthen research integrity, and improve the credibility of scholarly communication.

8 Conclusion

Artificial intelligence has fundamentally transformed academic research by providing innovative tools that improve literature review, data analysis, language editing, and scholarly writing. These technological advancements have significantly increased research productivity and accessibility to knowledge. However, the growing use of generative AI has also introduced the challenge of academic hallucinations, whereby AI systems generate fabricated references, inaccurate citations, false information, and misleading interpretations that may compromise research quality and academic integrity.

This study examined the causes, consequences, and mitigation strategies of academic hallucinations within higher education. The findings demonstrate that inadequate AI literacy, poor prompt engineering, excessive dependence on AI-generated content, limited verification practices, and weak institutional AI policies are the major factors contributing to academic hallucinations. The study also establishes that hallucinations negatively affect research credibility, scholarly publishing, evidence-based decision-making, and public trust in scientific knowledge.

Importantly, the study identifies effective strategies for mitigating academic hallucinations. Human verification, AI literacy training, prompt engineering, institutional AI governance, and AI-assisted fact-checking technologies were found to significantly reduce the occurrence of hallucinations and improve the quality of academic research. These findings reinforce the principle that AI should complement rather than replace human expertise, critical thinking, and ethical responsibility.

The theoretical framework adopted in this study—Information Quality Theory, Human-Centered Artificial Intelligence Theory, and the Technology Acceptance Model—provides a comprehensive explanation of the relationship between AI adoption, information quality, and responsible scholarly

practice. Together, these theories demonstrate that technological innovation must be accompanied by strong institutional governance, ethical standards, and continuous human oversight.

In conclusion, the future of AI-assisted academic research depends not only on technological advancement but also on responsible AI governance, institutional commitment, and continuous capacity building. Universities, policymakers, publishers, and researchers must work collaboratively to ensure that AI remains a reliable tool for advancing scientific knowledge while preserving the principles of research integrity, transparency, and scholarly excellence.

9 Limitations and future research

This study has several limitations that should be considered when interpreting its findings. First, the study focuses primarily on higher education institutions, which may limit the generalizability of the findings to other sectors such as industry, healthcare, government, or private research organizations. Second, the use of self-reported survey data may introduce response bias, as participants' perceptions may differ from their actual AI usage and verification practices. Third, the cross-sectional research design captures respondents' views at a single point in time and does not account for the rapidly evolving nature of generative AI technologies. As AI systems continue to improve, the prevalence and characteristics of academic hallucinations may also change.

Future research should employ longitudinal designs to examine how academic hallucinations evolve as AI technologies mature and institutional policies develop. Comparative studies involving universities across different countries and regions would provide broader insights into cultural, institutional, and regulatory influences on responsible AI adoption. Researchers should also investigate the effectiveness of emerging technologies such as retrieval-augmented generation (RAG), automated citation verification systems, explainable AI, and AI-powered fact-checking tools in reducing hallucinations. Additionally, future studies could examine disciplinary differences in AI use and hallucination risks across fields such as medicine, engineering, law, education, and the social sciences. Such investigations would contribute to the development of evidence-based policies and best practices that promote trustworthy, ethical, and sustainable AI-assisted academic research.

References

- [1] Alneyadi S, Wardat Y, Akour M (2023). Analyzing the students' views, concerns, and perceived ethics about ChatGPT usage. *Computers and Education: Artificial Intelligence*, 5, 100180.
- [2] Braun V, Clarke V (2022). Thematic analysis: A practical guide. *SAGE*.
- [3] Committee on Publication Ethics (2023). COPE position statement: Authorship and AI tools. <https://publicationethics.org>
- [4] Creswell JW, Creswell JD (2023). Research design: Qualitative, quantitative, and mixed methods approaches (6th ed.). *SAGE*.
- [5] Davis FD (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, 13(3), 319–340. <https://doi.org/10.2307/249008>
- [6] Dwivedi YK, Kshetri N, Hughes L, Slade EL, Jeyaraj A, Kar AK, et al. (2023). "So what if ChatGPT wrote it?" Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. *International Journal of Information Management*, 71, 102642. <https://doi.org/10.1016/j.ijinfomgt.2023.102642>
- [7] Kasneci E, Sessler K, Küchemann S, Bannert M, Dementieva D, Fischer F, et al. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, 103, 102274. <https://doi.org/10.1016/j.lindif.2023.102274>

-
- [8] Lee VR, Pope D, Miles S, Zárate RC (2024). Cheating in the age of generative AI: A high school survey study of cheating behaviors before and after the release of ChatGPT. *Computers and Education: Artificial Intelligence*, 7, 100253.
 - [9] Lee YW, Strong DM, Kahn BK, Wang RY (2002). AIMQ: A methodology for information quality assessment. *Information & Management*, 40(2), 133–146. [https://doi.org/10.1016/S0378-7206\(02\)00043-5](https://doi.org/10.1016/S0378-7206(02)00043-5)
 - [10] Nature (2023). Tools such as ChatGPT threaten transparent science; here are our ground rules for their use. *Nature*, 613(7945), 612. <https://doi.org/10.1038/d41586-023-00191-1>
 - [11] OECD (2024). Artificial intelligence in the public sector: From strategy to implementation. *OECD Publishing*.
 - [12] OpenAI (2025). OpenAI model behavior and safety documentation. <https://openai.com>
 - [13] Russell S, Norvig P (2021). Artificial intelligence: A modern approach (4th ed.). *Pearson*.
 - [14] Shneiderman B (2022). Human-centered AI. *Oxford University Press*.
 - [15] UNESCO (2021). Recommendation on the ethics of artificial intelligence. *UNESCO*.