

Volume 2, Issue 1, October 2026: DOI: <https://doi.org/10.5281/zenodo.22869678>. Article history:
Received: December 14, 2025; Revised: March 9, 2026; Accepted: June 10, 2026

AI Literacy and Critical Information Evaluation: Assessing the Role of Academic Librarians in Mediating Student Verification of Synthetic and Hallucinated Scholarly Content

***Kingdom H. Kari**

Department of Library and Information Science, Federal University, Otuoke, Nigeria

ORCID ID: <https://orcid.org/0000-0001-7475-5172>

*Corresponding author email: karihk@fuotuo.ke.edu.ng

Victoria Ale

Department of Library and Information Technology The Federal Polytechnic,
Nasarawa, Nasarawa State, Nigeria.

ORCID ID: <https://orcid.org/0009-0001-7407-3419>

Abstract

Background: The proliferation of generative artificial intelligence across higher education has heightened the risk of algorithmic hallucinations and phantom citations contaminating student research. Despite the severity of these epistemic risks to scholarly integrity, empirical assessments evaluating how academic librarians actively mediate student verification of synthetic scholarly content remain limited.

Objective: Anchored in frameworks of cognitive risk, AI-mediated information literacy, and epistemic authority, this study evaluated the reported prevalence of phantom citations in reference consultations, assessed academic librarians' self-perceived AI literacy and mediation practices, and tested the predictive and moderating relationships governing student verification adoption.

Methodology: A cross-sectional survey design sampled 246 academic librarians across federal, state, and private universities. Data gathered via a validated questionnaire were analysed using

Page | 93

© 2026 The Author(s). Published by Gever Verlun Publishers LLP under a Creative Commons Attribution-NonCommercial-ShareAlike (CC BY-NC-SA 4.0) licence. Registered with the Corporate Affairs Commission, Nigeria (Registration No. 9817916).

descriptive statistics, Pearson product-moment correlation, standard multiple linear regression, and hierarchical moderated regression with simple slopes probing.

Result: Librarians reported a high prevalence of student encounters with synthetic literature (Mean = 3.73), dominated by citations pairing authentic authors with fabricated titles. Technical AI literacy significantly correlated with proactive verification mediation ($r = 0.584$, $p < 0.001$). Instructional guidance intensity strongly predicted students' adoption of lateral verification strategies (Beta = 0.592, $p < 0.001$), explaining 37.5% of the variance. Furthermore, perceived cognitive authority significantly moderated this relationship ($B = 0.174$, $p = 0.002$), more than doubling the positive impact of instructional frequency on phantom citation detection under high authority conditions.

Conclusion: Academic librarians serve as vital institutional safeguards against generative hallucinations. Enhancing librarians' technical AI literacy, instructional intensity, and cognitive authority is essential for instilling robust lateral verification habits and preserving scholarly integrity.

Keywords: Academic librarians, AI literacy, Generative AI, Hallucinations, Information integrity, Lateral reading, Phantom citations, Scholarly verification.

Introduction

The rapid integration of generative artificial intelligence into higher education learning and research environments has transformed traditional modes of academic knowledge acquisition, synthesis, and writing. Large language models and AI-powered conversational systems are increasingly deployed by undergraduate and postgraduate students to conduct preliminary literature scans, generate bibliographies, summarise scholarly discourse, and draft analytical arguments (Huang et al., 2024; Naamati-Schneider & Alt, 2024). While these algorithmic tools offer substantial efficiencies in information organisation, they present severe epistemic challenges that directly threaten academic standards and research integrity (Giray et al., 2026; Silva-Atencio, 2025). The most prominent among these challenges is the phenomenon of algorithmic hallucination: the generation of factually incorrect assertions, fabricated data, non-existent sources, and phantom bibliographic references presented with persuasive linguistic fluency (Han et al., 2026; Huang et al., 2025; Spennemann, 2026).

Algorithmic hallucinations in academic contexts do not merely represent benign computational errors; they constitute systemic distortions that compromise the cumulative architecture of scientific inquiry (Boretti, 2026). Recent empirical investigations demonstrate that large language models frequently invent plausibly sounding journal articles, complete with fabricated titles, mismatched author affiliations, and non-existent Digital Object Identifiers (Han et al., 2026; Naser, 2026). Because generative tools operate on probabilistic token prediction rather than genuine epistemic understanding, their outputs often mirror legitimate scholarly apparatus without grounding in verified literature (Huang et al., 2025; Rahman et al., 2026). When students

unreflectively assimilate these synthetic outputs into academic papers and dissertations, the downstream consequences undermine research competence, contaminate scientific repositories, and violate core principles of attribution and evidentiary validation (Boretti, 2026; Giray et al., 2026; Latif, 2025).

A fundamental driver of this vulnerability is the persistent deficit in students' critical AI literacy and the prevalence of cognitive over-trust (Kulal, 2025; Saddhono et al., 2024). Students frequently exhibit automation bias, projecting epistemic authority onto generative models due to their coherent syntax and authoritative tone (Kulal, 2025; Li & Xu, 2026). Traditional digital and information literacy competencies, which historically focused on navigating indexical search engines and evaluating pre-existing, static web resources, are insufficient for discerning synthetic fabrications dynamically constructed by non-deterministic models (Hossain, 2026; Naamati-Schneider & Alt, 2024). Consequently, scholars argue that information literacy must serve as the missing foundational core of artificial intelligence literacy, ensuring that operational tool fluency is subordinated to rigorous evidentiary verification and human accountability (Hossain, 2026; Ismail et al., 2026).

Within this shifting technological landscape, the role of academic librarians requires fundamental reassessment. Far from becoming obsolete intermediaries or mere ghosts in an automated machine, academic librarians occupy an indispensable pedagogical and evaluative niche in higher education (McCrary, 2026). Through instructional mediation, reference interactions, and information literacy curricula, academic librarians are strategically positioned to cultivate students' critical evaluation capabilities, instil verification routines, and recalibrate algorithmic trust (Aeni et al., 2026; Li & Xu, 2026; Offem & Arikpo, 2026). However, empirical understanding regarding how librarians actively mediate student verification of synthetic scholarly outputs, and the extent to which their instructional intensity influences student verification practices, remains fragmented.

Addressing this empirical gap is essential. This study investigates the role of academic librarians in mediating student verification of synthetic and hallucinated scholarly content. Specifically, it examines the relationship between librarian instructional guidance, librarian-perceived cognitive authority, and students' critical verification capabilities within university environments. By establishing a descriptive and correlational quantitative framework, this research provides empirical insights into how academic libraries can defend research integrity against algorithmic hallucinations in the generative era.

Literature Review

Conceptual Foundations: Generative AI, Hallucination Dynamics, and Synthetic Information Risks

The proliferation of generative artificial intelligence across academic institutions has introduced profound transformations to the mechanics of scholarly communication. Foundational to these systems are transformer-based large language models trained on massive, heterogeneous textual

corpora to predict sequential tokens probabilistically (Huang et al., 2025). Despite their remarkable syntactic coherence and contextual versatility, these architectures are fundamentally decoupled from verifiable reality, creating widespread vulnerability to algorithmic hallucinations (Huang et al., 2025; Naser, 2026). In technical taxonomy, hallucinations encompass both intrinsic distortions, where generated text contradicts source input, and extrinsic fabrications, where the model generates plausible assertions, empirical claims, or bibliographic entities that lack any real-world foundation (Huang et al., 2025; Rahman et al., 2026).

In scholarly domains, hallucination dynamics present acute risks to knowledge production and validation. Silva-Atencio (2025) demonstrates that while synthetic text generation delivers notable productivity gains, it exponentially elevates hallucination risks, introducing subtle informational confabulations that resist routine detection. In academic writing and legal research, where propositional truth and formal precedent are foundational, AI-generated factual inaccuracies directly distort decision-making frameworks and analytical reasoning (Latif, 2025). The danger of these hallucinations is exacerbated by the absence of epistemic uncertainty signalling within current commercial models; synthetic claims are articulated with identical semantic confidence as empirically verified truths (Giray et al., 2026; Kulal, 2025).

The literature on factuality evaluation highlights the technical difficulty of identifying and mitigating these errors. Rahman et al. (2026) review automated fact-checking pipelines, noting that while multi-agent cross-examination and external knowledge retrieval improve accuracy, generative models remain prone to subtle semantic drift and unverifiable citations. Naser (2026) systematically catalogues benchmark datasets, detection metrics, and post-hoc correction strategies, concluding that automated computational filters alone cannot eliminate hallucinated statements from generative outputs. Consequently, external human verification and rigorous institutional oversight remain indispensable counterbalances to synthetic misinformation within academic ecosystems (Boretti, 2026; Naser, 2026).

Phantom Citations and the Threat to Scholarly Integrity

A particularly damaging manifestation of generative hallucination in higher education is the fabrication of academic literature, termed phantom citations (Han et al., 2026). When prompted to support claims with peer-reviewed scholarship, large language models frequently synthesise references by amalgamating real author surnames, existing journal titles, authentic publication years, and plausible keywords into completely fictitious bibliographic entities (Giray et al., 2026; Han et al., 2026).

In a focused empirical investigation, Spennemann (2026) examined the uniqueness and reproducibility of hallucinated citations generated across diverse generative platforms, revealing that models routinely produce novel, unverifiable citations that vary widely across repeated prompts. This non-deterministic generation of fictitious scholarship frustrates conventional verification workflows and complicates bibliographic verification (Spennemann, 2026). Expanding on this phenomenon, Han et al. (2026) conducted an extensive empirical study of non-

existent and unverifiable references across scholarly repositories, documenting how phantom citations compromise the citation network, lead novice researchers into epistemic dead ends, and inflate perceived scholarly consensus around non-existent findings.

The intrusion of synthetic references into student dissertations, seminar papers, and journal submissions presents an existential challenge to research integrity (Giray et al., 2026). Boretti (2026) underscores that generative hallucinations threaten the core epistemological contract of academic publishing. Boretti argues that unverified, AI-generated citations create a cascade of fabricated authority, necessitating publisher-led, academically supervised verification mechanisms at every tier of the research pipeline. In pedagogical settings, when students submit assignments containing phantom citations, they commit unintentional academic fraud, exposing deep vulnerabilities in their information-evaluation protocols and revealing an over-reliance on conversational interfaces for literature retrieval (Giray et al., 2026; Han et al., 2026).

Cognitive Risks, Trust, and Critical Thinking in AI Adoption

The student tendency to accept synthetic fabrications without interrogation is deeply rooted in cognitive vulnerabilities. Kulal (2025) models the cognitive risks of artificial intelligence adoption, conceptualising how operational fluency and conversational responsiveness cultivate inappropriate epistemic trust. When students experience low cognitive friction in obtaining immediate answers from generative tools, their intrinsic critical thinking routines are dampened, leading to automated cognitive offloading and uncritical acceptance of generated claims (Kulal, 2025).

This dynamic is closely tied to students' underlying research self-efficacy and analytical maturity. In an empirical study of graduate students, Zhu and Yang (2026) established that while generative AI usage can enhance research output, its positive contribution to overall research competence is mediated by critical thinking. Furthermore, research self-efficacy acts as a vital moderator: students lacking robust methodological grounding and independent evaluative confidence are significantly more prone to accepting AI outputs uncritically, whereas students with high self-efficacy actively deploy critical scrutiny to interrogate algorithmic claims (Zhu & Yang, 2026).

The necessity of cultivating critical cognitive faculties alongside technological fluency is reinforced by Saddhono et al. (2024), who argue that AI literacy cannot be reduced to procedural programming or prompt design. Rather, it requires targeted analytical skills that enable users to identify algorithmic bias, question source provenance, and recognise the structural boundaries of machine learning outputs (Saddhono et al., 2024). In instructional environments, Huang et al. (2024) demonstrated that deliberately using ChatGPT as a foil—prompting students to critique, deconstruct, and verify model outputs—serves as an effective pedagogical intervention for cultivating higher-order evaluative literacy. Without such explicit interventions, students naturally default to cognitive ease, allowing persuasive synthetic text to bypass analytical scrutiny (Huang et al., 2024; Kulal, 2025).

Evolving Literacies: Moving from Digital Fluency to AIMIL

Traditional frameworks of digital and media literacy are fundamentally strained by generative technologies. Naamati-Schneider and Alt (2024) observe that the evolution of AI-powered assistants requires skills far beyond conventional digital literacy. While conventional digital literacy focused on navigating static hypertext and locating existing knowledge artefacts, the generative era demands dynamic evaluative capabilities capable of assessing real-time, autonomously synthesised content whose truth value is indeterminate (Naamati-Schneider & Alt, 2024).

In response to this pedagogical gap, Hossain (2026) asserts that information literacy constitutes the missing epistemic core of AI literacy. Hossain contends that contemporary institutional initiatives frequently over-emphasise tool proficiency at the expense of critical reasoning, source credibility evaluation, and ethical reflexivity. By repositioning classic information literacy principles—such as authority as constructed and contextual, and format as a process—at the centre of artificial intelligence instruction, educational institutions can equip learners with the epistemic tools required to interrogate algorithmic assertions (Hossain, 2026).

Extending this conceptual synergy, Ismail et al. (2026) propose the AI-Mediated Information Literacy (AIMIL) framework. Grounded in information science principles, AIMIL provides an architectural model that links technical AI mechanics with intelligent decision-making, provenance tracking, verifiable evidence, and human accountability (Ismail et al., 2026). Similarly, Aeni et al. (2026) operationalise an integrated, multidimensional framework for critical engagement with generative AI, adapting classical evaluation metrics (currency, relevancy, authority, accuracy, and purpose) to algorithmic outputs. Their findings demonstrate that when instructional curricula explicitly scaffold AI verification around formal information evaluation criteria, students exhibit significantly higher accuracy in identifying biased, fabricated, or synthetic academic claims (Aeni et al., 2026).

Academic Librarians as Cognitive and Epistemic Intermediaries

As higher education navigates the disruptive effects of generative technologies, academic librarians are redefining their professional agency and instructional mandate. McCrary (2026) directly challenges the techno-determinist view that artificial intelligence renders academic libraries redundant, examining whether librarians are becoming ghosts in the machine. McCrary asserts that the rise of synthetic environments enhances rather than diminishes the necessity of human agency, ethical oversight, and pedagogical intermediation provided by professional librarians (McCrary, 2026).

Academic librarians serve as frontline educators capable of converting generative AI from an integrity threat into a platform for enhanced critical inquiry (Offem & Arikpo, 2026). Offem and Arikpo (2026) show that when academic libraries integrate artificial intelligence into user education, they create guided learning environments where students actively explore database

structures, citation indexation, and algorithmic mechanics. By demonstrating the structural contrast between curated bibliographic databases (such as Scopus and Web of Science) and probabilistic language models, librarians provide students with the operational knowledge needed to execute lateral reading and verify synthetic references (Hossain, 2026; Offem & Arikpo, 2026).

The empirical efficacy of librarian intervention in algorithmic environments is substantiated by Li and Xu (2026). Investigating AI-mediated information intermediation during collaborative problem-solving, Li and Xu discovered that librarian guidance intensity and user perceptions of the librarian's cognitive authority significantly enhanced students' evaluative rigour and creative problem-solving outcomes. Their structural modelling demonstrated that when librarians intervene actively and authoritatively, students abandon uncritical reliance on AI outputs, increase cross-referencing against primary sources, and show greater discernment regarding algorithmic limits (Li & Xu, 2026).

Synthesis and Conceptual Hypotheses Formulation

The literature collectively reveals a crucial intersection: generative AI creates systemic hallucination risks and phantom citations (Boretti, 2026; Han et al., 2026; Spennemann, 2026), which students are cognitively primed to accept due to automated over-trust and insufficient critical AI literacy (Kulal, 2025; Zhu & Yang, 2026). While technical fact-checking algorithms remain imperfect (Huang et al., 2025; Naser, 2026; Rahman et al., 2026), grounding AI education in robust information literacy frameworks provides students with the structural tools to interrogate synthetic outputs (Aeni et al., 2026; Hossain, 2026; Ismail et al., 2026). Academic librarians, acting through intensive pedagogical mediation and established cognitive authority, represent a primary institutional defence mechanism against the spread of synthetic scholarly misinformation (Li & Xu, 2026; McCrary, 2026; Offem & Arikpo, 2026).

Despite these theoretical and empirical foundations, there remains an absence of quantitative research systematically evaluating academic librarians' self-efficacy, instructional practices, and direct mediation of synthetic verification in tertiary institutions. To address this empirical gap, the present study formulates the following operational research questions and hypotheses:

Research Question 1: What is the reported prevalence of student encounters with hallucinated and phantom citations during reference consultations, as observed by academic librarians?

Research Question 2: What is the current level of academic librarians' self-perceived AI literacy competence and critical verification mediation practices?

Hypothesis 1 (H1): There is a statistically significant positive relationship between academic librarians' technical AI literacy competence and their proactive mediation of critical verification frameworks.

Hypothesis 2 (H2): Librarian instructional guidance intensity significantly predicts students' adoption of lateral verification strategies for synthetic scholarly content.

Hypothesis 3 (H3): Librarians' perceived cognitive authority positively moderates the relationship between user instruction frequency and student efficacy in identifying phantom citations.

Methodology

Research Design

A quantitative descriptive and correlational cross-sectional survey design was employed for this investigation. This design was deemed appropriate because it enabled the systematic collection of standardised data to ascertain the current prevalence of student encounters with synthetic scholarly content, assess librarians' self-reported AI literacy proficiencies, and evaluate the statistical associations and predictive pathways between instructional guidance intensity, perceived cognitive authority, and verification mediation practices without experimental manipulation.

Population and Sampling Technique

The target population comprised professional academic librarians serving in public and private universities. The inclusion criteria were restricted to reference librarians, instruction librarians, subject liaison librarians, and digital scholarship specialists whose primary job descriptions entailed direct reference assistance, information literacy instruction, or research support services to undergraduate and postgraduate students. Administrative and technical processing staff who had no direct instructional or consultative contact with students were excluded from the sampling frame.

A multi-stage sampling procedure combining stratified and purposive sampling techniques was utilised. In the first stage, universities were stratified based on institutional ownership (federal, state, and private institutions) to guarantee institutional diversity. In the second stage, purposive sampling was applied to identify and recruit practising reference, liaison, and instruction librarians across these institutions. To ensure adequate statistical power for correlational and multivariate regression analyses, an a priori sample size estimation was conducted using G-Power software. Specifying a medium effect size of f^2 equal to 0.15, an alpha level of 0.05, and a statistical power of 0.95 across five main predictors, the minimum required sample size was calculated as 138 participants. To compensate for unreturned, incomplete, or aberrant responses typical of online web surveys, the questionnaire was distributed to 320 academic librarians, yielding 246 completely filled and usable responses, which satisfied the statistical requirements for descriptive, bivariate, and regression analyses.

Instrumentation

Data were gathered using a self-administered, structured questionnaire titled the Academic Librarians AI Literacy and Synthetic Content Verification Scale (ALAIL-SCVS). The instrument was developed following an extensive synthesis of validated scales and conceptual models derived from contemporary literature on cognitive risks, AI-mediated information literacy, and verification dynamics. The questionnaire was structured into five distinct sections:

Section A captured the demographic and professional attributes of respondents, including gender, institutional affiliation, years of professional librarianship experience, primary functional role (reference, instruction, or liaison), and weekly contact hours dedicated to research consultations and bibliographic instruction.

Section B evaluated the reported prevalence and operational typology of student encounters with hallucinated and phantom citations during reference consultations, measured across a 5-point frequency scale ranging from never (1) to very frequently (5). Items in this section captured instances of students presenting non-existent citations, fabricated Digital Object Identifiers, artificial abstracts, and misattributed author affiliations generated by conversational language models.

Section C measured academic librarians' self-perceived AI literacy and technical competence. Items were adapted from established digital literacy and artificial intelligence assessment frameworks, operationalised across five dimensions: understanding large language model architecture, prompt formulation awareness, algorithmic bias recognition, hallucination mechanics comprehension, and awareness of verification technologies. Responses were scored on a 5-point Likert scale ranging from strongly disagree (1) to strongly agree (5).

Section D assessed instructional guidance intensity and proactive mediation of critical verification frameworks. This section measured the frequency and depth with which librarians instructed students on lateral reading, authoritative database cross-referencing (such as Scopus, Web of Science, and PubMed), reverse citation tracing, and algorithmic interrogation techniques. Items were scored on a 5-point Likert scale ranging from never (1) to always (5).

Section E evaluated the perceived cognitive authority of academic librarians during instructional encounters. Items assessed the degree to which librarians perceived that students deferred to librarian guidance, trusted human bibliographic expertise over automated algorithmic outputs, and adjusted their research verification habits based on librarian interventions, measured on a 5-point agreement scale from strongly disagree (1) to strongly agree (5).

Validity and Reliability of the Instrument

To establish content and face validity, the initial draft of the questionnaire was subjected to expert scrutiny by a panel of four senior academics, comprising two professors of library and information

science and two senior measurement and evaluation specialists. The experts critically reviewed each item for clarity of expression, relevance to the research questions, construct alignment, and absence of ambiguity. Feedback regarding item redundancy and technical wording was incorporated, resulting in minor revisions to the wording of several indicator items.

Construct validity and internal consistency were evaluated through a pilot study administered to 35 academic librarians who were not part of the final sample. The pilot data were subjected to exploratory factor analysis using principal axis factoring with varimax rotation to establish the dimensionality and factorial integrity of each subscale. All items loaded satisfactorily onto their designated constructs with factor loadings exceeding the recommended threshold of 0.50. Internal consistency reliability was computed using Cronbach's alpha coefficients. The overall scale demonstrated excellent internal reliability with an alpha coefficient of 0.91. The individual subscales yielded the following Cronbach alpha values: reported prevalence of phantom citations (0.84), librarians' AI literacy competence (0.88), instructional guidance intensity (0.87), and perceived cognitive authority (0.85). Because all reliability coefficients exceeded the universally accepted benchmark of 0.70, the instrument was adjudged robust, reliable, and appropriate for final field administration.

Data Collection Procedure

The instrument was converted into an electronic survey hosted on an institutional Google Forms platform. A recruitment message detailing the purpose of the study, eligibility criteria, institutional approval details, and a direct web link to the survey was circulated via official academic library association mailing lists, professional liaison forums, and verified university library consortia channels.

The data collection phase lasted for six weeks. To maximise response rates, polite follow-up reminders were dispatched to the professional distribution lists at the end of the second and fourth weeks. Participation was entirely voluntary, and respondents were required to provide informed consent by ticking a mandatory affirmation box before accessing the survey questions. No personally identifiable markers, such as names, IP addresses, or personal telephone numbers, were recorded, ensuring complete anonymity and confidentiality of the participants.

Methods of Data Analysis

The collected quantitative data were cleaned, coded, and entered into the Statistical Package for the Social Sciences (SPSS) software for Windows. Both descriptive and inferential statistical techniques were deployed to analyse the research questions and test the research hypotheses at a 0.05 level of statistical significance.

Descriptive statistics, including frequencies, percentage distributions, arithmetic means, and standard deviations, were calculated to address Research Question 1 and Research Question 2, establishing the baseline distribution, central tendencies, and variability of reported prevalence

rates and librarian competency profiles. Mean values were interpreted against established criterion benchmark levels.

Inferential statistics were utilised to test the research hypotheses. Prior to hypothesis testing, preliminary diagnostic checks were performed to confirm that the data satisfied the assumptions of multivariate analysis, including normality (via skewness and kurtosis evaluation), linearity, homoscedasticity, and absence of multicollinearity (evaluated through variance inflation factors and tolerance values).

Hypothesis 1 was tested using Pearson product-moment correlation analysis to determine the direction and strength of the linear association between academic librarians' technical AI literacy competence and their proactive mediation of critical verification frameworks.

Hypothesis 2 was analysed using standard multiple linear regression to evaluate the predictive capacity of librarian instructional guidance intensity on students' adoption of lateral verification strategies, assessing the unstandardised regression coefficients, standardised beta weights, and overall coefficient of determination (R-squared).

Hypothesis 3 was evaluated using hierarchical moderated regression analysis and Hayes PROCESS macro to ascertain the moderating influence of librarians' perceived cognitive authority on the direct pathway between user instruction frequency and student efficacy in identifying phantom citations. Moderation effects were further probed through simple slopes analysis evaluated at conditional values of the moderator (mean minus one standard deviation, mean, and mean plus one standard deviation).

Results

Demographic and Professional Profile of Respondents

A total of 246 academic librarians completed the survey. Female respondents comprised 54.1% (n = 133) of the sample, while male respondents represented 45.9% (n = 113). In terms of institutional ownership, 48.8% (n = 120) were affiliated with federal universities, 34.1% (n = 84) with state universities, and 17.1% (n = 42) with private universities. Regarding primary functional roles, reference librarians constituted 41.5% (n = 102), instruction librarians represented 35.8% (n = 88), and subject liaison librarians accounted for 22.8% (n = 56). The mean duration of professional librarianship experience was 9.42 years (SD = 5.16), and respondents reported an average of 14.80 weekly contact hours (SD = 6.24) dedicated to direct student research consultations and bibliographic instruction.

Research Question 1: Reported Prevalence and Typology of Student Encounters with Hallucinated and Phantom Citations

The first research question examined the prevalence and operational manifestations of student encounters with AI-generated synthetic scholarly materials during library consultations. Responses were scored on a 5-point scale: Never (1), Rarely (2), Sometimes (3), Frequently (4), and Very Frequently (5). A criterion mean of 3.00 served as the baseline for moderate prevalence.

Table 1: Prevalence and Typology of Student Encounters with Synthetic and Hallucinated Content

Manifestation of Synthetic Content	Mean	SD	Decision
Students presenting non-existent citations generated by AI tools	3.88	0.89	High Prevalence
Citations with real author names paired with fabricated titles/findings	3.94	0.81	High Prevalence
Completely fabricated Digital Object Identifiers (DOIs)	3.65	1.02	High Prevalence
Inaccurate or synthetic journal publication volumes and page numbers	3.79	0.86	High Prevalence
Fabricated abstracts presenting non-existent empirical findings	3.42	0.98	High Prevalence
Misattribution of genuine scientific findings to wrong authors/sources	3.71	0.90	High Prevalence
Overall Cluster Mean	3.73	0.69	High Prevalence

The overall cluster mean of 3.73 (SD = 0.69) exceeds the 3.00 criterion benchmark, establishing that academic librarians regularly observe students presenting hallucinated and phantom scholarly content during reference and research consultations. The most prevalent manifestation was the presentation of citations pairing authentic author names with entirely fabricated article titles or findings (Mean = 3.94, SD = 0.81), followed closely by completely non-existent citations generated by conversational AI platforms (Mean = 3.88, SD = 0.89). Fabricated DOIs (Mean = 3.65, SD = 1.02) and artificial abstracts (Mean = 3.42, SD = 0.98) were also encountered with notable frequency.

Research Question 2: Academic Librarians' Self-Perceived AI Literacy and Verification Mediation Practices

The second research question assessed librarians' self-reported technical AI literacy proficiencies and their active mediation of verification strategies. Items were rated on a 5-point scale from Strongly Disagree / Never (1) to Strongly Agree / Always (5), with a criterion benchmark of 3.00 denoting moderate proficiency or practice.

Table 2: *Librarians' Self-Perceived Technical AI Literacy and Verification Mediation Practices*

Competency and Practice Indicator	Mean	SD	Level
<i>Technical AI Literacy Competence</i>			
Comprehension of large language model generative mechanisms	3.62	0.84	High
Ability to recognise signs of algorithmic hallucination	3.81	0.76	High
Knowledge of prompt formulation for verification purposes	3.19	0.91	Moderate
Understanding of algorithmic bias and training data limitations	3.55	0.83	High
Familiarity with automated fact-checking and AI detection tools	3.08	0.95	Moderate
Subscale Mean	3.45	0.68	Moderate-to-High
<i>Verification Mediation Practices</i>			
Instructing students to cross-reference AI citations in index databases	4.12	0.79	High
Teaching lateral reading techniques to evaluate source provenance	3.96	0.82	High
Demonstrating reverse DOI and metadata lookups	3.85	0.91	High
Guiding students to interrogate the credibility of AI claims	3.78	0.88	High
Embedding critical AI evaluation frameworks into library instruction	3.39	0.97	Moderate
Subscale Mean	3.82	0.71	High

Librarians reported moderate-to-high technical AI literacy (Mean = 3.45, SD = 0.68) and high engagement in verification mediation practices (Mean = 3.82, SD = 0.71). Within technical literacy, librarians rated their ability to recognise algorithmic hallucinations highest (Mean = 3.81, SD = 0.76), while familiarity with automated verification tools scored lowest (Mean = 3.08, SD = 0.95). In terms of active mediation, librarians reported a strong commitment to directing students to cross-reference AI citations against established indexing databases such as Scopus and Web of Science (Mean = 4.12, SD = 0.79) and imparting lateral reading techniques (Mean = 3.96, SD = 0.82).

Hypothesis 1: Relationship Between Librarians' Technical AI Literacy and Proactive Verification Mediation

Hypothesis 1 stated that there is a statistically significant positive relationship between academic librarians' technical AI literacy competence and their proactive mediation of critical verification frameworks. This was tested using Pearson product-moment correlation analysis.

Prior to analysis, the assumption of normality was confirmed, with skewness (-0.31 to 0.18) and kurtosis (-0.44 to 0.22) falling well within acceptable parameters (-1.00 to +1.00).

Table 3: *Pearson Correlation Between Technical AI Literacy and Verification Mediation Practices*

Variable	Mean	SD	1	2
1. Technical AI Literacy Competence	3.45	0.68	1	
2. Verification Mediation Practices	3.82	0.71	0.584*	1

Note: $p < 0.001$ (two-tailed), $N = 246$.

The bivariate correlation revealed a statistically significant, strong positive association between technical AI literacy competence and verification mediation practices ($r = 0.584$, $p < 0.001$). Librarians with higher levels of technical AI literacy exhibited greater active involvement in mediating critical verification strategies during student interactions. Therefore, Hypothesis 1 was supported.

Hypothesis 2: Librarian Instructional Guidance Intensity as a Predictor of Lateral Verification Adoption

Hypothesis 2 stated that librarian instructional guidance intensity significantly predicts students' adoption of lateral verification strategies for synthetic scholarly content. A standard multiple linear regression was conducted where students' lateral verification strategy adoption served as the dependent variable, and instructional guidance intensity served as the primary predictor while controlling for librarian years of experience and weekly contact hours.

The regression diagnostics satisfied all critical assumptions: the Durbin-Watson statistic was 1.94, confirming independence of residuals; variance inflation factor (VIF) values ranged from 1.04 to 1.32, indicating no multicollinearity; and visual inspection of P-P plots verified homoscedasticity and normality of residuals.

Table 4: Multiple Regression Analysis Predicting Student Lateral Verification Adoption

Model Predictor	B	SE B	Beta	t	p	95% CI
(Constant)	1.182	0.241	—	4.905	< 0.001	[0.707, 1.657]
Years of Experience	0.012	0.008	0.088	1.542	0.124	[-0.003, 0.027]
Weekly Contact Hours	0.009	0.007	0.076	1.338	0.182	[-0.004, 0.022]
Instructional Guidance Intensity	0.591	0.054	0.592	10.944	< 0.001	[0.485, 0.697]

Model Summary: $R = 0.612$, $R\text{-squared} = 0.375$, $Adjusted\ R\text{-squared} = 0.367$, $F(3, 242) = 48.36$, $p < 0.001$.

The overall regression model was statistically significant ($F(3, 242) = 48.36$, $p < 0.001$), explaining 37.5% of the total variance in students' adoption of lateral verification strategies ($R\text{-squared} = 0.375$). Instructional guidance intensity was a strong and statistically significant positive predictor ($B = 0.591$, $Beta = 0.592$, $t = 10.944$, $p < 0.001$). Conversely, neither professional years of experience ($Beta = 0.088$, $p = 0.124$) nor weekly contact hours ($Beta = 0.076$, $p = 0.182$) exerted a significant independent influence. Consequently, Hypothesis 2 was supported.

Hypothesis 3: Moderating Role of Perceived Cognitive Authority

Hypothesis 3 stated that librarians' perceived cognitive authority positively moderates the relationship between user instruction frequency and student efficacy in identifying phantom citations. This hypothesis was evaluated using hierarchical moderated regression via Hayes PROCESS macro (Model 1) with 5,000 bootstrap resamples. Continuous variables were mean-centred prior to model estimation to prevent multicollinearity.

Table 5: Hierarchical Moderated Regression Analysis with Perceived Cognitive Authority as Moderator

Model Step and Variable	B	SE B	t	p	95% CI
Step 1: Direct Effects					
Constant	3.648	0.038	95.82	< 0.001	[3.573, 3.723]
User Instruction Frequency (X)	0.312	0.051	6.118	< 0.001	[0.212, 0.412]

Perceived Cognitive Authority (W)	0.284	0.049	5.796	< 0.001	[0.187, 0.381]
Step 2: Interaction Effect					
Interaction Term (X * W)	0.174	0.057	3.053	0.002	[0.062, 0.286]

Moderation Summary: Model R-squared = 0.418, $F(3, 242) = 57.94$, $p < 0.001$; Interaction R-squared change = 0.022, $F(1, 242) = 9.32$, $p = 0.002$.

The addition of the interaction term accounted for a statistically significant increment in variance explained (R-squared change = 0.022, $F(1, 242) = 9.32$, $p = 0.002$), confirming a significant moderation effect.

To probe the nature of this moderation, a simple slopes analysis was performed across three conditional levels of perceived cognitive authority: low (one standard deviation below the mean), moderate (at the mean), and high (one standard deviation above the mean).

Table 6: *Conditional Effects of User Instruction Frequency at Levels of Perceived Cognitive Authority*

Level of Perceived Cognitive Authority	Conditional Effect (B)	SE	t	p	95% CI
Low (-1 SD = -0.72)	0.187	0.062	3.016	0.003	[0.065, 0.309]
Moderate (Mean = 0.00)	0.312	0.051	6.118	< 0.001	[0.212, 0.412]
High (+1 SD = +0.72)	0.437	0.065	6.723	< 0.001	[0.309, 0.565]

The simple slopes analysis demonstrated that while user instruction frequency significantly predicted student identification of phantom citations across all levels, the magnitude of the effect intensified progressively as perceived cognitive authority increased: from $B = 0.187$ ($p = 0.003$) at low cognitive authority, to $B = 0.312$ ($p < 0.001$) at the mean, and reaching $B = 0.437$ ($p < 0.001$) under high cognitive authority conditions. These results confirm that librarians' cognitive authority strengthens the positive influence of instructional frequency on students' phantom citation detection. Thus, Hypothesis 3 was supported.

Discussion

The findings of this study provide empirical substantiation of the critical, expanding role academic librarians occupy in mitigating the epistemic risks posed by generative artificial intelligence in

higher education. The investigation established a high reported prevalence of student encounters with synthetic and hallucinated scholarly literature (overall mean = 3.73), confirming that algorithmic confabulation is not an isolated technical anomaly, but an immediate operational challenge in tertiary learning environments. The most prevalent manifestation observed was the blending of authentic author identities with fabricated paper titles, methodology, or findings (mean = 3.94), closely followed by completely non-existent citations (mean = 3.88). These outcomes corroborate the observations of Han et al. (2026) and Spennemann (2026), who demonstrated that large language models systematically generate plausible yet fictitious bibliographic references that defy conventional superficial checks. Furthermore, as Boretti (2026) and Giray et al. (2026) warned, the uncritical assimilation of these phantom citations by students threatens the cumulative integrity of scientific literature, making expert intermediation indispensable.

Regarding the competency profiles of academic librarians, the results revealed moderate-to-high technical AI literacy (mean = 3.45) alongside high engagement in verification mediation practices (mean = 3.82). Librarians demonstrated substantial confidence in detecting algorithmic hallucinations and understanding underlying model mechanisms, though their self-reported familiarity with automated verification tools was comparatively moderate (mean = 3.08). The testing of Hypothesis 1 revealed a strong, statistically significant positive correlation between librarians' technical AI literacy and their proactive verification mediation ($r = 0.584$, $p < 0.001$). This finding aligns directly with the conceptual arguments advanced by Hossain (2026) and Ismail et al. (2026), who posited that information literacy serves as the essential core of AI literacy. Librarians who grasp the probabilistic nature and operational boundaries of generative models are markedly more inclined to deploy targeted pedagogies, steering students away from passive automation bias and towards rigorous evidence verification.

The multiple regression analysis supporting Hypothesis 2 demonstrated that librarian instructional guidance intensity is a robust predictor of students adoption of lateral verification strategies (Beta = 0.592, $p < 0.001$), accounting for 37.5% of the variance. Notably, structural demographic factors such as professional years of experience and weekly contact hours did not exert significant independent effects. This indicates that pedagogical approach and deliberate focus on verification strategies outweigh mere tenure or instructional volume. These findings empirically validate the arguments of Aeni et al. (2026) and Offem and Arikpo (2026), who highlighted that explicitly scaffolding evaluation criteria and teaching lateral interrogation routines—such as cross-referencing claims in curated citation indices—enables students to actively counter synthetic fabrications. Rather than rendering librarians redundant, the generative turn elevates the necessity of their direct instructional agency, as McCrary (2026) theorised.

Finally, the moderated regression analysis supporting Hypothesis 3 confirmed that librarians perceived cognitive authority significantly strengthens the relationship between user instruction frequency and student efficacy in identifying phantom citations (interaction $B = 0.174$, $p = 0.002$). Simple slopes probing showed that the positive effect of instruction frequency more than doubled from low cognitive authority conditions ($B = 0.187$) to high cognitive authority conditions ($B = 0.437$). This finding expands upon the conclusions of Li and Xu (2026) and Kulal (2025). When

students view academic librarians as authoritative, expert epistemic intermediaries, they are far more likely to internalise cautionary guidance, question conversational AI responses, and abandon uncritical cognitive offloading. Cognitive authority functions as a vital catalyst, translating instructional contact into genuine critical discernment and sustained verification habits.

Conclusion

This study demonstrates that the unchecked diffusion of generative artificial intelligence into student scholarship presents systemic threats to scholarly integrity, characterised prominently by phantom citations, fabricated Digital Object Identifiers, and synthetic abstracts. The findings establish that academic librarians serve as pivotal institutional safeguards against these epistemic hazards. Academic librarians' technical understanding of artificial intelligence directly drives their proactive mediation of critical verification techniques, while the intensity of their instructional guidance substantially promotes students' adoption of lateral reading and database cross-referencing. Crucially, the effectiveness of this instructional mediation is magnified when librarians hold strong cognitive authority among students. Ultimately, combating synthetic misinformation in the generative era cannot rely exclusively on algorithmic detection filters; it demands the revitalisation of information literacy frameworks, reinforced by the active, authoritative pedagogical agency of academic librarians.

Recommendations

To operationalise these findings within higher education institutions, the following policy and instructional recommendations are proposed:

- University library administrations should institute mandatory, continuous professional development programmes focused on technical AI literacy, algorithmic hallucination mechanics, and advanced lateral verification methodologies for reference, instruction, and liaison librarians.
- Academic libraries should revise standard information literacy curricula to formally integrate AI-Mediated Information Literacy (AIMIL) models, ensuring that instruction moves beyond basic database retrieval to address prompt interrogation, synthetic source deconstruction, and reverse metadata tracing.
- Higher education faculties should mandate structured, credit-bearing library verification workshops for undergraduate and postgraduate researchers prior to dissertation proposals, establishing collaborative instructional partnerships between subject lecturers and liaison librarians.
- Academic libraries should actively cultivate and project their institutional cognitive authority by establishing dedicated AI Verification and Research Integrity Clinics, positioning librarians as the primary institutional arbiters for evaluating problematic, AI-generated scholarly evidence.
- University management and library consortia should establish institutional protocols that mandate multi-database verification (using curated index repositories such as Scopus and

Web of Science) before any AI-assisted literature synthesis is accepted in academic assessment.

Declarations

Ethical Approval: This study was conducted in accordance with the Declaration of Helsinki. Formal institutional ethical approval was not required for this study as the survey was completely anonymous, voluntary, and carried no risk to participants.

Artificial Intelligence (AI) Use Statement: During the preparation of this work, the authors used Claude (Anthropic) to improve language, style, and clarity. Following the use of this tool, the authors thoroughly reviewed, edited, and verified the content, and take full responsibility for the final intellectual integrity of the publication.

Data Availability Statement: The datasets generated and analysed during the current study are available from the corresponding author upon reasonable request.

Funding Statement: The authors did not receive any specific grant or financial support from funding agencies in the public, commercial, or not-for-profit sectors for this study.

Conflict of Interest: The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Authors' Contributions: K. H. Kari conceptualised the study, designed the theoretical framework and survey instrumentation, supervised data collection, and drafted the initial manuscript. V. Ale contributed to the literature review, performed data cleaning and statistical analyses, and critically revised the manuscript for substantial intellectual content. Both authors read and approved the final version of the manuscript.

References

- Aeni, N., Hasan, M., & Dahri, N. A. (2026). Integrating information literacy and AI literacy: A multidimensional framework for critical engagement with generative AI in English language teaching. *Journal of Applied Research in Higher Education*, 1–16. <https://doi.org/10.1108/JARHE-04-2026-0768>
- Boretti, A. (2026). Hallucinations in generative AI: A threat to scholarly integrity and the urgent need for publisher-led academically supervised verification. *Energy Research & Social Science*, 136, Article 104720. <https://doi.org/10.1016/j.erss.2026.104720>

- Giray, L., Islam, M. E., & Glo, J. A. (2026). AI hallucinations in academic writing: Implications for research integrity. *Naunyn-Schmiedeberg's Archives of Pharmacology*, 399, 16827–16839. <https://doi.org/10.1007/s00210-026-05494-4>
- Han, C., Manoharan, S., Ye, X., & Speidel, U. (2026). Phantom citations: An empirical study of non-existent and unverifiable references in scholarly literature. *Journal of the Association for Information Science and Technology*. <https://doi.org/10.1002/asi.70104>
- Hossain, Z. (2026). Information literacy as the missing core of AI literacy: The evolving role of librarians. *Library Hi Tech News*, 1–3. <https://doi.org/10.1108/LHTN-12-2025-0220>
- Huang, C.-W., Coleman, M., Gachago, D., & Van Belle, J.-P. (2024). Using ChatGPT to encourage critical AI literacy skills and for assessment in higher education. In H. E. Van Rensburg, D. P. Snyman, L. Drevin, & G. R. Drevin (Eds.), *ICT education* (Communications in Computer and Information Science, Vol. 1862, pp. 105-118). Springer, Cham. https://doi.org/10.1007/978-3-031-48536-7_8
- Huang, L., Yu, W., Ma, W., Zhong, W., Feng, Z., Wang, H., Chen, Q., Peng, W., Feng, X., Qin, B., & Liu, T. (2025). A survey on hallucination in large language models: Principles, taxonomy, challenges, and open questions. *ACM Transactions on Information Systems*, 43(2), Article 42, 1–55. <https://doi.org/10.1145/3703155>
- Ismail, S. A., Yatin, S. F. M., Salleh, M. I. M., Shaharudin, S. M., Hassan, S., & Zain, W. M. (2026). From AI literacy to AIMIL: An information science framework for intelligent decision-making, evidence and human accountability. *The Journal of Intelligent Decision Making and Information Science*, 3(10S). <https://jidmis.org/index.php/jidmis/article/view/2813>
- Kulal, A. (2025). Cognitive risks of AI: Literacy, trust, and critical thinking. *Journal of Computer Information Systems*, 1–13. <https://doi.org/10.1080/08874417.2025.2582050>
- Latif, Y. A. (2025). Hallucinations in large language models and their influence on legal reasoning: Examining the risks of AI-generated factual inaccuracies in judicial processes. *Journal of Computational Intelligence, Machine Reasoning, and Decision-Making*, 10(2), 10–20. <https://morphpublishing.com/index.php/JCIMRD/article/view/2025-02-07>
- Li, R., & Xu, X. (2026). AI-mediated information intermediation in collaborative problem-solving: Effects of librarian guidance intensity and cognitive authority perceptions on student teams' creative outcomes. *Library & Information Science Research*, 48(2–3), Article 101415. <https://doi.org/10.1016/j.lisr.2026.101415>

- McCrary, Q. D. (2026). Are we ghosts in the machine? AI, agency, and the future of libraries. *The Journal of Academic Librarianship*, Article 103181. <https://doi.org/10.1016/j.acalib.2025.103181>
- Naamati-Schneider, L., & Alt, D. (2024). Beyond digital literacy: The era of AI-powered assistants and evolving user skills. *Education and Information Technologies*, 29, 21263–21293. <https://doi.org/10.1007/s10639-024-12694-z>
- Naser, M. Z. (2026). Hallucinations in generative artificial intelligence and large language models: Tests, datasets, detection and correction methods. *Language Resources and Evaluation*, 60, Article 64. <https://doi.org/10.1007/s10579-026-09938-4>
- Offem, C. O., & Arikpo, S. O. (2026). Artificial intelligence as a tool for promoting information literacy among library users. *International Journal of Jurisprudence, School Management and Civic Education*, 1(1). <https://unilaws.org/INTERJACE/article/view/460>
- Rahman, S. S., Islam, M. A., Alam, M. M., Zeba, M., Rahman, M. A., Chowa, S. S., Raiaan, M. A. K., & Azam, S. (2026). Hallucination to truth: A review of fact-checking and factuality evaluation in large language models. *Artificial Intelligence Review*, 59, Article 70. <https://doi.org/10.1007/s10462-025-11454-w>
- Saddhono, K., Suhita, R., Rakhmawati, A., Rohmadi, M., & Sukmono, I. K. (2024). AI and literacy: Developing critical thinking and analytical skills in the digital era. In *2024 International Conference on IoT, Communication and Automation Technology (ICICAT)* (pp. 360–365). IEEE. <https://doi.org/10.1109/ICICAT62666.2024.10922871>
- Silva-Atencio, G. (2025). Synthetic Data in AI: Performance Gains versus Hallucination Risk. *Artificial Intelligence and Applications*. <https://doi.org/10.47852/bonviewAIA52026620>
- Spennemann, D. H. R. (2026). How Unique Are Hallucinated Citations Offered by Generative Artificial Intelligence Models? *Publications*, 14(3), 38. <https://doi.org/10.3390/publications14030038>
- Zhu, H., & Yang, S. (2026). The Impact of Generative AI Use on Graduate Students' Research Competence: The Mediating Role of Critical Thinking and the Moderating Role of Research Self-Efficacy. *Behavioral Sciences*, 16(2), 304. <https://doi.org/10.3390/bs16020304>