Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 4, Issue 1 (2026)

Online ISSN 3007-3197
Print ISSN 3007-3189

http://amresearchreview.com/index.php/Journal/about

Investigating Lived Experiences of Trusting AI to Misconceptions in Karachi, Pakistan, at the University level: An Interpretative Analysis

Huma Safdar
PhD Scholar, Department of Education, Institute of Business Management, Karachi
Email: huma.safdar33@gmail.com

Dr. Kiran Hashmi
Head of Department, Department of Education, Institute of Business Management, Karachi Email: kiran.hashmi@iobm.edu.pk

Mazhar Ali Shah
Assistant Director, STEDA, School Education and Literacy Department, Government of Sindh, Sindh Teacher Education Development Authority
Email: mazharsheerazi@gmail.com

ABSTRACT

In this study, university students in Karachi are implicitly analyzed in terms of beliefs in, information interpretation, and response to artificial-intelligence-generated fake content in the academic setting. The purposive sampling made possible the stimulated-recall type of interviews based on the interpretative phenomenological analysis (IPA) when each participant read one or two academic responses provided by AI and then thought over them in semi-structured Zoom interview. Thematic analysis, in keeping with the Brain and Clark paradigm, results in students considering AI to be a convenient scholarly resource and they trust the results of the latter due to their disguised character hence breeding the formation of misconceptions at the unconscious level. The cases of error detection occurred when the responses of AI were opposed to the existing information. These results indicate the need to introduce AI-literacy courses in universities and set ethical standards to students. This study shows the necessity of empirical testing in the research of the future with the use of interventions related to literacy by referring to the example of Karachi.

Keywords: AI-Generated Fake News, Misconceptions, Digital Literacy, Scholarly Integrity, Generative AI, Critical Thinking.

Introduction

Artificial intelligence (AI) has become a revolutionary force in different disciplines of life, including education. The rising usage of AI for educational purposes, like assignment completion and exam preparation, has increased concerns about the critical thinking of students at the university level. Overreliance on AI can lead students to develop unquestioning trust in AI-generated information without judging its accuracy (Zaman et al., 2025). This overdependence may create a false sense of surety about AI outputs, ultimately weakening students' critical thinking abilities

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 4, Issue 1 (2026)

Online ISSN 3007-3197
Print ISSN 3007-3189
http://amresearchreview.com/index.php/Journal/about

(Gillespie, 2018). Moreover, accepting AI-generated content without evaluation can contribute to inaccurate information (Abuzar et al., 2025). When students fail to critically evaluate AI outputs, they internalize and even share incorrect or misleading information (Brennen et al., 2021).

The rapid rise of misinformation is one of the most prominent challenges in today's digital era (Lewandowsky, 2023; Swire-Thompson & Lazer, 2020). Misinformation, defined as inaccurate or misleading content shared without ill purposes, has taken on a new aspect with the development of generative artificial intelligence (Ireton & Posetti, 2018; Ecker et al., 2022). AI tools like Gemini, Copilot, ChatGPT, and Google Bard are not only changing the dimension of knowledge production but also how it is trusted and used. These technologies generate convincing, confident outputs that may contain factually inaccurate information or "hallucinations" (Bandara, 2024), which can mislead students to accept faulty content as a credible information (Sundar, 2008). While researchers in health and political disciplines have examined the impact of misinformation, generative AI has been given limited consideration, especially regarding its educational implications and how it affects students' learning behaviors, reasoning, and trust in digital content when students are overdependent on AI (Kim et al., 2024; Pérez-Escolar et al., 2023).

This study addresses a critical and underexplored gap by examining how university students unknowingly trust AI, embrace incorrect information, form misconceptions, and later on discover or fail to discover these errors in educational settings. Though Pakistan is growing digital integration at higher education, there is limited research available on how students unconsciously trust false output generated by AI, and navigate AI-generated misinformation once identified as produced by AI when they use it for the completion of various academic tasks. Based on the theoretical frameworks of digital literacy (Ng et al., 2023) and critical thinking (Facione, 2023), this research explores how students unintentionally trust AI, assimilate incorrect information, create misconceptions, and later recognize or fail to recognize these errors. By exploring these experiences through an interpretative phenomenological lens, the study offers insight into the challenges and opportunities that AI presents for academic integrity and construction of knowledge in a fast-evolving educational lexicon. Therefore, it contributes to global discussions while stressing the need for localized strategies in AI literacy and responsible use of AI.

Background of the Research

In Sindh province, university students in Karachi over rely on AI-generated content to complete academic work. Though these tools can provide personalized learning experiences, their outputs contain inaccurate, biased, and incomplete information, which hinders students' analytical and critical thinking skills (Fasahat et al., 2025; Stuchlíková et al., 2024). Digital incompetence, combined with an exam-driven, memorization-focused educational culture, increases students' vulnerability to misinformation, raising concerns to academic integrity and reflective practices (Kanwal et al., 2023). Although the swift adoption of AI in urban universities in Karachi, no study has yet examined how students unknowingly trust AI, accept false content as true, leading to misconceptions, and later detect or fail to detect these inaccuracies in academic settings, revealing a significant empirical gap.

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 4, Issue 1 (2026)

Online ISSN 3007-3197
Print ISSN 3007-3189

http://amresearchreview.com/index.php/Journal/about

Likewise, trends are observed across developing countries of Asia, where students are increasingly interacting with generative AI tools at the higher education level. Research indicated that although AI can provide personalized learning and timely feedback, students' ability to critically evaluate information generated by AI remains diminutive (Moustaghfir & Birgui, 2024). Effective use of AI in education demands stringent content verification. Students and educators need to recognize the limitations of AI outputs, verify the information with reliable sources, and promote digital literacy and critical thinking skills among students to reduce the risks of misinformation (Geoffrey et al., 2024) particularly in countries where exam driven culture is the focus of education.

Globally, Western studies in higher education contexts have analyzed the effects of AI-generated misinformation on students' learning and reasoning. While AI tools can improve student academic engagement and performance, misinformation has been shown to weaken critical and analytical thinking skills (Bender et al., 2021; Dipto et al., 2024). These findings highlight the double-edged nature of AI in education. On one side it offers meaningful learning opportunities, while on the other side it poses major concerns to intellectual growth and academic honesty.

In short, while AI has the ability to transform learning, its potential to generate misinformation poses a critical challenge**.** No qualitative research has yet explored lived experiences of students with AI-generated misinformation in Karachi Pakistan, stressing an immediate need for context-specific studies to inform educators and policymakers.

The misinformation created by AI is a concern to critical and analytical thinking of students, academic integrity in Pakistani higher education, and students, who are not aware of AI outputs, tend to trust them excessively. In an exam-oriented culture with a high emphasis on rote learning and low levels of digital literacy, misleading AI-generated content will affect the students (Rehman Uddin et al., 2025; Irshad et al., 2025). Even though AI tools can help with the implementation of customised learning and feedback, over-reliance on AI tools without critical evaluation can compromise critical thinking and valuable learning (He Li et al., 2025; Muhammad Aqeel et al., 2025). As a result, the problem should be subject to academic research.

Unless this problem is properly investigated, the reliance of students on misinformation generated by AI could result in lower academic levels, the reduction of true learning, and the influence on ethical behavior. As one of the centres of fast digital growth, Karachi highlights the pros and cons of AI usage in education. However, empirical research specifying the unintentional trust of AI by students, internalisation of false information, development of misconceptions, and the about-to-be-identified-or-not-identified nature of the errors does not exist, which indicates a significant gap in the existing knowledge (Rehman Uddin et al., 2025).

Research Questions

This interpretative phenomenological study aimed at exploring how students unconsciously place their trust in AI, internalise false information, develop misconceptions, and later end up finding (or not finding) these misconceptions. The paper also examined how these phenomena affect the critical thinking and the

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 4, Issue 1 (2026)

Online ISSN 3007-3197
Print ISSN 3007-3189

http://amresearchreview.com/index.php/Journal/about

learning of students in learning institutions. Although the world is worried about the potential of AI to spread false information (Zou et al., 2023; Hou et al., 2025), limited studies have looked at how students in developing countries like Pakistan are dealing with the threat. The gap in this work is that it examines the experiences of students living in the fast-digitising academic environment in Karachi with inadequate digital literacy. Therefore, this study explored,

How do university students in Karachi unknowingly trust and use AI-generated answers in their academic work without realizing some information may be wrong?

How do students internalize wrong AI information and develop misconceptions during learning?

How do students later discover (or fail to discover) AI errors, and what strategies do they use to check the information?

Research Objectives

To explore how university students in Karachi unknowingly trust and use AI-generated academic answers.

To examine how wrong AI information becomes part of students' understanding and leads to misconceptions.

To identify how students discover (or fail to discover) AI errors and what verification strategies they use.

Significance of the Study

This study contributes to the body of research because it fills in a significant gap on how students unwillingly depend on the conclusions provided by AI and internalise false information as a result, a phenomenon which has not been examined sufficiently in this area (Rehman Uddin et al., 2025). It builds upon modern perspectives on AI-related digital literacy and misinformation by exploring how custom-made AI-generated deception can affect the critical thinking, learning behaviors, and academic activity of students (Geffrey et al., 2024), particularly in the case of poor digital literacy.

The paper highlights the importance of preparing learners with strong digital literacy skills in order to be critical of AI-generated information. Through cultivating an understanding of AI systems and the constraints they necessarily introduce, students will be able to use these tools in a responsible manner without falling into overdependence and encourage students to adopt a more critical and thoughtful approach to digital technologies (Ng et al., 2021). In addition, the results can be used by policymakers and educators to develop strategies to strengthen digital literacy, advance critical thinking, and reduce the risks of misinformation by AI in higher education (Rehman Uddin et al., 2025).

Scope of the Study

The qualitative investigation focused on the B.Ed learners in three different universities in Karachi using a qualitative phenomenological design. Due to the semester final examination, the the semi structure interviews via zoom with the participants did not come easy. This study outlines the ways students unsuspectingly use AI-generated content, introduce errors to their learning process, create the false

Online ISSN 3007-3197
Print ISSN 3007-3189

beliefs, and become conscious or even unconscious of their mistakes in academic settings. The studies consider such constructs as inadvertent trust by students in AI-generated academic material, internalisation of misinformation, misconception formation, and the following finding (or lack thereof) of the errors. By making use of an interpretative phenomenological approach and interviewing based on stimulated recall, the study provides a clear delimitation in terms of population, setting, time and methodology, which, in turn, guarantee both feasibility and analytical depth

.

Limitations & Ethics

Though such a qualitative study does not aim at being generalisable, its strength is the depth and richness of the insights about university students in Karachi Pakistan who unwillingly rely on AI-generated information, internalise falsehood, form misconceptions, and ultimately (or fail to realise) errors. The purposive sampling technique provided an in-depth understanding of the cognitive engagement, digital literacy behavior, and critical thinking of students in a lowly researched setting.

Since it is such a sensitive topic to analyze the reliance on AI and possible academic malpractice, the ethical aspect needs be considered the most important. The informed consent were obtained, and the confidentiality was ensured to the participants, who would also be allowed to withdraw without any consequences at any moment. The information would be safely kept and only utilised to conduct the research, and thus, the participants would be free to share their experiences without compromising their educational status.

Operational Definitions

Artificial Intelligence (AI):

Refers to computer systems or programs capable of performing tasks that typically require human intelligence, such as reasoning, problem-solving, and generating text or responses

AI-Generated Text-Based Misinformation:

Any inaccurate, misleading, or false information produced by AI tools in written form, which users may interpret as factual or reliable.

Literature Review

Theoretical Framework

The current research is grounded in two conceptually connected theories Digital Literacy (Ng et al., 2023) and Critical Thinking (Facione, 2023). These theories are a conceptual construct to explain the process by which students access, assess, and react to AI-generated misinformation in academic settings.

Digital literacy, especially in the scope of generative artificial intelligence, goes beyond technical skills that are primitive. It preempts the ability to critically read, evaluate, and morally handle digital content (Baskara, 2025). With the growing introduction of AI into the academic world, students will need to develop advanced skills such as the ability to detect bias, verify information, and determine the quality of AI results. However, institutionalization of the described frameworks is not extensive and learners are not sufficiently equipped to spot misinformation (Tiernan

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 4, Issue 1 (2026)

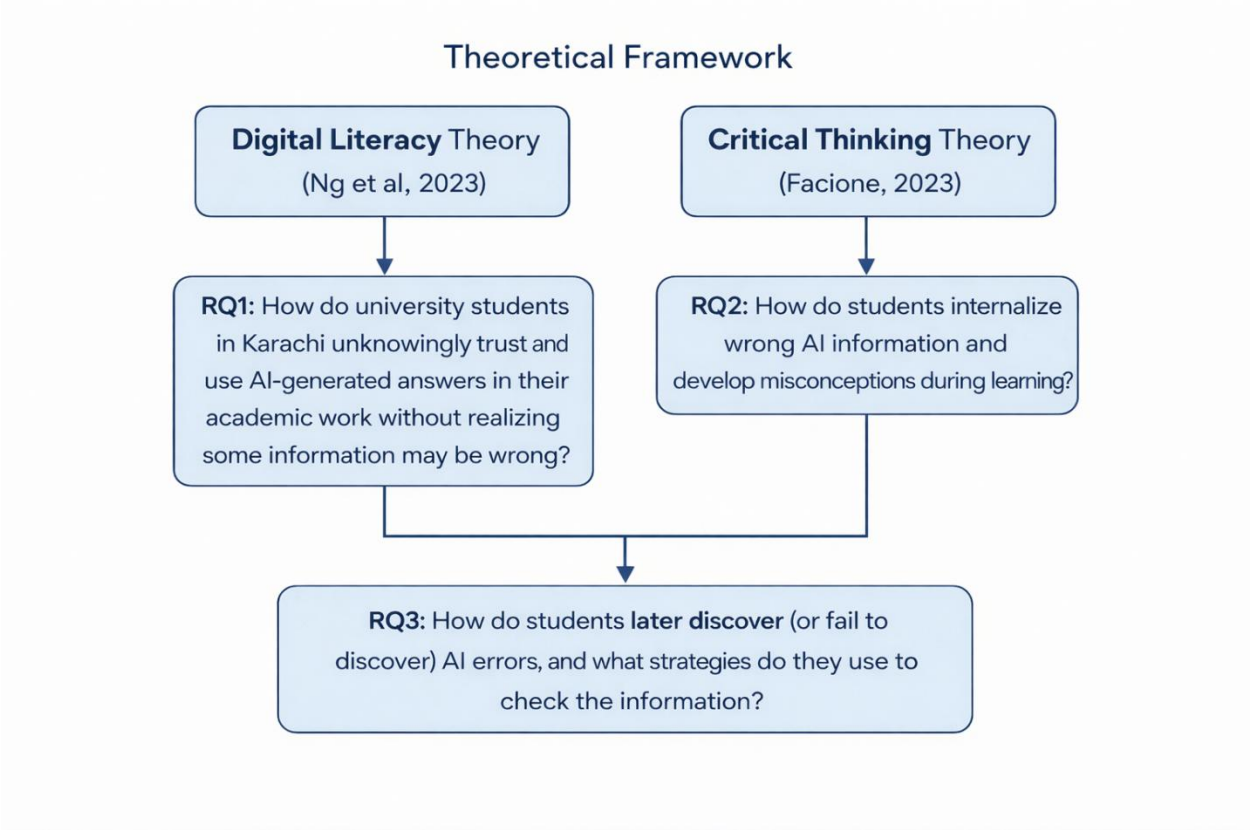
Online ISSN 3007-3197
Print ISSN 3007-3189

http://amresearchreview.com/index.php/Journal/about

et al., 2023). In addition, the misleading nature of generative AI can destroy the trust and confidence of the students (Schmitt and Elechais, 2023). Therefore, media-literacy programs are necessary to training students to identify AI-created false information, especially in settings with rote learning (Washington, 2023). Digital literacy synthesis and critical pedagogy can promote the capacity of learners to question and challenge AI products (Chomintra, 2023).

Critical Thinking framework supplements the digital literacy aiming at guiding students towards logical conclusions and problem-solving (Hou et al., 2025; Essien et al., 2024). Empirical research shows that cognitive engagement can diminish when there is a dependency on AI to produce scholarly work (Zou et al., 2023), and this effect is particularly strong in systems that are centered on assessment, like the one in Pakistan, where the habit of producing AI-generated plausible but inaccurate information has been reported (Schmitt and Elechais, 2023). Hence, the frameworks underpin the present investigation in analyzing how students assess and react to AI-generated misinformation in their academic environment.

Refer to Figure 1.



Theme 1: The experiences of the university students with AI-generated content.

In various learning environments, students are increasingly using AI to solve their academic work; although, in most cases, they do not have the critical ability to assess the content produced by AI. Mosae and Kaushal (2025) conducted a research in India and found that students were open to AI support but were unable to detect false information. Mohammed et al. (2024) suggested that AI did not improve critical and

Online ISSN 3007-3197
Print ISSN 3007-3189

reflective thinking in students in Iraq; however, it was useful in compositions. The same trends occurred in the West. The study conducted by Choi et al. (2024) established that students were overconfident in AI, thus reducing their enthusiasm to check the results. Chen et al. (2024) highlighted the importance of digital literacy education in ethical and critical interaction with AI.

The misunderstandings may emerge in the background, when the students make use of AI. Artificial intelligence produces false information that seems persuasive and thus influences the users to think in a certain way by demonstrating compatibility with the already existing beliefs or expectations. Therefore, the students might internalise false information unwillingly (Kidd et al., 2023). As an example, AI-assisted misinformation can include more details, sound rather personal, and offer conclusions in a logical form, which makes it hard to distinguish between a reliable and misleading one (Zhou et al., 2023). This is more dangerous, as AI can generate information with high certainty, capitalize on the tendency of users to hold onto the beliefs, and create text which might be regarded as credible, similar to that produced in a scholarly writing (Xu et al., 2023). Therefore, the unspoken creating of misconceptions is one of the biggest challenges in the era of generative AI.

Learners often do not have systematic practices to question or contextualize the outputs of AI (Sedlbauer et al., 2024). Although Karachi is prominent in the digital realm of Pakistan, no previous research has explored the experience of students receiving the misinformation of AI in this area, which makes the gap a priority of the current study.

Theme 2: Students Have Confidence in AI Responses and Memorize the wrong data.

Misinformation created by AI affects the cognition of students in both complicated and contradictory ways. Overly trusting AI, students are inclined to lessen personal cognitive activities, which leads to cosmetic learning (Lee et al., 2025). Although AI technologies can help in some simple tasks, like memory and summarization, they can spread false information (Essien et al., 2024). Even though personalized learning capabilities enable metacognition (Sardi et al., 2025; Saude et al., 2024), excessive dependence on them compromises independent thinking (Zhai et al., 2024) since the content generated by them has the potential to look valid. Ruiz–Rojas et al. (2024) have observed that although AI can help to reduce order, it requires constant critical analysis. Moreover, according to the study by Safarov et al. (2025), people constitute a significant portion of the non-experts who believe that AI is more knowledgeable, and thus, tend to place more trust in AI responses when the result is polished and well-written, which paves the way to believing that AI information requires no verification. In a similar manner, Zhai et al. (2024) warn against overreliance on AI as it may destroy the ability of the students to think critically.

Cognitive theories, including the Taxonomy created by Bloom, show that AI completes tasks that require lower-order thinking, which makes students rely more on higher-order thinking skills analysis and assessment (Gonsalves, 2024). However, the trust that students have towards AI depends on their AI literacy (Hou et al., 2025). Numerous students are aware of the shortcomings of AI, and they are ready to obtain effective digital literacy education (Zou et al., 2023). It is confirmed by international

Online ISSN 3007-3197
Print ISSN 3007-3189
http://amresearchreview.com/index.php/Journal/about

studies that the integration of digital and critical literacy is the key to reducing the effect of misinformation (Gonsalves, 2024; Hou et al., 2025).

Theme 3: Coping and Digital Literacy Practices.

Students are also reporting the combination of individual response tactics and institutional aid to cope with AI misinformation. One of the Chinese universities reported that students were aware of the danger of AI but did not have the skills to act effectively (Zou et al., 2023). According to Wang (2024), students in the United States are concerned about fake news and also require formal education in AI literacy. These results point to the need of direct training in AI content evaluation.

Institutions are reacting with specific schemes. According to Chaudhuri and Terrones (2024), AI-oriented information-literacy programs are created to strengthen AI skills of students. Akakpo (2023) suggests workshops and systematic assistance to develop the habit of critical evaluation. Revised models of literacy that include academic ethics, fact-checking, and discernment are offered by Latin American scholars (Ponce Rojo et al., 2025). These models deal with threats to AI and attempt to build cognitive autonomy. However, tertiary education in Pakistan is still under-represented in this study area.

Critical Summary

The literature constantly proves that, despite the widespread use of AI tools, students often need digital and critical literacy to distinguish and cope with AI-produced fake news (Mosae & Kaushal, 2025). Such a shortage can lead to development of overdependence on AI, loss of analytical skills and tolerance of fake information. Whereas several countries have launched AI literacy programmes, the bulk of empirical studies are Western or East Asian-based, and the evidence of Pakistan, especially in the fast-digitising cities like Karachi is limited. In this light, the current research aims to fill this gap by exploring the perception and reaction of students in Karachi to AI-generated fake academic content using paradigms of digital literacy and intellectual thinking.

Methodology

The research employing an interpretivist paradigm is aimed at understanding how people create meaning out of their experience living in particular social and cultural milieu (Creswell and Poth, 2023). The use of AI-generated academic content and its impact on university students, including their unintentional trust in it, internalisation of misleading information, development of misconceptions, and their subsequent or lack thereof awareness of having been misled in an academic context made the phenomenological research design the choice.

Interpretative Phenomenological Analysis (IPA) is especially suitable when investigating the experiences of students in relation to using emergent technologies like generative AI because it predicts the mechanisms that allow individuals to make sense of the events and create meaning. Modern IPA advice emphasizes the ability to represent personal interpretations and sense-making in natural settings in detail (Smith et al., 2022; Woodend, 2024). IPA was used in this research to explore the trust and the use of AI generated answers among the students in detail, development of

Online ISSN 3007-3197
Print ISSN 3007-3189

misconceptions, and the identification of errors and information verification in the end. The stimulated recall interviews supported this methodological position by using the own AI-generated examples of students as a technique to encourage them to think rigorously of the interpretation, trust, and fact-checking strategies.

Participants and Sampling

The sample was comprised of eight B.Ed. (four years) students of three different Karachi universities. The sample of participants was selected purposely because they use AI tools, including ChatGPT and Grammarly, in their studying in real life, which guarantees that people who have had first-hand experience with depending on AI-generated information, including situations when they have accepted false information without knowing and changed their mind later (or not at all). The fact that the researcher was recruited via his network of academic contacts also brought about a level of trust and familiarity with the researcher and this enabled the researcher to get deep insights.

Data Collection

The collection of the data was done in two stages. During Phase 1, participants were to bring one to two AI-generated responses that they had used in the past to carry out academic assignments. Phase 2 involved semi-structured Zoom stimulated recall interviews, whereby the AI-generated samples were used as a cue to discuss the interpretations, trust and misconception formation and the discovery (or lack thereof) of inaccuracies among participants. The interviews were held in a period of about thirty to forty minutes, and the participants were reminded to share their experience with AI generated contents openly. All recordings of interviews were made audio and transcribed to be analyzed later.

Data of the qualitative interviews were evaluated by adopting the thematic analysis technique as suggested by Braun and Clarke (2006). After collecting data, they were repeatedly transcribed and listened to after recording. Immersion was then achieved by repeating the transcripts. Ai tools such as ChatGPT and grammarly were also used in this procedure. Applicable text sections were found and coded and the codes were then divided into more comprehensive themes. These themes were constantly reconsidered and updated to provide a real image of the leading tendencies in the dataset and correspond with the research questions of the study, thus helping to discover the important thematic findings presented by the participants.

Ethical Considerations

Ethical procedures were given much attention. Participants were provided with an informed consent form before the interviews, where the goals of the research were explained to them, their right to leave the interviews at any time, and guaranteed anonymity and confidentiality. Since such a topic as the use of AI and the aspect of academic honesty may be rather sensitive, the interviews were held in the non-judgmental and supportive atmosphere. All information was safely kept and used solely in academic purpose by the policy of the institution.

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 4, Issue 1 (2026)

Online ISSN 3007-3197
Print ISSN 3007-3189

http://amresearchreview.com/index.php/Journal/about

Data Analysis

The paper questions the phenomenon whereby university students unconsciously believe the information created by AI without understanding the possibility of inaccuracies and fake information. The participants were expected to give one to two answers generated by AI that have been used in scholarly situations, and these answers were applied in the stimulated recall interviews. In these interviews, the participants elaborated on the ways they interpreted the AI responses, why they trusted them, whether they identified by the inaccuracies (or not) and the methods they used to confirm the information.

The AI as an Academic Solution.

The participants found AI tools expedient and efficient and they conceptualised these tools as immediate academic support. Students were using AI as a way to complete tasks bypassing and reducing the amount of work and effort. One interviewee commented: ChatGPT was a one- click answer (Respondent #1, Student). This quote shows that AI was not viewed as a helpful assistive tool but as an answer in a readymade version; the convenience of AI therefore influenced academic habits of students, encouraging them to consume information at a rapid pace instead of building knowledge step-by-step. The analysis also shows that students think that the constant use of AI lowered their effort and gave them a stress-free and relaxed learning experience. One of the participants mentioned: I got lazy (Respondent #8, Student). This reaction shows that the convenience of AI affected academic behaviors, leading to the decrease in the independent work and the decline of critical thinking activities.

Trust in AI Answers

The analysis shows that students relied on the output of AI based on the fact that they were well-structured and scholarly fluent. The participants were keen to note that the way the answers of AI are presented made them believable.

One participant explained:

It answers in a very proper manner. (Respondent # 2, Student).

This reaction refers to the fact that the trust was created as a result of the tone and format of AI output as opposed to confirmed accuracy. The students, therefore, were prone to consider information as true, especially when the answer was in an authoritative or formal manner.

Quite to the contrary, one of the participants answered that they do employ AI in academia, but they do not necessarily accept the information literally. I have already used AI-generated content more than once, not necessarily to copy-paste something, but to ensure that I am indeed able to organize my thoughts or provide me with an alternative perspective on what I am thinking.

Subconscious Misperception Development.

The discussion demonstrates that the misperceptions were developed because students accepted AI-generated information without doubts. According to the participants, they have at times found themselves after the fact discerning that AI content was incorrect or misguided.

One of the participants posted their response when exposed to false material:

I was appalled at it. (Respondent # 3, Student).

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 4, Issue 1 (2026)

Online ISSN 3007-3197
Print ISSN 3007-3189

http://amresearchreview.com/index.php/Journal/about

This reaction points out that the initial misconceptions were subconscious. Students were not ready to get AI to give wrong answers and this anticipation predisposed them to adopt misinformation in their knowledge until it was corrected by previous knowledge or contradiction.

Realization of Errors

The respondents clarified that they were more prone to AI mistake when they already possessed knowledge about the subject. In cases where the students were not familiar with the subject matter, it was hard to determine whether the output of AI was correct or not.

A participant stated:

And when you know about something like that then you can easily make a judgment (Respondent # 4, Student).

Another respondent expressed his or her mind as:

In the beginning, when I was using it, I depended more on it. However, over time, with a constant usage, I noticed that I am superior to AI. There are answers that are so ambiguous to the extent that it does not even know I am querying something. Second, it is unable to solve mathematical duties- it is not familiar with mathematics (Respondent # 7, Student).

It means that the detection of errors relied on the background knowledge of students. The only way that students were able to notice misinformation was that they needed to have sufficient knowledge to compare the answers given by AI to those they already possessed.

AI Inaccuracy as an Academic Risk.

The other discovery is that students felt that the inaccuracies of AI was a genuine academic problem. Participants provided some examples of AI providing wrong responses.

One participant stated:

It is not able to do mathematical problems at all (Respondent # 5, Student).

Another shared:

It combines numerous indistinct things and presents them to me. That is why I have gone through it several times when I heard that the information is not right (Respondent #8, Student).

This reaction indicates that even the confident production of AI may be erroneous, posing a threat to academic education. Incorrect information is liable to be transferred into the assignment and the learning outcomes when the students do not check the answers provided by AI.

Verification Strategies in order to check information.

The participants talked about how to prove the information generated by AI. The review demonstrates that verification was often done by cross-checking with Google and other sources, but the checking was often done under some doubtful circumstances and was not a regular activity.

A participant stated:

I check cross-referencing Google. (Respondent # 6, Student).

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 4, Issue 1 (2026)

Online ISSN 3007-3197
Print ISSN 3007-3189

http://amresearchreview.com/index.php/Journal/about

Another voiced:
I just use more books instead of internet information or JPT whichever it is providing to me. I use a lot of textbooks. There, I discover that I do not necessarily get all the information right since I check it." (Respondent # 1, Student).
The answer provided means that students tried to minimize misinformation by checking content, but they did it selectively. The verification was mainly done when the AI output was suspicious or where the accuracy of the task was needed.
Meaning-Making When One realizes There are mistakes.
Based on the analysis, it can be seen that, having been exposed to misinformation, students reconsidered their assumptions about the role of AI in learning. Some of the respondents explained that they reduced the role of AI and redefined it as an aiding resource.
One participant explained:
I consider it a point of departure, rather than the ultimate solution (Respondent # 7, Student).
Another participant stated,
What impact does misinformation on AI have on critical thinking and analytical thinking by Pakistani students? Not everybody is that intelligent to know that this is misinformation that I have to cross-check and verify the same (Respondent # 2, Student).
This reaction proves that students acquired reflective meaning-making patterns when they found out that they were wrong. They did not rely on AI entirely but rather restructured AI as a draft and took the responsibility of checking, learning, and final decision-making. Nonetheless, students do not possess the skills that are required to detect fake information.

DISCUSSION

The paper analysed how the students believe AI without questioning the level of its accuracy, which may lead to internalized falsehood, misperceptions, and slow (or no) realization of errors. In the interview, participants were asked to think about AI-generated content performed in academic work using stimulated recall IPA interviews. The results showed that students perceive AI as a fast study aid and believe it because it is presented in a structured manner and seems to be accurate. Nevertheless, it is insidious and thus generates misinformation which results in the development of misconceptions among the students which they later realize. Once an error is detected, students explain how they change their ways of making sense by developing partial verification and how they reframe AI as an entry point but not an authority figure. The findings are addressed within the frames of the theoretical frameworks of Digital Literacy (Ng et al., 2023) and Critical Thinkingary (Facione, 2023) that enables the research to place its contribution into the context of current research on AI misinformation and learning.

RQ1: Trust and Use of AI-Generated Answers.

An interesting observation was also made on the basis of the data: students habitually interpreted the outputs produced by AI as a time-saving academic tool, assuming the role of AI as a supplementary tool designed to provide pre-crafted solutions. Thus, AI

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 4, Issue 1 (2026)

Online ISSN 3007-3197
Print ISSN 3007-3189
http://amresearchreview.com/index.php/Journal/about

seems not only as a convenient, but as a pervasive part of academic life of students, as a replacement of work-intensive research and synthesis. This fact helps confirm global research that students are using AI mostly to be faster and more convenient shortcuts, as well as to accomplish assignments instead of boosting their knowledge level (Mosae and Kaushal, 2025; Mohammed et al., 2024). In the current research, the researchers have gone further to build on the previous literature by incorporating the method of stimulated recall hence demonstrating not only the frequency of using AI, but also the interpretive processes whereby learners assign meaning to AI-generated answers.

This interpretive pattern highlighted an essential weakness regarding the digital literacy approach because it is associated with students who demonstrate operational digital competencies (i.e., the ability to access and use AI tools) but often lack evaluative digital literacy skills, namely, the ability to judge credibility and the accuracy of the context (Ng et al., 2023). The shortcoming applied consequentially in educational settings where students would not have to go through basic learning activities like comparison, synthesis and independent reasoning (activities that are inherently associated with critical engagement) when AI is interpreted as a solution, as opposed to an origin.

The results showed that the trust of student in AI is highly mediated by the professional image of AI output. Instead of being based on a credibility test that has been made out of evidence, it is the product of surface sensing such as fluency, confident tone, presentation structure. These findings were consistent with the studies showing the tendency of students to over-trust AI-generated content due to the persuasiveness of the language and the inclusiveness of answers (Choi et al., 2024) and supported the literature that explains the tendency of the generative AI to create texts that are similar to academic traditions, which consequently hides misinformation (Chen et al., 2024).

The current literature explained the fact that students tend to adopt a heuristic form of shortcut: they evaluate the acceptability of an answer based on its aesthetic value as opposed to its factual value. The practice is a threat in the academic scenes, as most students might not be adequately knowledgeable in the domain to detect the correctness of information availed. The core thinking would entail active judgment and scrutiny of information (Facione, 2023). But in case the AI provides assertive and well-constructed answers, the students might stop questioning and trust as well. In that regard, the crucial aspect of AI literacy interventions needs to be the development of the idea that stylistic fluency might not always be associated with informational correctness.

RQ2: Misinformation and Misconception Development

One of the main contributions of this research is the discovery of the development of unconscious misconceptions as a logical extension of the unquestioning use of AI. These statistics demonstrated that misinformation spreads when students are ready to absorb AI information unintentionally, incase when they are not familiar with the subject or when the response is similar to the seemingly respectable scholarly position. This result is in line with the research that argues that generative AI is reliable in generating misinformation, which a user internalizes without knowing

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 4, Issue 1 (2026)

Online ISSN 3007-3197
Print ISSN 3007-3189
http://amresearchreview.com/index.php/Journal/about

(Kidd et al., 2023; Zhou et al., 2023; Xu et al., 2023). The current inquiry not only verifies these statements but goes further to show that the formation of misconceptions is not only a consequence of the AI error, but it also depends on the learning environment where students are in search of quick solutions, as well as less effort.

The evidence highlighted the dangers of AI misinformation, which need not take the form of blatant lies but instead, it can be structured as smoothly integrated, scholarly-appearing replies, which is incorporated into the knowledge systems of learners. The study of digital literacy highlighted the importance of critical interpretation and moral analysis of harmless online information (Tiernan et al., 2023). The results created the sense that without the mentioned competencies, AI may creepily disrupt the process of building knowledge in academics. In practice, this means that the risks associated with AI-related misconceptions should be dealt with by higher education institutions in Pakistan as systemic curriculum problems that need to be addressed with the help of structured intervention and not on the self-awareness of individual students (Tiernan et al., 2023; Chen, 2023; UNESCO, 2024).

The data also explained that the students detected AI errors when outputs opposed their previous knowledge or when they gave inconsistencies that were confirmed through experience. This supported the previous studies that show that the error rate of the AI outputs is imbalanced and highly dependent on the expertise of the user, with non-experts being less susceptible to misinformation (Sedlbauer et al., 2024). The value of this work is the exegesis of the meaning-making process: once students recognize such mistakes, they reevaluate the role of AI in education and redirect it as a starting point, thus separating themselves to the point of passive acceptance.

The previous studies hypothesize that overuse of AI may reduce cognitive involvement of students and undermine critical thinking, especially in the cases where the students use AI as a source of unquestionable information (Zou et al., 2023; Zhai et al., 2024). Nonetheless, the results provided a more subtle description. Respondents mentioned that they became more cautious and verified more after realizing that AI made mistakes and that this led to more thoughtful and systematic thinking. Notably, learners with prior knowledge were more likely to compare the outputs of AI with their prior knowledge, identify inconsistencies, and perform a critical evaluation. Therefore, it is proposed in this study that AI does not have a homogenous negative impact on critical thinking; instead, critical thinking may be positively impacted with the help of verification and reflective evaluation, when students are critical in using AI and apply previous knowledge to critical thinking.

RQ3: Verification and Detection of Errors.

Students shared that they follow the verification practices that include checking AI materials within the Google, academic literature, and instructor or peer consultation. The practices are in line with previous research that learners use coping mechanisms to reduce risks of AI misinformation although these are informally and inconsistently applied (Wang, 2024; Zou et al., 2023). The current results, however, make this literature more complicated because they show that verification is often selective and doubt-, instead of systematic, based. Students are more likely to check when something is suspicious, whereas AI misinformation can appear convincing enough to

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 4, Issue 1 (2026)

Online ISSN 3007-3197
Print ISSN 3007-3189
http://amresearchreview.com/index.php/Journal/about

not arouse doubts at all.

In terms of digital literacy, the given tendencies point towards the fact that the assessment abilities are partially formed: the students realize the need to use the verification, but do not have a systematic system to implement (Ng et al., 2023).

This is a relevant dilemma especially in the higher education system in Pakistan where the culture of exam based assessment is strong and facilitates the learning of right answers rather than learner assessment abilities. Institutional policies, therefore, should support the development of AI -specific information-literacy-practices, including the compulsory triangulation of sources and credibility scoring, and strictly preset referencing guidelines, to be used in case of the use of AI technologies in academic work.

Conclusion

The current study enhances the available literature by providing context-specific findings, based on Karachi, thus filling a significant gap in Pakistani scholarship on higher education that regards the lived experiences of students in terms of AI misinformation, a subject that is not being examined adequately. Although the effects of AI overtrust and misinformation are outlined in the international studies, the paper is unique to explain the process by which misinformation is internalized: students accept misconceived information not only due to the fact that AI sources are defective, but because they are accepted as authoritative and become replacements of academic information.

Future Implications

Implications (Students, teachers, Universities)

The study is, therefore, consequential to students since it highlights the fact that AI can be used as an academic support, but at the same time influences the learning experience by introducing misconceptions in the sublime. Students thus need specific instructions regarding the procedural application of AI, as well as regarding the systematic assessment of AI-generated material. To be more exact, it is possible to recommend learners to view AI output as a rough draft and cross-verify claims with reliable sources, as well as checking citations and factual soundness before further integration.

The findings should be paid attention to by teachers since the implementation of AI alters the student interaction with the learning activities. Since it is common that AI output can be readily accepted by students as authentic because of the apparent linguistic quality, the instructors can integrate verification exercises in the classroom (e.g., short-term AI verification activities where students will compare AI claims with academic works). This way redefines AI as an expedient rather than a shortcut of thinking.

Universities can pay attention to these implications since AI misinformation is not only an integrity problem but a learning problem that impairs cognitive growth and critical thinking. Uncritical reliance on AI can be reduced by institutional guidelines, AI literacy training, and redesign of assessment mechanisms. Otherwise, in the absence of such interventions, AI can support superficial learning, especially in exam-driven systems that reward the correct answer more than the process itself.

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 4, Issue 1 (2026)

Online ISSN 3007-3197
Print ISSN 3007-3189

http://amresearchreview.com/index.php/Journal/about

Policy Implications

At the policy level, universities would be able to:

* set rules of AI use including disclosure policies and ethical limits;
* organize AI-literacy seminars, the main activity of which is the detection of misinformation;
* require authentication procedures of assignments, including, but not limited to, AI output having to be backed up by no less than a couple of scholarly sources;
* educate faculty to come up with examinations that test reasoning power and not just confirming terminal answers.

Future Research

This study can be further developed in a number of ways by future research. To begin with, cross-city and cross-institutional comparative studies in Pakistan could be conducted to investigate whether trust patterns among students and verification vary over institutional culture or academic major. Second, the mixed-method study can combine qualitative interviews and experimental tasks that will help to measure the rate of misconceptions adoption against persuasive AI output. Third, intervention studies are justified, e.g. to determine whether structured AI-literacy training can increase the verification practices of students and reduce the development of misconceptions. Last but not least, the future studies must investigate the cognitive impact of persistent AI dependence on a long-term basis to determine whether long-term dependence on AI impacts children and their ability to think critically in the long-term, particularly in assessment-based academic settings.

References

Akakpo, B. (2023). Enhancing AI literacy through student workshops in higher education. Journal of Digital Education, 17(2), 112–125.

Baskara, F. X. R. (2025). Conceptualizing digital literacy for the AI era: A framework for preparing students in an AI-driven world. Data and Metadata, 4, 530. https://doi.org/10.56294/dm2025530

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? In Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (FAccT '21) (pp. 610–623). Association for Computing Machinery. https://doi.org/10.1145/3442188.3445922

Braun, V., & Clarke, V. (2021). Thematic analysis: A practical guide. SAGE Publications.

Chan, C. K. Y. (2024). Students' perceptions of "AI-giarism": Investigating changes in academic integrity in higher education. Education and Information Technologies, 29, 13151–13178. https://doi.org/10.1007/s10639-024-13151-7

Chaudhuri, S., & Terrones, L. (2024). Building AI information literacy: Lessons from three U.S. university libraries. Journal of Academic Librarianship, 50(2), 245–257.

Chen, C., Lin, M., & Zhao, L. (2024). AI and student motivation: A longitudinal study. British Journal of Educational Technology, 55(1), 33–48.

Choi, H., Park, J., & Kim, S. (2024). Overtrust in AI: Risks for student reasoning and

Online ISSN 3007-3197
Print ISSN 3007-3189

digital responsibility. Educational Technology Research & Development, 72(1), 22–39.

Chomintra, H. (2023). Digital literacy through critical pedagogy: A Freirean approach in AI contexts. Asian Journal of Educational Technology, 15(3), 144–160.

Creswell, J. W., & Poth, C. N. (2023). Qualitative inquiry and research design: Choosing among five approaches (5th ed.). SAGE Publications.

Dipto, P., Sharma, R., & Adhikari, M. (2024). AI-generated content and its impact on academic reasoning. Journal of Digital Scholarship, 6(1), 77–90.

Du, T. T. (2025). Students' perceptions of artificial intelligence in academic writing: A case at a university in Vietnam. Eurasian Journal of Applied Linguistics, 11(2), 122–134. https://doi.org/10.32601/ejal.11210

Essien, R., Narayanan, K., & Fang, W. (2024). Cognitive shifts in AI-supported learning environments. Journal of Educational Psychology, 119(3), 225–239.

Facione, P. A. (2023). Critical thinking: What it is and why it counts (2023 ed.). Insight Assessment.

Fasahat, M., Noor, A., & Alam, S. (2025). University students' perceptions of AI-generated content in learning. Pakistan Journal of Education and Technology, 9(1), 13–27.

Geoffrey, A., Singh, A., & Malik, A. (2024). Addressing AI misinformation in Asia's higher education. Asian Journal of Digital Pedagogy, 11(4), 234–248.

Gonsalves, J. (2024). Rethinking Bloom in AI-supported classrooms. International Journal of Cognitive Education, 10(1), 19–32.

He Li, Y., Tran, K., & Wang, Z. (2025). AI and learning autonomy: An empirical review. Educational Review Quarterly, 41(2), 178–194.

Hou, Y., Zhang, X., & Liu, Y. (2025). The role of AI literacy in students' critical AI engagement. Technology and Education Review, 18(2), 75–93.

Irshad, T., Ahmed, A., & Batool, S. (2025). AI dependency and academic ethics: A Pakistani perspective. Journal of Educational Ethics, 6(2), 66–80.

Kanwal, S., Farooq, U., & Hussain, M. (2023). Digital literacy and misinformation challenges in Pakistan. South Asian Journal of Educational Development, 8(1), 44–59.

Lee, S., Nakamura, Y., & O'Brien, T. (2025). AI and student cognition: A cross-cultural study. Computers & Education, 175, 104345.

Mohammed, H., Al-Bayati, R., & Salman, D. (2024). Iraqi university students' engagement with ChatGPT. Middle East Journal of Digital Learning, 7(1), 88–105.

Mosae, R., & Kaushal, P. (2025). Evaluating the credibility of AI-generated academic content: A case study. Indian Journal of Digital Education, 13(2), 55–70.

Moustaghfir, K., & Birgui, M. (2024). Strategic integration of AI literacy in higher education policy. Journal of Innovation in Education, 19(1), 11–28.

Muhammad Aqeel, A., Jamil, S., & Naveed, H. (2025). Critical thinking decline in AI-reliant classrooms. Journal of Emerging Educational Research, 12(1), 91–108.

Ng, W., Chai, C. S., & Lee, M. (2023). Conceptualizing digital literacy in the AI era. Educational Technology & Society, 26(2), 7–19.

Ponce Rojo, A., Morales, J., & Rivera, C. (2025). MIL-AI competencies: A

Online ISSN 3007-3197
Print ISSN 3007-3189

lexicometric analysis. Latin American Journal of Digital Education, 10(1), 21–40.

Rehman Uddin, M., Sadiq, T., & Haider, A. (2025). Misinformation and academic integrity in Pakistani universities. Journal of South Asian Higher Education, 3(2), 129–143.

Ruiz-Rojas, J., López-Gómez, M., & Torres, C. (2024). AI and academic task organization. Iberian Journal of Learning Sciences, 18(3), 201–218.

Sardi, M., Di Martino, E., & Santoro, G. (2025). Personalized AI in education: Opportunities and risks. European Journal of Digital Pedagogy, 14(1), 37–52.

Saúde, A., Lima, R., & Jorge, A. (2024). Self-regulated learning and AI: A review. Portuguese Journal of Educational Technology, 9(2), 55–70.

Schmitt, J., & Elechais, R. (2023). Digital deception and AI: Rebuilding student trust. Journal of Cyber Ethics in Education, 6(2), 23–39.

Sedlbauer, J., Novak, P., & Helcelet, M. (2024). Cognitive risks of ChatGPT in academic settings. Czech Journal of Educational Psychology, 10(1), 66–84.

Stuchlíková, I., Holeček, M., & Konečný, M. (2024). Generative AI and student cognition. Central European Journal of Digital Education, 15(1), 15–30.

Tiernan, P., Costello, E., & Brennan, M. (2023). AI and the future of digital literacy: Are students ready? Irish Educational Studies, 42(2), 245–263.

Wang, J. (2024). ChatGPT in writing classrooms: A qualitative study. Journal of Writing and Technology in Higher Education, 10(1), 51–69.

Washington, A. (2023). Media literacy in the age of generative AI. American Journal of Media Education, 8(3), 111–122.

Zaman, M., Chandio, A. A., & Noor, H. (2025). Evaluating the influence of Meta-AI on enhancing English reading comprehension proficiency: An experimental study within a social media application framework. Social Science Review Archives, 3(1), 2223–2233. https://doi.org/10.70670/sra.v3i1.533

Zhai, X., Yin, H., & Xu, X. (2024). Cognitive erosion in AI-enhanced classrooms. AI & Education Journal, 7(2), 99–117.

Zou, L., Zhang, R., & Cheng, M. (2023). Students' perceptions of AI misinformation: A Chinese university study. Asia-Pacific Journal of Education Technology, 11(4), 142–160.