E-ISSN: 3032-6753
Vol. 3, No. 3, September-December (2026), pp.374-388

STUDENTS' PERCEPTIONS OF THE CREDIBILITY OF INFORMATION GENERATED BY AI

Nasaruddin Siregar[1]
Universitas Bhayangkara Jakarta Raya, Indonesia

Sari Endah Nursyamsi[2]
Universitas Bhayangkara Jakarta Raya, Indonesia

Correspondence : Sari Endah Nursyamsi (sari.endah@dsn.ubharajaya.ac.id)

ABSTRACT

Generative artificial intelligence (GenAI) tools such as ChatGPT have quickly become part of students' everyday information practices, yet these tools can produce fluent text that is inaccurate or entirely fabricated. This study examines how university students perceive the credibility of AI-generated information, which factors shape those perceptions, and how students verify what they receive. An integrative literature review was conducted on empirical studies published between 2021 and 2026. Fourteen studies met the inclusion criteria, covering about 14,000 students in at least ten countries. Data were extracted into a structured matrix and analysed through thematic synthesis. Five themes emerged: (1) students hold cautious and conditional trust, with trust levels ranging from 14% to 47% across studies and credibility ratings slightly above the scale midpoint; (2) credibility is judged through surface cues and heuristics such as fluency, tone, and consistency; (3) credibility is task-dependent, high for summarising and explaining but low for references and high-stakes work; (4) experience, domain knowledge, and academic level shape credibility judgements; and (5) verification practices are uneven, relying heavily on intuition and re-prompting. Evidence that 47% to 55% of references generated by earlier ChatGPT versions were fabricated shows that misplaced confidence carries real academic risk. The findings suggest that higher education institutions should move from banning or blindly accepting AI toward building calibrated trust through AI literacy, explicit verification routines, and clear policies.

Keywords: Artificial Intelligence, ChatGPT, Information Credibility, Trust, Higher Education Students, AI Literacy

INTRODUCTION

Since the public release of ChatGPT in late 2022, generative artificial intelligence (GenAI) has spread through higher education at a pace rarely seen with earlier technologies. A nationwide survey of more than 6,300 students in Germany found that almost two-thirds already used AI-based tools in their studies, most often to clarify concepts and obtain explanations of subject-specific content (von Garrel & Mayer, 2023). In the United Kingdom, only 7% of 2,555 surveyed students had never heard of any GenAI tool, and more than half had used or considered using such tools for academic purposes (Johnston et al., 2024). Similar adoption patterns have been reported in Sweden, Poland, Hong Kong, and Indonesia (Amin et al., 2026; Chan & Hu, 2023; Stöhr et al., 2024; Strzelecki, 2024).

For many students, GenAI is no longer only a writing assistant. It has become a source of information that sits alongside, and sometimes replaces, search engines, textbooks, and library databases. This shift raises a central question of information credibility. Large language models generate text by predicting likely word sequences rather than by retrieving verified facts, which

E-ISSN: 3032-6753
Vol. 3, No. 3, September-December (2026), pp.374-388

makes them prone to hallucination, that is, the production of fluent and confident content that is nonsensical or unfaithful to the source (Ji et al., 2023). The problem is not hypothetical. When ChatGPT was asked to write short literature reviews, 55% of the citations produced by GPT-3.5 and 18% of those produced by GPT-4 were fabricated (Walters & Wilder, 2023). In medical content, only 7% of 115 generated references were both authentic and accurate (Bhattacharyya et al., 2023).

Because AI-generated text is often clear, well-organised, and confident, users may find it as credible as content written by humans. An experimental study found no difference in perceived competence or trustworthiness between human and AI-generated content, while AI-generated text was rated as clearer and more engaging (Huschens et al., 2023). This combination of high fluency and uncertain accuracy creates a risk of over-trust, especially for students who are still developing disciplinary knowledge and information literacy (Kasneci et al., 2023).

Research on students and GenAI has grown rapidly, but most early studies focused on adoption, usage frequency, and academic integrity rather than on how students judge the credibility of the information they receive (Strzelecki, 2024; von Garrel & Mayer, 2023). Studies that do address credibility are scattered across disciplines such as education, information science, computer science, and health, use different measures, and have not yet been brought together. A synthesis is therefore needed to understand what is currently known about students' credibility perceptions and to inform teaching and institutional policy.

This study addresses that gap through an integrative review of empirical studies published between 2021 and 2026. It is guided by three research questions:

- RQ1: How do university students perceive the credibility of AI-generated information?
- RQ2: What factors shape students' credibility judgements of AI-generated information?
- RQ3: How do students verify AI-generated information, and what gaps exist in their verification practices?

LITERATURE REVIEW

Information credibility

Information credibility refers to the extent to which a source or a message is perceived as believable. It is usually understood as a perception held by the receiver rather than as an objective property of the information. Classic credibility research identifies two core dimensions: trustworthiness, meaning the perceived honesty and good intention of the source, and expertise, meaning the perceived competence of the source. In digital environments, people

E-ISSN: 3032-6753
Vol. 3, No. 3, September-December (2026), pp.374-388

rarely evaluate every item systematically. Instead, they rely on heuristics such as the appearance of a website, the consistency of information across sources, or endorsements by others (Choi et al., 2024).

GenAI complicates these established practices. A chatbot answer usually has no visible author, publisher, or date, and it often presents information without any traceable source. Choi et al. (2024) found that students therefore developed new constructs to judge GenAI credibility, such as the refinedness and explainability of an answer, and new cues such as the version of the model and the tone of the response. Credibility judgement in the GenAI era thus combines traditional dimensions with new, AI-specific signals.

Generative AI and the problem of hallucination

Large language models are trained on very large text corpora to predict the next token in a sequence. This design makes them highly fluent but does not guarantee factual accuracy. Ji et al. (2023) describe hallucination as one of the main limitations of natural language generation, distinguishing between intrinsic hallucinations, which contradict the source, and extrinsic hallucinations, which cannot be verified from the source. In academic settings, fabricated references are the most visible form of hallucination. Walters and Wilder (2023) examined 636 citations generated for 84 short literature reviews and found that, besides fabrication, 43% of the real citations produced by GPT-3.5 and 24% of those produced by GPT-4 contained substantive errors. Newer models have reduced but not eliminated these problems, which means that verification remains essential.

Trust in AI and technology acceptance

Trust is closely related to credibility but focuses on the user's willingness to rely on a system under uncertainty. In technology acceptance research, trust has emerged as a key factor in GenAI adoption. Among 607 ChatGPT users, trust had a strong effect on intention to use ($\beta = 0.711$), which in turn shaped actual use (Choudhury & Shamszare, 2023). Among students in China, perceived trust moderated the relationships between ChatGPT awareness and perceived ease of use, usefulness, and intelligence (Shahzad et al., 2024). In Poland, habit, performance expectancy, and hedonic motivation explained 73.4% of the variance in students' intention to use ChatGPT (Strzelecki, 2024). Students' perception of information quality, including its reliability and relevance, also increases their readiness to adopt GenAI (Essien et al., 2025). These studies show that credibility perceptions are not only a cognitive judgement but also a driver of behaviour.

Trust in AI is desirable only when it is calibrated, that is, when the level of trust matches the actual reliability of the system. Over-trust can lead to automation bias, where users accept

E-ISSN: 3032-6753
Vol. 3, No. 3, September-December (2026), pp.374-388

AI output without sufficient scrutiny, while under-trust can cause users to reject useful assistance. For students, the risk of over-trust is especially serious because errors may enter assignments, theses, and eventually professional practice.

AI literacy

AI literacy has been proposed as the competence that allows people to use AI critically and responsibly. Ng et al. (2021) conceptualise AI literacy in four dimensions: knowing and understanding AI, using and applying AI, evaluating and creating AI, and AI ethics. The evaluation dimension is directly related to credibility, because it includes the ability to recognise when AI output may be wrong and to check it against other sources. Kasneci et al. (2023) argue that the educational value of large language models depends on whether students learn to treat them as starting points rather than final authorities.

Research gap

Existing reviews and commentaries on GenAI in education mostly focus on opportunities, risks, and integrity (Kasneci et al., 2023). Empirical studies on students' credibility perceptions do exist, but they differ in country, discipline, method, and measures, and their findings have not been systematically compared. In addition, few studies connect students' perceptions with evidence on the actual accuracy of AI output. This review fills that gap by synthesising recent empirical evidence and placing it next to evidence on AI hallucination.

METHOD

Research design

This study used an integrative literature review design. Integrative reviews allow researchers to combine studies with different methodologies, including quantitative, qualitative, and mixed-methods research, in order to build a comprehensive understanding of a phenomenon (Cronin & George, 2023). This design was chosen because research on students' perceptions of AI-generated information is new and methodologically diverse. Reporting of the search and selection process was informed by the PRISMA 2020 statement (Page et al., 2021).

Search strategy

Literature was searched in journals indexed by Scopus, the Directory of Open Access Journals (DOAJ), Google Scholar, and arXiv for preprints that had not yet been formally published. The search combined keywords related to the technology ("generative AI", "ChatGPT", "AI chatbot", "large language model"), the population ("student", "university", "higher education"), and the outcome ("credibility", "trust", "accuracy", "reliability",

E-ISSN: 3032-6753
Vol. 3, No. 3, September-December (2026), pp.374-388

"perception", "hallucination"). Reference lists of relevant articles were also checked to find additional studies.

Inclusion and exclusion criteria

Table 1. Inclusion and exclusion criteria

Criterion	Inclusion	Exclusion
Publication year	2021 to 2026	Published before 2021
Population	University students (undergraduate or postgraduate)	School pupils, general public, or professionals only
Focus	Perceptions of credibility, trust, accuracy, or reliability of AI-generated information	Studies focusing only on usage frequency or academic integrity policy
Study type	Empirical studies (quantitative, qualitative, or mixed methods)	Editorials, opinion papers, and non-empirical commentaries
Language	English	Other languages

Source: Developed by the authors.

Studies that measured the actual accuracy of AI output without involving students (Bhattacharyya et al., 2023; Walters & Wilder, 2023) and studies on adoption without a credibility focus (Strzelecki, 2024; von Garrel & Mayer, 2023) were not included in the synthesis but were used as contextual evidence in the discussion.

Study selection, data extraction, and analysis

Records were screened by title and abstract and then assessed in full text against the criteria in Table 1. Fourteen studies were included in the final synthesis. The review process is summarised in Figure 1.

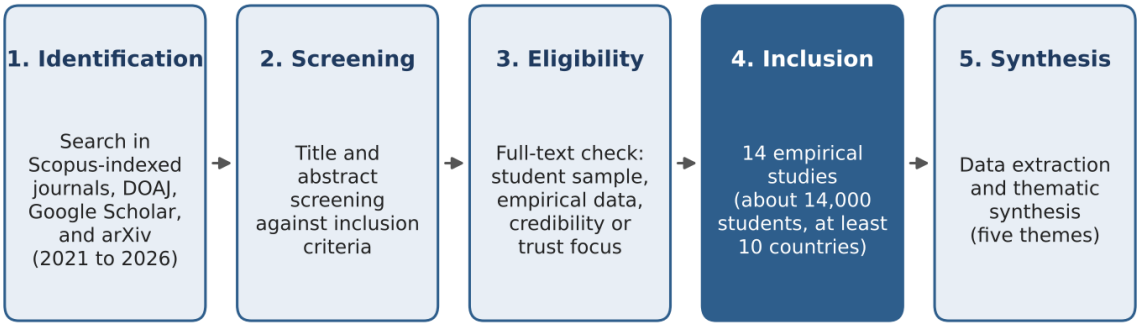

Figure 1. Stages of the integrative review process

Source: Developed by the authors based on Page et al. (2021).

For each included study, the following data were extracted: author and year, country, sample size and characteristics, research design, and key findings related to credibility, trust, or verification. The extracted findings were then analysed through thematic synthesis following the reflexive thematic analysis approach of Braun and Clarke (2021). Findings were first coded

E-ISSN: 3032-6753
Vol. 3, No. 3, September-December (2026), pp.374-388

line by line, then grouped into descriptive themes, and finally developed into analytical themes that answer the research questions. Because the studies used different instruments and scales, the numerical results were compared descriptively rather than pooled statistically.

RESULTS AND DISCUSSION

Characteristics of the included studies

The fourteen included studies involved a combined total of about 14,043 students from at least ten countries: Indonesia, the United States, India, the United Kingdom, Hong Kong, China, Vietnam, the United Arab Emirates, Spain, and Sweden. Three studies were published in 2023, seven in 2024, two in 2025, and two in 2026. Six studies used quantitative surveys or structural equation modelling, six used mixed methods, and two were qualitative. Sample sizes ranged from 25 interviewees (Choi et al., 2024) to 5,894 survey respondents (Stöhr et al., 2024). The characteristics and main findings of each study are presented in Table 2.

Table 2. Characteristics and key findings of the included studies

Author (year)	Country	Sample	Design	Key finding on credibility or trust
Amin et al. (2026)	Indonesia	56 students (UG to doctoral)	Survey	ChatGPT seen as easy and helpful; concerns about integrity and over-reliance; postgraduates more cautious than undergraduates
Amoozadeh et al. (2024)	US, India	253 students	Survey	47% trust, 36% neutral, 16% distrust; 103 of 216 rejected believing output when its correctness was uncertain
Baek et al. (2024)	US	1,001 students	Mixed methods	Skepticism theme in which ChatGPT "seems too good to be true"
Bikanga Ada (2024)	UK	335 computing students	Mixed methods	57.4% do not trust ChatGPT information, 14% trust it; postgraduates trust more (M = 3.50 vs 2.97); reports of fake literature
Bouyzourn & Birch (2025)	UK	115 students and 4 interviews	Mixed methods	Perceived expertise and ethical risk are strongest trust predictors; trust lowest for citation generation
Chan & Hu (2023)	Hong Kong	399 students	Survey	Generally positive attitude; accuracy is the first concern listed, followed by privacy and ethics
Choi et al. (2024)	US	25 students	Interviews	Students used 9 credibility constructs, 5 heuristics, and 10 cues to judge GenAI output
Essien et al. (2025)	Not reported	522 students	SEM	Perceived information quality (quality, reliability, relevance) increases readiness to adopt GenAI
Finkel et al. (2024)	Spain	4,800 students	Survey and expert analysis	Mean ratings (0 to 10): correct 6.01, reliable 5.75, unbiased 5.71; ratings depend strongly on experience
Ngo (2023)	Vietnam	200 survey and 30 interviews	Mixed methods	Most serious concern: inability to assess the quality and reliability of sources; difficulty citing accurately
Shahzad et al. (2024)	China	320 students	PLS-SEM	Trust moderates the effect of awareness on ease of use, usefulness, and perceived intelligence
Shoufan (2023)	UAE	56 engineering students	Thematic analysis and survey	Answers "not always accurate"; solid background knowledge needed to use ChatGPT well
Shoufan & Esmaeil (2026)	UAE	63 engineering students	Qualitative	Fabricated citations and confident wrong answers; about 46% of comments described active verification
Stöhr et al. (2024)	Sweden	5,894 students	Survey	Over half positive, nearly as many concerned; female, humanities, and medicine students more concerned

E-ISSN: 3032-6753
Vol. 3, No. 3, September-December (2026), pp.374-388

Source: Compiled by the authors from the included studies. UG = undergraduate; SEM = structural equation modelling; PLS = partial least squares.

Thematic synthesis produced five themes. Table 3 shows the themes and the studies that support each of them.

Table 3. Themes from the thematic synthesis

Theme	Description	Supporting studies
T1. Cautious and conditional trust	Students value GenAI but do not fully trust its information; credibility ratings are moderate	Amoozadeh et al.; Baek et al.; Bikanga Ada; Chan & Hu; Finkel et al.; Stöhr et al.
T2. Credibility judged through cues and heuristics	Fluency, tone, consistency, and model version are used as shortcuts	Choi et al.; Bouyzourn & Birch; Shoufan & Esmaeil
T3. Task-dependent credibility	Trust is higher for summarising and explaining, lower for references and high-stakes work	Bikanga Ada; Bouyzourn & Birch; Finkel et al.
T4. Experience, knowledge, and academic level matter	Experience raises ratings; knowledge and postgraduate status shape caution	Amin et al.; Bikanga Ada; Finkel et al.; Shoufan; Stöhr et al.
T5. Uneven verification, weak citation checking	Verification relies on intuition and re-prompting; source checking is the weakest area	Ngo; Shoufan & Esmaeil; Bouyzourn & Birch; Bikanga Ada

Source: Thematic synthesis by the authors.

Theme 1: Cautious and conditional trust

Across studies, students showed a positive attitude toward GenAI as a learning aid but held clear reservations about the credibility of its information. Chan and Hu (2023) reported a generally positive attitude among 399 students in Hong Kong, yet accuracy was the first concern listed. In Sweden, more than half of 5,894 students expressed positive attitudes, but nearly as many voiced concerns about the future use of chatbots in education (Stöhr et al., 2024).

Direct measures of trust varied considerably between populations (Figure 2). In a sample of 253 students in the United States and India, 47% reported trusting GenAI, 36% were neutral, and only 16% expressed distrust (Amoozadeh et al., 2024). Among 335 computing students in the United Kingdom, the pattern was almost reversed: 57.4% did not trust the information provided by ChatGPT, 27.8% were undecided, and only 14% trusted it (Bikanga Ada, 2024).

E-ISSN: 3032-6753
Vol. 3, No. 3, September-December (2026), pp.374-388

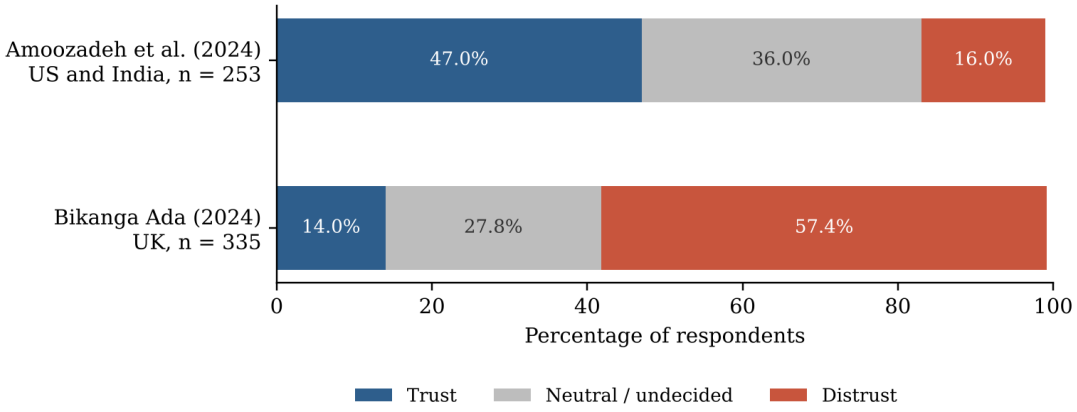

Figure 2. Distribution of students' trust in AI-generated information in two studies
Source: Visualised by the authors from Amoozadeh et al. (2024) and Bikanga Ada (2024).

Even where trust was higher, it was not blind. In the study by Amoozadeh et al. (2024), 103 of 216 respondents disagreed with the statement that they believe the system's output even when they do not know for certain that it is correct, while only 45 agreed. The largest study on rated credibility, with 4,800 students in Spain, found that students rated ChatGPT's information only slightly above the midpoint of a 0 to 10 scale. Completeness (6.62) and currency (6.44) received the highest ratings, while correctness (6.01), reliability (5.75), and lack of bias (5.71) received the lowest (Finkel et al., 2024), as shown in Figure 3. A qualitative strand in a study of 1,001 American students also revealed a skeptical view that ChatGPT "seems too good to be true" (Baek et al., 2024).

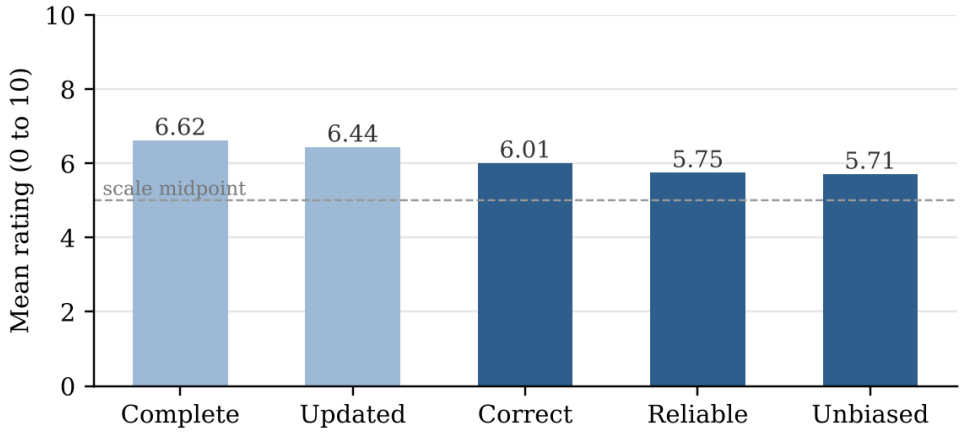

Figure 3. Students' mean ratings of ChatGPT information quality (n = 4,800)
Source: Visualised by the authors from Finkel et al. (2024).

Taken together, these findings answer RQ1: students perceive AI-generated information as moderately credible, useful enough to rely on for many tasks, but not reliable enough to accept without question. The large difference between the two trust distributions in Figure 2

E-ISSN: 3032-6753
Vol. 3, No. 3, September-December (2026), pp.374-388

also suggests that trust is highly context-dependent, influenced by discipline, institutional culture, and the specific tasks students perform.

Theme 2: Credibility judged through cues and heuristics

Because AI answers lack the traditional markers of credibility, students developed new ways of judging them. In interviews with 25 American students, Choi et al. (2024) identified nine credibility constructs, including refinedness and explainability, five heuristics, including consistency within and between AI systems, and ten cues, including the model version and the tone of the response. Students in the United Arab Emirates reported that AI "can sound very convincing even when the information is wrong" and often flagged answers only when they seemed illogical or irrelevant (Shoufan & Esmaeil, 2026).

Reliance on such cues is understandable, but it is also risky. The very features that make AI text persuasive, namely clarity, structure, and confidence, are not reliable indicators of accuracy. Huschens et al. (2023) showed that readers perceived AI-generated content as being as competent and trustworthy as human-written content and rated it as clearer. Bouyzourn and Birch (2025) found that perceived expertise was one of the strongest predictors of students' overall trust, while human-likeness and reputation were not significant. When perceived expertise is inferred from fluent writing, students may attribute expertise to the system that it does not actually have.

Theme 3: Task-dependent credibility

Students did not judge AI credibility uniformly. They distinguished between tasks. Computing students in the United Kingdom trusted ChatGPT more for summarising, paraphrasing, and explaining concepts than for high-stakes academic work (Bikanga Ada, 2024). Similarly, Bouyzourn and Birch (2025) found that trust was highest for coding and summarising and lowest for entertainment and citation generation. This pattern indicates that students possess some intuitive sense of where AI is strong and where it is weak.

However, the same study revealed an important contradiction. Confidence in ChatGPT's referencing ability was the strongest correlate of overall trust, even though referencing is the task with the best-documented inaccuracies (Bouyzourn & Birch, 2025). As Figure 4 shows, objective evaluations found that 55% of references produced by GPT-3.5 and 18% of those produced by GPT-4 were fabricated (Walters & Wilder, 2023), and that among 115 references generated for medical papers, 47% were fabricated and 46% were real but inaccurate (Bhattacharyya et al., 2023). The gap between students' confidence and the documented accuracy of AI references is therefore one of the most important findings of this review.

E-ISSN: 3032-6753
Vol. 3, No. 3, September-December (2026), pp.374-388

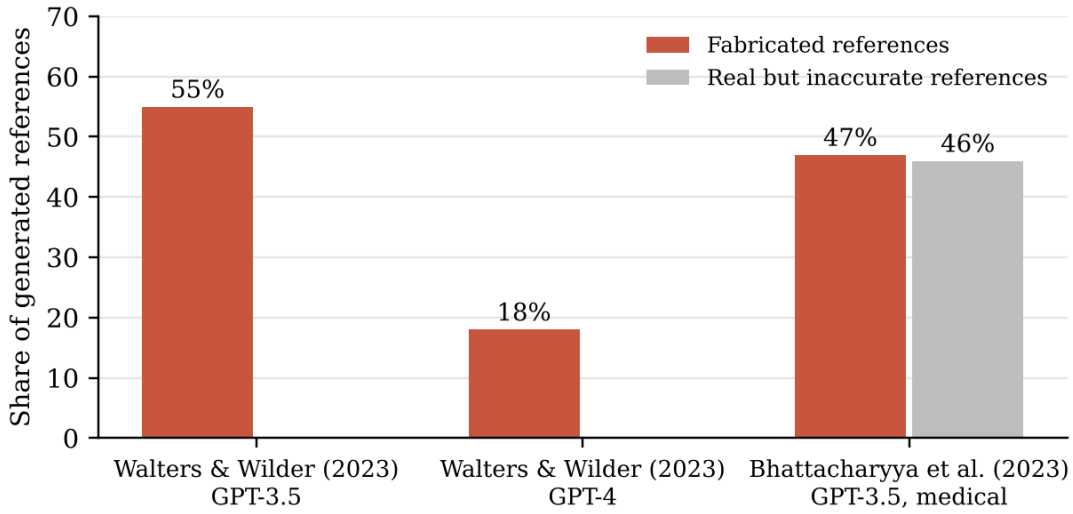

Figure 4. Objective evidence on the accuracy of references generated by ChatGPT

Source: Visualised by the authors from Walters and Wilder (2023) and Bhattacharyya et al. (2023).

Theme 4: Experience, knowledge, and academic level

Several individual factors shaped students' credibility judgements, answering RQ2. The first factor is experience. In the Spanish study, students with a good experience rated the reliability of ChatGPT at 6.26 out of 10, whereas those with a bad experience rated it at only 2.40 (Finkel et al., 2024). Bouyzourn and Birch (2025) likewise found that frequent use increased trust, while self-reported understanding of how large language models work reduced it. Experience can therefore either build justified confidence or reinforce over-reliance, depending on whether students have encountered and recognised errors.

The second factor is domain knowledge. Engineering students in the United Arab Emirates believed that using ChatGPT effectively requires solid background knowledge, because only knowledgeable users can detect wrong answers (Shoufan, 2023). The third factor is academic level, although the direction of its effect differs across studies. Postgraduate computing students in the United Kingdom reported significantly more trust than undergraduates (M = 3.50 vs 2.97) (Bikanga Ada, 2024), whereas in Indonesia, master's and doctoral students were more cautious about ethical and pedagogical implications than undergraduates (Amin et al., 2026). The fourth factor is discipline and gender. Female students and students in the humanities and medicine were more negative and more concerned about AI in learning and assessment, while male students and students in technology and engineering were more optimistic (Stöhr et al., 2024).

E-ISSN: 3032-6753
Vol. 3, No. 3, September-December (2026), pp.374-388

Table 4. Factors shaping students' credibility judgements of AI-generated information

Factor	Direction of influence	Evidence
Experience and frequency of use	Good experience and frequent use increase perceived credibility	Finkel et al. (2024); Bouyzourn & Birch (2025)
Understanding of AI and domain knowledge	Greater understanding reduces trust and improves error detection	Bouyzourn & Birch (2025); Shoufan (2023)
Academic level	Mixed: higher trust among UK postgraduates, more caution among Indonesian postgraduates	Bikanga Ada (2024); Amin et al. (2026)
Discipline and gender	Humanities, medicine, and female students more concerned	Stöhr et al. (2024)
Task type	Higher trust for summarising and explaining, lower for references	Bikanga Ada (2024); Bouyzourn & Birch (2025)
Output cues	Fluency, tone, consistency, and version used as shortcuts	Choi et al. (2024)

Source: Synthesised by the authors.

Theme 5: Uneven verification and weak citation checking

Regarding RQ3, most studies showed that students recognise the need to verify AI output, but their verification practices are uneven. Vietnamese students identified the inability to assess the quality and reliability of sources, together with difficulty in citing sources accurately, as their most serious concerns, and they proposed checking answers against reliable sources as a solution (Ngo, 2023). Among senior engineering students in the United Arab Emirates, about 46% of comments described active verification; of these, roughly 70% involved cross-checking with external sources, while about 30% involved re-prompting the same model (Shoufan & Esmaeil, 2026). Many other students relied on intuition, which the authors considered a weak safeguard because convincingly wrong answers can pass a quick perceptual check.

Re-prompting the same system is a particularly limited strategy because it tests the model against itself rather than against independent evidence. Moreover, Shoufan and Esmaeil (2026) found that some students held a search-engine-like mental model of GenAI, assuming that the system retrieves information rather than generates it. Such misconceptions help explain why fabricated citations can be accepted. Students in the United Kingdom also reported that "ChatGPT sometimes makes up fake literature" (Bikanga Ada, 2024), which suggests that some students learn the limits of AI only after encountering errors themselves.

Integrated discussion

The five themes can be combined into a conceptual framework of students' credibility judgements of AI-generated information (Figure 5). Individual and contextual antecedents, including experience, knowledge, academic level, discipline, task type, and output cues, shape the perceived credibility of AI output. This perception then leads to one of three outcomes: calibrated trust, in which students use AI but verify and cite carefully; over-trust, in which fluent but incorrect information or fabricated citations are accepted; or under-trust, in which students

E-ISSN: 3032-6753
Vol. 3, No. 3, September-December (2026), pp.374-388

avoid AI and lose its potential learning benefits. Institutional guidance and AI literacy training can moderate this process by helping students align their trust with the actual reliability of the system.

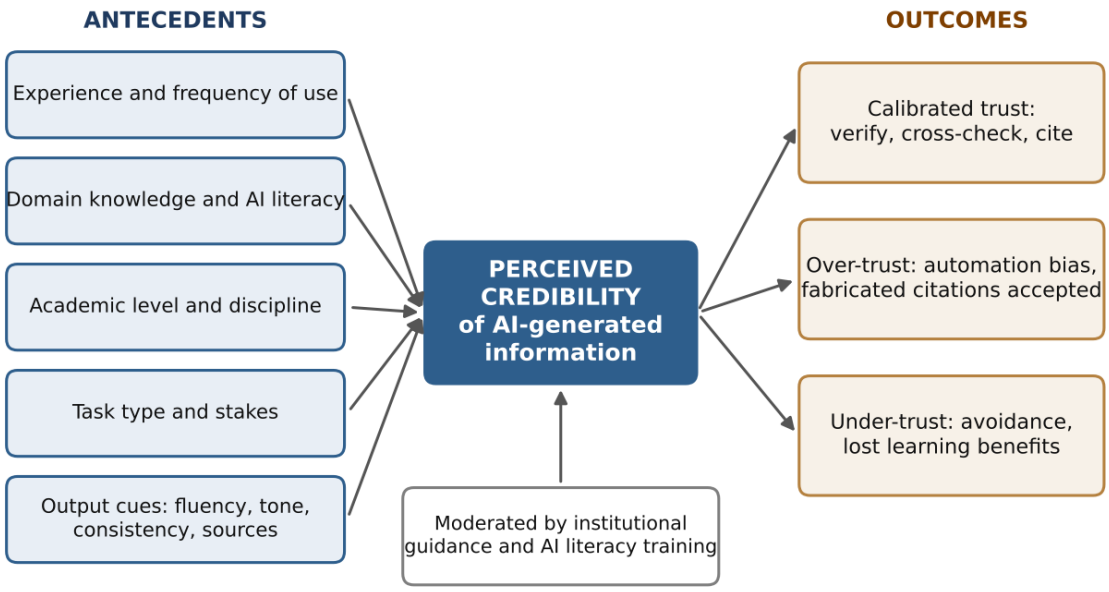

Figure 5. Conceptual framework of students' credibility judgements of AI-generated information

Source: Developed by the authors from the thematic synthesis.

This framework connects the reviewed findings with trust and acceptance research. Trust strongly predicts the intention to use GenAI (Choudhury & Shamszare, 2023; Shahzad et al., 2024), and perceived information quality increases readiness to adopt it (Essien et al., 2025). Therefore, credibility perceptions do not only affect what students believe but also how much they rely on AI in their studies. The central problem is not whether students trust AI, but whether their trust is calibrated. The review shows two forms of miscalibration: students may distrust AI broadly while still trusting it for one of its weakest tasks, namely referencing, and students may judge credibility from fluency, which AI produces easily regardless of accuracy.

These findings are consistent with the AI literacy framework of Ng et al. (2021), which places evaluation as a core competence. Most students in the reviewed studies had some awareness that AI can be wrong, but fewer had systematic strategies for evaluation. This gap between awareness and practice suggests that AI literacy programmes should focus not only on knowing that AI can hallucinate, but on concrete verification routines, such as checking every reference in a library database, comparing AI answers with course materials, and asking AI for

E-ISSN: 3032-6753
Vol. 3, No. 3, September-December (2026), pp.374-388

sources that can then be independently confirmed. Practical implications are summarised in Table 5.

Table 5. Practical implications for higher education

Stakeholder	Recommended action
Lecturers	Design assignments that require students to verify and annotate AI output; demonstrate examples of hallucination and fabricated references in class
Librarians	Integrate GenAI evaluation into information literacy programmes, including reference verification in library databases
Institutions	Provide clear policies on when and how AI may be used and how its use should be disclosed, rather than blanket bans
Curriculum developers	Embed AI literacy, especially the evaluation dimension, into first-year courses across disciplines
Students	Treat AI output as a starting point; verify facts with independent sources; never cite a reference without locating the original

Source: Developed by the authors.

This review has several limitations. First, the included studies used different instruments and scales, so their results could not be pooled statistically. Second, some studies were preprints that had not yet completed peer review at the time of writing (Bouyzourn & Birch, 2025; Shoufan & Esmaeil, 2026). Third, most studies relied on self-reported perceptions and did not measure whether students could actually detect inaccurate information. Fourth, GenAI models change rapidly, so perceptions measured in 2023 may not reflect experiences with newer models. Finally, only English-language publications were included, which may under-represent research from non-English-speaking countries, including Indonesia.

CONCLUSION

This integrative review of fourteen empirical studies involving about 14,000 students shows that university students perceive AI-generated information as moderately credible. They value its usefulness but hold cautious and conditional trust, with credibility ratings only slightly above the midpoint and trust levels varying widely between populations. Students judge credibility through new heuristics and cues, such as fluency, tone, and consistency, and their trust depends on the task, their experience, domain knowledge, academic level, and discipline. Verification practices are uneven and often rely on intuition or re-prompting, while the area where students' confidence is most misplaced, namely references, is the area where AI is objectively least reliable.

The main implication is that higher education should aim for calibrated trust rather than prohibition or uncritical acceptance. This requires AI literacy programmes that emphasise evaluation skills, assignments that make verification visible, and institutional policies that guide responsible use. Future research should use validated credibility scales, combine perception

E-ISSN: 3032-6753
Vol. 3, No. 3, September-December (2026), pp.374-388

measures with objective tests of students' ability to detect AI errors, conduct longitudinal and experimental studies on the effects of AI literacy interventions, and include more studies from developing countries such as Indonesia, where research on this topic remains limited.

REFERENCES

Amin, M., Mustaqim, B., Pratama, W., & Sibuea, A. M. (2026). Assessing student perspectives on ChatGPT in higher education: A quantitative analysis. *IAES International Journal of Artificial Intelligence (IJ-AI), 15*(2), 1062–1070. https://doi.org/10.11591/ijai.v15.i2.pp1062-1070

Amoozadeh, M., Daniels, D., Nam, D., Kumar, A., Chen, S., Hilton, M., Srinivasa Ragavan, S., & Alipour, M. A. (2024). Trust in generative AI among students: An exploratory study. In *Proceedings of the 55th ACM Technical Symposium on Computer Science Education (SIGCSE 2024)* (pp. 67–73). ACM. https://doi.org/10.1145/3626252.3630842

Baek, C., Tate, T., & Warschauer, M. (2024). "ChatGPT seems too good to be true": College students' use and perceptions of generative AI. *Computers and Education: Artificial Intelligence, 7*, 100294. https://doi.org/10.1016/j.caeai.2024.100294

Bhattacharyya, M., Miller, V. M., Bhattacharyya, D., & Miller, L. E. (2023). High rates of fabricated and inaccurate references in ChatGPT-generated medical content. *Cureus, 15*(5), e39238. https://doi.org/10.7759/cureus.39238

Bikanga Ada, M. (2024). It helps with crap lecturers and their low effort: Investigating computer science students' perceptions of using ChatGPT for learning. *Education Sciences, 14*(10), 1106. https://doi.org/10.3390/educsci14101106

Bouyzourn, K., & Birch, A. (2025). *What shapes user trust in ChatGPT? A mixed-methods study of user attributes, trust dimensions, task context, and societal perceptions among university students* (arXiv:2507.05046). arXiv. https://doi.org/10.48550/arXiv.2507.05046

Braun, V., & Clarke, V. (2021). One size fits all? What counts as quality practice in (reflexive) thematic analysis? *Qualitative Research in Psychology, 18*(3), 328–352. https://doi.org/10.1080/14780887.2020.1769238

Chan, C. K. Y., & Hu, W. (2023). Students' voices on generative AI: Perceptions, benefits, and challenges in higher education. *International Journal of Educational Technology in Higher Education, 20*, 43. https://doi.org/10.1186/s41239-023-00411-8

Choi, W., Bak, H., An, J., Zhang, Y., & Stvilia, B. (2024). College students' credibility assessments of GenAI-generated information for academic tasks: An interview study. *Journal of the Association for Information Science and Technology*. Advance online publication. https://doi.org/10.1002/asi.24978

Choudhury, A., & Shamszare, H. (2023). Investigating the impact of user trust on the adoption and use of ChatGPT: Survey analysis. *Journal of Medical Internet Research, 25*, e47184. https://doi.org/10.2196/47184

E-ISSN: 3032-6753
Vol. 3, No. 3, September-December (2026), pp.374-388

Cronin, M. A., & George, E. (2023). The why and how of the integrative review. *Organizational Research Methods, 26*(1), 168–192. https://doi.org/10.1177/1094428120935507

Essien, A. E., Crawford, J., & Kremantzis, M. D. (2025). Evaluating the impact of perceived information quality on GenAI technology adoption in higher education students. *Journal of Applied Learning & Teaching, 8*(2), 24–34. https://doi.org/10.37074/jalt.2025.8.2.5

Finkel, L., Parra-Contreras, P., Martínez-Solana, Y., & Matos-Mejías, C. (2024). ChatGPT as a source of information in higher education: How are the results provided by generative AI valued? *Profesional de la Información, 33*(6), e330604. https://doi.org/10.3145/epi.2024.0604

Huschens, M., Briesch, M., Sobania, D., & Rothlauf, F. (2023). *Do you trust ChatGPT? Perceived credibility of human and AI-generated content* (arXiv:2309.02524). arXiv. https://doi.org/10.48550/arXiv.2309.02524

Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., Ishii, E., Bang, Y. J., Madotto, A., & Fung, P. (2023). Survey of hallucination in natural language generation. *ACM Computing Surveys, 55*(12), 248. https://doi.org/10.1145/3571730

Johnston, H., Wells, R. F., Shanks, E. M., Boey, T., & Parsons, B. N. (2024). Student perspectives on the use of generative artificial intelligence technologies in higher education. *International Journal for Educational Integrity, 20*, 2. https://doi.org/10.1007/s40979-024-00149-4

Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., ... Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences, 103*, 102274. https://doi.org/10.1016/j.lindif.2023.102274

Ng, D. T. K., Leung, J. K. L., Chu, S. K. W., & Qiao, M. S. (2021). Conceptualizing AI literacy: An exploratory review. *Computers and Education: Artificial Intelligence, 2*, 100041. https://doi.org/10.1016/j.caeai.2021.100041

Ngo, T. T. A. (2023). The perception by university students of the use of ChatGPT in education. *International Journal of Emerging Technologies in Learning, 18*(17), 4–19. https://doi.org/10.3991/ijet.v18i17.39019

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., ... Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ, 372*, n71. https://doi.org/10.1136/bmj.n71

Shahzad, M. F., Xu, S., & Javed, I. (2024). ChatGPT awareness, acceptance, and adoption in higher education: The role of trust as a cornerstone. *International Journal of*

E-ISSN: 3032-6753
Vol. 3, No. 3, September-December (2026), pp.374-388

Educational Technology in Higher Education, 21. https://doi.org/10.1186/s41239-024-00478-x

Shoufan, A. (2023). Exploring students' perceptions of ChatGPT: Thematic analysis and follow-up survey. *IEEE Access, 11*, 38805–38818. https://doi.org/10.1109/ACCESS.2023.3268224

Shoufan, A., & Esmaeil, A.-A. (2026). *AI hallucination from students' perspective: A thematic analysis* (arXiv:2602.17671). arXiv. https://doi.org/10.48550/arXiv.2602.17671

Stöhr, C., Ou, W., & Malmström, H. (2024). Perceptions and usage of AI chatbots among students in higher education across genders, academic levels and fields of study. *Computers and Education: Artificial Intelligence, 7*, 100259. https://doi.org/10.1016/j.caeai.2024.100259

Strzelecki, A. (2024). To use or not to use ChatGPT in higher education? A study of students' acceptance and use of technology. *Interactive Learning Environments, 32*(9), 5142–5155. https://doi.org/10.1080/10494820.2023.2209881

von Garrel, J., & Mayer, J. (2023). Artificial intelligence in studies: Use of ChatGPT and AI-based tools among students in Germany. *Humanities and Social Sciences Communications, 10*, 799. https://doi.org/10.1057/s41599-023-02304-7

Walters, W. H., & Wilder, E. I. (2023). Fabrication and errors in the bibliographic citations generated by ChatGPT. *Scientific Reports, 13*, 14045. https://doi.org/10.1038/s41598-023-41032-5