Iranian Journal of Educational Sociology

Journal Website

Article history:
Received 08 June 2026
Revised 02 August 2026
Accepted 09 August 2026
Initial Published 11 August 2026
Final Publication 01 May 2027

Iranian Journal of Educational Sociology

Volume 10, Issue 3, pp 1-21

Quarterly
Iranian Journal of Educational Sociology

E-ISSN: 2645-3460

Explaining the Process of University Students' Epistemic Trust in AI-Generated Content: A Grounded Theory Study from the Perspective of Philosophy of Education

Sakineh. Rastegar Moghadam Ebrahimian[1*]

[1] Assistant Professor, Department of Theology, Farhangian University, Tehran, Iran

*** Corresponding author email address**: rastegar@cfu.ac.ir

A r t i c l e I n f o

Article type:
Original Research

How to cite this article:
Rastegar Moghadam Ebrahimian, S. (2027). Explaining the Process of University Students' Epistemic Trust in AI-Generated Content: A Grounded Theory Study from the Perspective of Philosophy of Education. *Iranian Journal of Educational Sociology, 10*(3), 1-21.
https://doi.org/10.61838/kman.ijes.1534

© 2027 the authors. Published by Iranian Association for Sociology of Education, Tehran, Iran. This is an open access article under the terms of the Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0) License.

A B S T R A C T

Purpose: This study aimed to explain the process through which university students develop, evaluate, recalibrate, and regulate epistemic trust in AI-generated content from the perspective of philosophy of education.

Methods and Materials: This qualitative study was conducted using a grounded theory approach among 24 university students studying at universities and higher education institutions in Tehran, Iran. Participants were initially selected through purposive sampling and subsequently through theoretical sampling until theoretical saturation was achieved. Data were collected through individual semi-structured, in-depth interviews focusing on participants' experiences of using, evaluating, trusting, questioning, and verifying AI-generated content in academic contexts. Interviews were transcribed verbatim and analyzed concurrently with data collection using initial coding, focused coding, axial categorization, constant comparison, memo writing, theoretical sampling, and selective integration. Trustworthiness was enhanced through prolonged engagement with the data, peer examination, participant validation, negative-case analysis, reflexive memo writing, and maintenance of an audit trail.

Findings: The analysis generated six main categories: pragmatic entry into AI-mediated knowing, heuristic attribution of epistemic credibility, epistemic friction and disruption of trust, verification and calibration of trust, educational and social regulation of epistemic trust, and consequences for epistemic agency. These categories were integrated around the core category of "calibrating epistemic delegation under conditions of algorithmic opacity." The findings indicated a dynamic process progressing from pragmatic initiation and heuristic credibility attribution to provisional reliance, epistemic friction, active verification and calibration, and conditional integration. Contextual and intervening factors, including prior knowledge, task stakes, disciplinary norms, AI literacy, information literacy, teacher guidance, and previous negative experiences, shaped this process. Two longer-term trajectories emerged: reflective epistemic augmentation and epistemic dependency.

Conclusion: Epistemic trust in AI-generated content is a conditional, dynamic, and revisable process in which educationally responsible use depends on maintaining learner

responsibility, critical verification, evidential judgment, and appropriate limits on epistemic delegation.

Keywords: *Epistemic Trust; Generative Artificial Intelligence; University Students; Grounded Theory; Philosophy of Education; Epistemic Agency; AI-Generated Content*

1. Introduction

The rapid expansion of generative artificial intelligence (AI) has introduced a new epistemic actor into contemporary higher education. Unlike earlier educational technologies that primarily delivered, stored, or organized information, generative AI systems can produce explanations, arguments, summaries, interpretations, examples, recommendations, and apparently authoritative responses in natural language. This transformation has altered the relationship between students and knowledge because learners increasingly encounter not merely technological tools but systems that appear capable of participating in processes traditionally associated with knowledgeable human agents. Artificial intelligence is now embedded in educational activities ranging from tutoring and feedback to writing support, problem solving, assessment, and research assistance, creating new opportunities for individualized learning while simultaneously generating questions about accuracy, authority, transparency, responsibility, and the legitimacy of machine-mediated knowledge. The broader diffusion of AI across sectors has been shaped by technological capability, perceived usefulness, organizational readiness, and users' willingness to adopt intelligent systems (Fulton et al., 2022), yet educational adoption raises distinctive concerns because decisions about trusting AI directly influence how students learn, justify beliefs, evaluate evidence, and construct academic knowledge. Research on students' attitudes toward AI-based innovations has already shown that acceptance is related not only to functionality but also to expectations, perceived benefits, and users' understanding of the technology (Le et al., 2022). Similarly, early studies of student attitudes and trust have indicated that learners may develop relatively favorable orientations toward AI even while remaining uncertain about its reliability and consequences (Arif, 2023). These developments suggest that the educational significance of AI cannot be adequately understood through adoption or usability alone; rather, it requires examination of the epistemic relationship that develops when students begin to treat algorithmically generated content as a source worthy of belief, use, verification, or rejection.

This issue becomes especially important because trust in AI differs fundamentally from trust in conventional educational sources. A student who consults a textbook, scholarly article, lecturer, or disciplinary expert typically operates within recognizable structures of authorship, accountability, expertise, and evidential provenance. Generative AI, by contrast, produces highly fluent outputs without always making transparent how particular claims were derived, what evidence supports them, or whether the apparent certainty of the response corresponds to actual reliability. The problem is therefore not merely one of technological trust but of epistemic trust: the willingness to treat information originating from an external source as sufficiently credible, relevant, and reliable to influence one's beliefs, judgments, or actions. Studies of trustworthiness in AI increasingly emphasize that trust is multidimensional and depends on users' perceptions of competence, transparency, predictability, fairness, explainability, and accountability (Zhang et al., 2024). Explainable AI has consequently become a major field of research because users' ability to understand or interrogate the basis of algorithmic outputs is considered essential to responsible reliance (Mohammed, 2023). In educational settings, calls for interpretable algorithmic mechanisms have similarly emphasized that opacity can undermine meaningful oversight and make it difficult to determine whether automated systems should be trusted in consequential decisions (Liu & Cheng, 2022). Research on automated educational feedback has also highlighted the importance of establishing trustworthy AI through mechanisms that support transparency, reliability, and appropriate interpretation of machine-generated guidance (Lin et al., 2022). These concerns become particularly acute in generative AI because the linguistic fluency of the output can create an impression of epistemic authority even when the underlying content is incomplete, uncertain, fabricated, or unsupported.

From the perspective of philosophy of education, this shift raises a fundamental question concerning the status of AI in the ecology of knowledge. Education is not reducible to acquiring correct answers; it involves learning how knowledge is justified, how evidence is weighed, how disagreement is evaluated, and how learners become intellectually responsible agents capable of exercising

Iranian Journal of Educational Sociology
E-ISSN: 2645-3460

independent judgment. Generative AI complicates this process because it can produce persuasive answers before the student has engaged in the cognitive and epistemic work through which understanding is ordinarily formed. Machin-Mastromatteo has argued that generative AI requires a reconsideration of knowledge, ethics, and research practices precisely because it blurs established boundaries between information production, authorship, interpretation, and epistemic responsibility (Machin-Mastromatteo, 2025). Ethical analyses of AI in education likewise indicate that the use of intelligent systems must be evaluated in relation to autonomy, fairness, transparency, privacy, responsibility, and the broader purposes of education (Yan & Chau, 2025). Similar arguments concerning Education 5.0 stress that AI should be integrated in ways that preserve human-centered values and ethical agency rather than allowing technological efficiency to displace educational judgment (Smyrnaiou et al., 2023). The emerging discussion of trustworthy learning environments therefore increasingly emphasizes reciprocal alignment between humans and AI, whereby systems should be designed to support human values and learners should simultaneously develop the capacity to interact with AI critically and responsibly (Shen, 2025). Under this view, epistemic trust is not simply a psychological attitude toward a tool but a philosophical and educational problem concerning who or what is granted authority to define, explain, and validate knowledge.

Empirical studies in higher education demonstrate that students already engage with generative AI in ways that make this problem concrete. Research on ChatGPT as a virtual tutor in physics has shown that students can perceive generative AI as useful for obtaining explanations and supporting problem solving, while their evaluations remain shaped by concerns about correctness and reliability (Ding et al., 2023). Studies of students' and faculty members' perceptions of generative AI similarly reveal a mixture of enthusiasm, uncertainty, and concern regarding appropriate use, academic integrity, accuracy, and educational value (Petricini et al., 2023). In college writing contexts, students' use of ChatGPT has been associated with complex judgments concerning learning, grading, fairness, and trust, indicating that perceived usefulness does not necessarily translate into uncomplicated confidence in generated outputs (Tossell et al., 2024). From a student-centered perspective, human–AI collaboration in higher education appears to be most productive when students remain active participants in interpreting, refining, and evaluating AI contributions rather than treating generated content as inherently authoritative (Razmerita, 2024). These findings suggest that students do not merely "use" AI; they negotiate different forms and degrees of reliance according to context, task, prior knowledge, perceived risk, and expected academic consequences. Such negotiation is particularly important because trust can develop incrementally through repeated successful interactions, and usefulness may gradually be generalized into a broader assumption of epistemic reliability even when the system's performance varies considerably across domains and tasks.

Trust is also shaped by the characteristics through which AI presents itself to users. Human beings frequently rely on heuristic cues when evaluating sources, especially under conditions of uncertainty or limited time. In digital environments, anthropomorphic qualities can influence trust, confidence, and willingness to disclose or rely on information (Wang et al., 2024). Generative AI intensifies this issue because its conversational design, apparent responsiveness, and capacity to imitate human explanatory styles can create a sense of interpersonal competence even though the system does not possess human understanding or accountability in the conventional sense. Trust can likewise be strengthened by external signals of reliability. Research on AI certification labels has shown that visible indicators of assurance can increase adoption by strengthening users' trust in otherwise opaque systems (Almeida et al., 2026). In education, however, such external assurances remain uneven, and students often confront AI-generated information without clear institutional mechanisms for evaluating its provenance or evidential status. The development of trustworthy AI for educational metaverses and other immersive learning environments has therefore emphasized the importance of security, transparency, reliability, and user confidence as foundational requirements for responsible educational deployment (Abdullakutty & Qayyum, 2024). At the same time, the difference between perception and expectation remains consequential: students and other users may expect AI systems to be more accurate, neutral, or capable than they actually are, creating a gap that can distort patterns of reliance (Skenderi et al., 2024). This gap is epistemically significant because trust based on perceived competence rather than demonstrated reliability can encourage students to accept content without sufficient scrutiny.

The educational consequences of such trust are not uniformly negative or positive. Properly calibrated reliance on AI may extend students' cognitive resources, support exploration, facilitate explanation, increase access to

Iranian Journal of Educational Sociology
E-ISSN: 2645-3460

feedback, and help learners identify gaps in their understanding. Yet excessive or poorly calibrated trust may produce overreliance, reduce engagement with primary sources, weaken independent problem solving, and make students less attentive to uncertainty or evidential justification. The quality of AI-generated feedback is especially important in this regard. Research comparing feedback perceptions suggests that students respond not only to the identity of the feedback provider but also to the perceived quality of the feedback itself, which can influence whether machine-generated guidance is accepted, acted upon, or rejected (Ruwe & Kuklick, 2025). In professional education, this problem has led to the development of entrustment-based frameworks designed to specify what kinds of tasks may appropriately be delegated to AI and under what levels of human oversight (Gin et al., 2024). Such frameworks are particularly relevant to epistemic trust because they shift the question from whether AI is trustworthy in general to whether it is trustworthy for a particular task, under particular conditions, with particular safeguards. Similar concerns appear in standardized assessment, where algorithmic systems must be judged according to validity, transparency, fairness, and the degree to which stakeholders can legitimately place confidence in automated scoring processes (Aloisi, 2023). These studies support the idea that trust should be calibrated rather than absolute, and that the educationally meaningful question concerns the conditions under which reliance is epistemically justified.

The distinction between calibrated and uncalibrated trust becomes even more important when students lack the expertise required to detect errors. In specialized domains, knowledge and experience strongly affect how users evaluate AI-assisted outputs. Research involving orthodontists and orthodontic students, for example, has shown that knowledge, experience, and attitudes toward AI-assisted analysis vary across levels of expertise, suggesting that familiarity with a domain influences how intelligent systems are perceived and used (Lin et al., 2023). This creates a paradox for students: AI may be most attractive precisely when they know the least about a topic, yet low prior knowledge also reduces their ability to evaluate whether the generated content is correct. Consequently, apparent coherence, detail, confidence, or stylistic sophistication may become substitutes for genuine verification. Public perceptions of AI in healthcare similarly demonstrate that trust is influenced by knowledge, familiarity, and perceived benefits and risks, rather than by technical performance alone (Syed et al., 2024). The same principle is relevant in educational settings, where students' judgments may be conditioned by their digital literacy, academic experience, disciplinary knowledge, previous encounters with AI errors, and exposure to institutional guidance. The need to distinguish perceived trustworthiness from warranted epistemic trust is therefore central to understanding how learners interact with AI-generated knowledge.

Another important dimension concerns the ethical and institutional environment surrounding AI use. Learning technologies increasingly collect, process, infer, and generate information in ways that challenge existing concepts of privacy, fairness, accountability, and informed participation. Ethical frameworks for multimodal learning analytics emphasize that responsible use requires continuous attention to transparency, stakeholder interests, proportionality, and potential harms (Alwahaby & Cukurova, 2024). Although generative AI differs technically from multimodal analytics, similar principles apply because students are required to make decisions about systems whose operations may be partly inaccessible to them. The broader development of national and institutional AI research infrastructures also shows that trustworthy AI is not merely an individual concern but a strategic societal issue requiring interdisciplinary research, governance, and sustained public investment (Donlon, 2024). In higher education, institutional responses to AI therefore shape the conditions under which epistemic trust develops. Policies that focus exclusively on prohibition may leave students without the conceptual tools needed to assess AI critically, whereas unregulated adoption may normalize reliance without adequate attention to verification or responsibility. What is required is an educational framework in which students learn not merely how to prompt AI effectively but how to interrogate claims, identify uncertainty, evaluate sources, and preserve their own epistemic agency.

Recent work increasingly suggests that trust in generative AI develops dynamically over time. Rather than beginning from either complete skepticism or complete acceptance, students often move through stages of experimentation, provisional confidence, disruption, and reconstruction. Zhou's study of upper secondary students' engagement with generative AI in science learning explicitly conceptualizes this movement as a transition from initial trust toward critical reconstruction, showing how encounters with limitations can prompt more reflective forms of engagement (Zhou, 2026). This developmental perspective is highly

Iranian Journal of Educational Sociology
E-ISSN: 2645-3460

relevant to higher education because university students repeatedly interact with AI across different tasks, accumulating experiences that may strengthen, weaken, differentiate, or recalibrate their trust. Çetinkaya's discussion of trust in AI in medical education similarly underscores that trust should not be understood as a simple binary but as a carefully managed relationship shaped by context, competence, transparency, and responsibility (Çetinkaya, 2025). The same insight can be extended to the broader educational environment: students may trust AI for brainstorming but distrust it for references, rely on it for language improvement but verify theoretical claims, or accept it as a first-stage assistant while refusing to grant it final epistemic authority. These patterns suggest that epistemic trust is processual, situational, and conditional, and that it is formed through repeated interaction between technological affordances, learner characteristics, educational norms, and experiences of confirmation or error.

Despite the growing literature on student perceptions, acceptance, AI ethics, explainability, educational applications, and trustworthy AI, an important conceptual gap remains. Much existing research examines whether students trust AI, how much they trust it, or what factors predict their willingness to use it, but fewer studies explain how epistemic trust is actually constructed, disrupted, recalibrated, and integrated into students' everyday academic practices. Existing studies illuminate important components of the problem, including students' attitudes (Arif, 2023), perceptions of generative AI (Petricini et al., 2023), task-specific trust (Tossell et al., 2024), human–AI collaboration (Razmerita, 2024), feedback quality (Ruwe & Kuklick, 2025), algorithmic transparency (Mohammed, 2023), ethical governance (Yan & Chau, 2025), and evolving forms of critical engagement (Zhou, 2026); however, these strands have not yet been sufficiently integrated into a process model grounded in students' lived academic experiences. Moreover, the philosophical dimension of this process remains underdeveloped. Questions concerning when information becomes knowledge, how justification is preserved when AI participates in reasoning, where epistemic responsibility resides, and how students maintain intellectual autonomy in the presence of persuasive algorithmic outputs require attention from the philosophy of education, not only from technology adoption or human–computer interaction research. A grounded theory approach is particularly appropriate for addressing this gap because it permits the development of an explanatory model from students' experiences rather than imposing a predetermined structure of trust. By examining how learners describe initial reliance, credibility judgments, encounters with error, verification practices, social and institutional influences, and longer-term changes in academic behavior, such an approach can clarify the mechanisms through which AI becomes either an epistemic scaffold or a source of dependency. Accordingly, the aim of this study was to explain the process through which university students develop, evaluate, recalibrate, and regulate epistemic trust in AI-generated content from the perspective of philosophy of education using a grounded theory approach.

2. Methods and Materials

2.1. Study Design and Participants

This study employed a qualitative research design based on grounded theory to explain the process through which university students develop, maintain, revise, or withdraw epistemic trust in content generated by artificial intelligence. Grounded theory was selected because the phenomenon under investigation involves a dynamic and context-dependent process rather than a fixed set of attitudes or behaviors. The study sought to move beyond merely identifying students' positive or negative perceptions of artificial intelligence and instead develop an explanatory model of how they judge the credibility, reliability, epistemic authority, and educational legitimacy of AI-generated information during their academic activities. The inquiry was informed by the perspective of philosophy of education, particularly questions concerning the nature of knowledge, epistemic authority, justification, intellectual autonomy, critical judgment, the relationship between learner and source of knowledge, and the educational consequences of delegating cognitive and interpretive activities to intelligent systems. Accordingly, students' trust in AI-generated content was conceptualized not simply as technological acceptance but as an epistemic process through which learners determine whether, under what conditions, and to what extent an artificial intelligence system can be regarded as a credible source of knowledge within educational contexts.

The participants consisted of 24 university students studying at universities and higher education institutions in Tehran, Iran. Participants were recruited through purposive sampling in the initial phase and theoretical sampling in subsequent phases of the study. The initial participants were selected to provide diverse experiences of using generative

Iranian Journal of Educational Sociology
E-ISSN: 2645-3460

artificial intelligence tools for educational and academic purposes, including searching for information, explaining difficult concepts, summarizing scholarly materials, generating ideas, preparing assignments, improving academic writing, translating texts, solving academic problems, and obtaining feedback on learning tasks. As data analysis progressed, theoretical sampling was employed to identify participants whose experiences could clarify emerging concepts, elaborate relationships among categories, or provide contrasting cases. Particular attention was given to variation in participants' academic discipline, level of study, frequency of artificial intelligence use, degree of familiarity with generative AI tools, and level of dependence on AI-generated information in academic activities. This diversity was considered important because epistemic trust may develop differently depending on students' disciplinary expectations, prior knowledge, digital literacy, academic experience, and perceived consequences of relying on artificial intelligence.

Eligibility criteria included being an enrolled university student in Tehran, having direct experience using generative artificial intelligence systems for academic or educational purposes, having used such systems sufficiently to be able to describe situations involving trust, doubt, verification, acceptance, or rejection of AI-generated content, and willingness to participate in an in-depth interview. Students whose experience with artificial intelligence was limited to non-academic entertainment or who had only superficial or one-time exposure to generative AI tools were not included because they could not provide sufficiently rich accounts of the epistemic processes under investigation. Sampling and data collection proceeded concurrently with analysis, in accordance with grounded theory methodology. Recruitment continued until theoretical saturation was achieved, meaning that additional interviews no longer produced substantively new properties of the main categories, the relationships among categories had become sufficiently developed, and the emerging explanatory model adequately accounted for variations in participants' experiences. The final sample of 24 participants was therefore determined by theoretical saturation rather than by a predetermined statistical requirement. Throughout the study, participation was voluntary, participants were informed about the purpose and procedures of the research, confidentiality and anonymity were maintained, and informed consent was obtained before the interviews. Participants were also informed that they could discontinue participation at any stage without any negative consequences.

2.2. Instruments

Data were collected primarily through individual semi-structured, in-depth interviews designed to elicit detailed narratives concerning students' experiences of encountering, evaluating, trusting, questioning, verifying, and using AI-generated content in educational settings. A semi-structured interview guide was developed in accordance with the research purpose and the conceptual concerns of philosophy of education, while sufficient flexibility was preserved to allow participants to describe their experiences in their own terms. The initial interview questions focused on participants' experiences of using artificial intelligence for academic purposes, circumstances in which they considered AI-generated information trustworthy or untrustworthy, criteria used to assess the validity of AI responses, experiences in which artificial intelligence provided inaccurate or misleading information, strategies adopted to verify generated content, the role of prior knowledge in evaluating AI outputs, and the perceived influence of artificial intelligence on independent thinking and learning. Participants were also encouraged to reflect on whether they considered artificial intelligence a tool, an assistant, a source of information, or a form of epistemic authority, and how their perception of its role influenced the extent to which they accepted its outputs.

As data collection and analysis progressed, the interview guide evolved in response to emerging categories, consistent with the logic of theoretical sampling and constant comparison. Later interviews therefore explored increasingly focused issues such as the transition from initial curiosity to habitual reliance, the role of repeated confirmation or error in shaping trust, the distinction between functional trust and epistemic trust, conditions under which students suspended their own judgment in favor of AI-generated responses, and factors that motivated them to independently verify information. Additional probes addressed the influence of teachers, peers, institutional expectations, academic assessment practices, disciplinary norms, and students' perceptions of the authority of conventional educational sources. Questions also explored whether students regarded transparency, citation provision, linguistic confidence, logical coherence, speed of response, personalization, and apparent expertise as indicators of epistemic credibility. Particular attention was paid to

Iranian Journal of Educational Sociology
E-ISSN: 2645-3460

situations in which persuasive or fluent AI-generated content was accepted despite uncertainty regarding its factual basis, because such experiences were theoretically relevant to understanding the distinction between perceived credibility and justified epistemic trust.

Interviews were conducted in a private and comfortable setting, either face-to-face or through secure online communication when necessary, according to participant preference. Each interview lasted approximately 45 to 75 minutes. With participants' permission, interviews were digitally audio-recorded and subsequently transcribed verbatim. Field notes were prepared immediately after each interview to document contextual observations, significant pauses, emphases, contradictions, emotional responses, preliminary analytic ideas, and possible directions for theoretical sampling. Analytical memos were also used throughout the data collection process to capture emerging interpretations and conceptual questions. Data collection was therefore not treated as a separate phase preceding analysis; rather, interviews, memo writing, comparison, conceptual development, and theoretical sampling were conducted iteratively. This recursive process allowed emerging theoretical insights to shape subsequent interviews and enabled the researchers to examine both common patterns and deviant or contrasting experiences.

To enhance the credibility and depth of the collected data, participants were encouraged to provide concrete examples rather than general opinions. When a participant stated that an AI-generated response was trustworthy, for example, follow-up questions explored what features produced that judgment, whether the information had been independently verified, what alternative sources were available, and whether the participant would have made the same judgment in a high-stakes academic situation. Similarly, when participants described distrust, they were asked to explain whether distrust resulted from previous factual errors, uncertainty about sources, disciplinary limitations, ethical concerns, contradictions with prior knowledge, or broader doubts about artificial intelligence as a legitimate knowledge-producing system. This approach facilitated the collection of data concerning not only what students believed about AI-generated content but also how epistemic judgments were formed and modified through interaction with artificial intelligence. The resulting interview material provided rich accounts of the cognitive, experiential, educational, social, and normative conditions underlying epistemic trust.

2.3. Data Analysis

Data analysis was conducted according to grounded theory procedures through an iterative process of coding, constant comparison, memo writing, category development, and theoretical integration. Analysis began immediately after the first interview and continued concurrently with subsequent data collection. Each transcript was read several times to obtain an overall understanding of the participant's experience before detailed coding commenced. During the initial coding stage, the interview transcripts were examined line by line and meaningful units were assigned conceptual labels that remained as close as possible to participants' experiences while simultaneously identifying underlying actions, judgments, conditions, and processes. Initial codes captured phenomena such as accepting responses because of linguistic confidence, comparing AI outputs with prior knowledge, checking responses against scholarly sources, becoming more cautious after encountering fabricated information, relying on AI for low-risk tasks, distinguishing explanation from authoritative knowledge, transferring cognitive responsibility to artificial intelligence, and developing conditional rather than absolute trust. Coding emphasized processes and actions rather than merely descriptive topics because the primary objective was to explain how epistemic trust developed and changed over time.

Constant comparative analysis was employed throughout the analytic process. Codes were continuously compared with other codes, incidents with other incidents, participant accounts with contrasting accounts, and emerging categories with new data. Through this process, conceptually similar codes were grouped into more abstract categories, while differences among experiences were examined to identify properties, dimensions, conditions, and variations within each category. As categories became more developed, analysis focused increasingly on relationships among contextual conditions, students' interpretive strategies, mediating factors, actions and interactions, and epistemic consequences. For example, experiences of factual accuracy, perceived coherence, responsiveness to user questioning, availability of external verification, disciplinary expertise, previous encounters with hallucinated information, and perceived academic risk were examined in relation to changes in the level and form of students' epistemic trust. Particular analytic attention was given to identifying circumstances under which students moved from cautious experimentation to selective reliance, from selective reliance

Iranian Journal of Educational Sociology
E-ISSN: 2645-3460

to habitual dependence, or from initial confidence to critical skepticism.

During focused and axial phases of analysis, the most analytically significant and frequently occurring codes were used to organize the data around higher-order categories. Relationships among categories were systematically examined in terms of conditions, actions and interactions, intervening influences, and consequences. The analysis sought to determine what prompted students to trust AI-generated content, what conditions reinforced or destabilized this trust, what strategies students used when uncertainty arose, and what educational consequences followed from different forms of reliance. Categories were continually refined through negative-case analysis and comparison with participants whose experiences differed from dominant patterns. Theoretical sampling was then directed toward filling conceptual gaps and elaborating categories that remained insufficiently developed. For example, when emerging analysis suggested that prior subject-matter knowledge substantially influenced the ability to evaluate AI-generated information, subsequent participants were selected and questioned to explore differences between students using artificial intelligence within familiar and unfamiliar domains. Similarly, emerging concerns regarding the relationship between convenience and epistemic dependence guided additional exploration of situations in which students accepted content despite recognizing uncertainty about its validity.

At the theoretical integration stage, a central category was identified that possessed sufficient explanatory power to connect the major categories and account for the overall process of epistemic trust. The core category was selected on the basis of its recurrence across participants' accounts, its relationship with the principal categories, its ability to explain variation in students' experiences, and its relevance to the research question. The relationships among the core category, causal and contextual conditions, intervening factors, students' epistemic strategies, and educational consequences were then progressively refined to construct an explanatory grounded theory model. Analytical memos played an important role in this stage by documenting conceptual decisions, emerging propositions, questions about category relationships, and interpretations informed by philosophy of education. The philosophical interpretation focused particularly on the tension between epistemic autonomy and cognitive delegation, the legitimacy of algorithmic authority, the learner's responsibility for justification, the difference between access to information and possession of knowledge, and the educational significance of maintaining critical judgment when interacting with artificial intelligence.

Several procedures were employed to enhance the trustworthiness of the qualitative analysis. Credibility was strengthened through prolonged engagement with the data, iterative interviewing, constant comparison, theoretical sampling, examination of contrasting cases, and participant validation of selected interpretations. Portions of the coding structure and category definitions were reviewed through peer examination to assess conceptual coherence and reduce the likelihood that categories reflected only a single researcher's interpretation. Dependability was supported by maintaining an audit trail that documented interview procedures, coding decisions, category revisions, theoretical sampling decisions, and analytic memos. Confirmability was enhanced through reflexive memo writing, through which the researchers explicitly examined their assumptions concerning artificial intelligence, education, trust, and epistemic authority and distinguished these assumptions from interpretations grounded in participants' accounts. Transferability was facilitated by providing detailed descriptions of the participants, educational context, AI-use experiences, and conceptual conditions associated with different forms of epistemic trust. Analysis continued until the categories were conceptually dense, their properties and relationships were sufficiently specified, and the resulting theoretical framework provided a coherent explanation of how university students construct and regulate epistemic trust in AI-generated content within contemporary educational environments.

3. Findings and Results

The final sample consisted of 24 university students studying at universities and higher education institutions in Tehran. Participants ranged in age from 19 to 37 years, with a mean age of 26.8 years (SD = 4.6). Thirteen participants were women and 11 were men. In terms of educational level, eight participants were undergraduate students, 10 were master's students, and six were doctoral students. The sample was intentionally heterogeneous with regard to academic discipline in order to capture variations in epistemic judgment across fields with different conventions of evidence, verification, and knowledge production. Seven participants studied humanities or social sciences, six studied engineering or computer-related disciplines, four studied medical or health sciences, three studied basic or

Iranian Journal of Educational Sociology
E-ISSN: 2645-3460

natural sciences, two studied management or economics, and two studied arts or architecture. Participants also differed substantially in their pattern of generative artificial intelligence use. Eleven reported using generative AI tools almost every day for academic purposes, eight used them several times per week, and five used them approximately once per week or less frequently. Four participants had less than six months of experience with generative AI, eight had between six and 12 months of experience, eight had between one and two years of experience, and four reported more than two years of experience. The most frequently reported academic applications included improving or generating academic text, obtaining explanations of difficult concepts, preliminary information searching, summarizing materials, translating academic content, generating ideas, solving disciplinary problems, and receiving feedback on assignments. Importantly, differences in frequency of use did not correspond directly to differences in epistemic trust. Some frequent users demonstrated highly conditional and reflective forms of trust, whereas some less experienced users showed relatively rapid acceptance of fluent AI-generated responses. The interviews therefore indicated from the outset that epistemic trust was better understood as a process of judgment and regulation than as a simple function of frequency of technology use.

Table 1

Development of the Coding Structure Through Grounded Theory Analysis

Analytic level	Analytic output	Number retained	Examples of analytic content	Function in theory development
Meaningful units	Segments of participants' accounts containing an identifiable experience, judgment, action, or consequence related to AI-generated content	1,276	"I usually believe it first and check later"; detecting incorrect references; comparing an answer with lecture notes; using AI when deadlines are close; doubting overly confident responses	Established the empirical foundation of the analysis
Initial codes	Close-to-data conceptual labels generated through line-by-line analysis	482	trusting fluent language; assuming expertise from completeness; checking unfamiliar claims; confidence reduced by fabricated references; delegating first-stage search; avoiding AI for high-stakes decisions	Captured actions, perceptions, conditions, and changes in epistemic judgment
Focused codes	More conceptually significant codes developed through constant comparison and consolidation	96	credibility through coherence; familiarity-based verification; error-triggered distrust; source opacity; risk-sensitive checking; iterative questioning; selective delegation; dependency awareness	Reduced descriptive redundancy and identified recurrent processes
Subcategories	Conceptual clusters representing specific dimensions of the emerging process	22	perceived efficiency; surface credibility cues; confirmation with prior knowledge; epistemic friction; provenance checking; human consultation; risk zoning; teacher regulation; reflective autonomy	Specified properties and dimensions of the major categories
Main categories	Higher-order categories explaining major components of epistemic trust formation	6	pragmatic entry into AI-mediated knowing; heuristic attribution of credibility; disruption through epistemic friction; verification and calibration; educational-social regulation; consequences for epistemic agency	Organized the explanatory structure of the emerging theory
Core category	Integrative concept linking conditions, actions/interactions, and consequences	1	Calibrating epistemic delegation under conditions of algorithmic opacity	Explained the central process through which students determined how much cognitive and epistemic authority could be delegated to AI

Analysis of the 24 interviews generated 1,276 meaningful units that were initially reduced to 482 substantive codes through line-by-line coding. Continuous comparison of incidents, expressions, and actions revealed considerable conceptual overlap among the initial codes, leading to their consolidation into 96 focused codes. These focused codes were subsequently organized into 22 subcategories and six higher-order categories. The analytic process indicated that the participants' experiences could not adequately be described through a simple trust–distrust dichotomy. Rather, trust changed according to task type, previous experiences with accuracy and error, students' own knowledge, the availability of alternative sources, academic consequences, and perceived responsibility for the final product. The core category, "calibrating epistemic delegation under conditions of algorithmic opacity," emerged as the concept with the

Iranian Journal of Educational Sociology
E-ISSN: 2645-3460

greatest explanatory capacity because it connected almost all major patterns in the data. Participants were repeatedly engaged in decisions about how much of the processes of searching, interpreting, explaining, composing, evaluating, or deciding could legitimately be transferred to an AI system whose internal basis for generating a particular response was not fully visible to them. The central problem was therefore not simply whether AI was regarded as trustworthy, but how students negotiated the boundary between using AI as a cognitively useful resource and permitting it to function as an unexamined epistemic authority. The coding structure further demonstrated a developmental tendency from initially pragmatic and heuristic judgments toward more conditional forms of trust among participants who had encountered errors, developed verification practices, or received explicit educational guidance concerning AI limitations.

Table 2

Main Categories and Subcategories of University Students' Epistemic Trust in AI-Generated Content

Main category	Subcategories	Central meaning	Typical manifestations in participants' experiences
Pragmatic entry into AI-mediated knowing	Perceived efficiency and immediacy; accessibility and personalization; academic time pressure; prior familiarity with digital assistance	Students usually entered the relationship with AI through practical usefulness rather than an explicit judgment regarding its epistemic legitimacy	Asking AI before consulting textbooks; using AI because it was immediately available; requesting simplified explanations; resorting to AI during deadlines; treating it as a convenient first point of inquiry
Heuristic attribution of epistemic credibility	Fluency and coherence; completeness and apparent confidence; perceived general expertise; consistency with prior knowledge	Surface and experiential cues were initially used as substitutes for direct access to the evidential basis of generated claims	Equating well-written answers with correctness; trusting structured responses; assuming that detailed explanations reflected expertise; accepting answers that "sounded right"; believing information that confirmed existing knowledge
Epistemic friction and disruption of trust	Factual or conceptual errors; fabricated or unreliable references; inconsistency across prompts; source and reasoning opacity	Encounters with error interrupted automatic acceptance and made the epistemic limitations of AI visible	Discovering nonexistent references; receiving different answers to similar prompts; identifying disciplinary mistakes; becoming uncertain because the origin of information could not be established
Verification and calibration of trust	Cross-source corroboration; provenance and citation checking; iterative prompting; consultation with knowledgeable humans; risk-sensitive boundary setting	Students developed procedures for deciding when AI could be relied upon and when independent verification was necessary	Checking textbooks and databases; opening cited sources; asking the same question in several forms; consulting instructors; verifying high-stakes claims more carefully; limiting AI use in unfamiliar domains
Educational and social regulation of epistemic trust	Teacher and institutional guidance; peer practices and normalization; disciplinary and assessment expectations	Trust was shaped by the social organization of knowledge within the university rather than by student–technology interaction alone	Adopting instructors' warnings; imitating peers' use patterns; differentiating acceptable use across courses; modifying reliance according to assignment rules and disciplinary expectations
Consequences for epistemic agency	Reflective augmentation of learning; cognitive outsourcing and dependency	AI use could either support students' epistemic autonomy or weaken active responsibility for knowing, depending on how trust was regulated	Using AI to initiate inquiry but independently evaluating results; improved questioning and comparison; accepting answers without understanding; reduced reading of original sources; difficulty completing tasks without AI

The category structure presented in Table 2 indicates that epistemic trust developed through a sequence of interacting practical, cognitive, social, and educational processes. The first category, pragmatic entry into AI-mediated knowing, showed that students generally did not begin using generative AI after consciously deciding that it constituted a reliable epistemic source. Instead, the relationship was initiated through accessibility, speed, personalization, and the ability to reduce the immediate cognitive cost of academic work. This pragmatic beginning created opportunities for the second category, heuristic attribution of epistemic credibility, in which properties such as linguistic fluency, organized presentation, rapid responsiveness, and apparent comprehensiveness were interpreted as indicators of knowledgeability. However, the interviews showed that this form of trust was inherently unstable. The third category, epistemic friction and disruption of trust, became visible when students encountered invented references, incorrect concepts, inconsistent answers, or information that conflicted with trusted disciplinary knowledge. These disruptions were important turning points because they transformed AI from an apparently transparent provider of answers into an epistemically uncertain system requiring active judgment. The fourth category, verification and

Iranian Journal of Educational Sociology
E-ISSN: 2645-3460

calibration of trust, represented the central set of actions through which students attempted to manage this uncertainty. Rather than permanently accepting or rejecting AI after encountering errors, many participants developed conditional verification routines. The fifth category showed that these routines were influenced by teachers, peers, academic norms, and assessment structures. Finally, the sixth category demonstrated that the educational consequences depended substantially on how successfully students maintained control over the epistemic relationship. When AI was incorporated into a process of questioning, comparison, source evaluation, and self-explanation, it tended to function as an epistemic scaffold. When convenience progressively replaced evaluation, reading, and independent reasoning, the same technology contributed to cognitive outsourcing and a weakening of epistemic agency.

Table 3

Causal, Contextual, and Intervening Conditions Shaping Epistemic Trust

Condition type	Category of condition	Specific dimensions	Effect on epistemic trust
Causal conditions	Perceived usefulness	Speed, availability, immediate explanation, capacity to reformulate information	Encouraged initial engagement and lowered the threshold for accepting AI as an informational resource
Causal conditions	Repeated satisfactory performance	Previous correct answers, successful completion of tasks, useful explanations	Produced accumulated confidence and encouraged broader delegation
Causal conditions	Cognitive and academic pressure	Time constraints, information overload, difficult assignments, simultaneous responsibilities	Increased willingness to accept responses without extensive verification
Causal conditions	Perceived informational breadth	Ability to respond across multiple topics and formats	Encouraged attribution of generalized competence beyond domains in which accuracy had actually been established
Contextual conditions	Familiarity with the subject	High versus low prior knowledge	Greater knowledge enabled error detection and conditional trust; low knowledge increased vulnerability to apparent credibility
Contextual conditions	Stakes of the task	Informal learning, routine assignment, examination, thesis, publication, clinical or technical implication	Higher stakes generally increased verification and reduced unconditional reliance
Contextual conditions	Type of academic task	Brainstorming, summarization, translation, explanation, factual retrieval, argument construction, numerical or technical problem solving	Trust thresholds differed by task; generative and preliminary tasks were typically delegated more readily than authoritative factual decisions
Contextual conditions	Disciplinary norms	Standards of evidence, citation practices, tolerance for ambiguity, dependence on current information	Determined which forms of AI output were perceived as acceptable or epistemically insufficient
Contextual conditions	Availability of alternative sources	Access to textbooks, databases, instructors, peers, or specialized resources	Greater access facilitated corroboration and reduced exclusive dependence on AI
Intervening conditions	AI literacy	Knowledge of hallucination, probabilistic generation, prompting limitations, and source uncertainty	Increased skepticism toward confident presentation and encouraged verification
Intervening conditions	Information and source literacy	Ability to evaluate references, search databases, distinguish primary from secondary evidence	Strengthened the transition from surface credibility to evidence-based judgment
Intervening conditions	Previous negative experiences	Exposure to fabricated references, factual errors, inconsistency, or misleading explanations	Frequently served as a trigger for recalibration and more selective trust
Intervening conditions	Teacher guidance	Explicit rules, modeling of critical AI use, discussion of limitations	Encouraged reflective use and clarified responsibility for verification
Intervening conditions	Peer norms	Normalization, collective experimentation, peer warnings, shared verification practices	Could either normalize rapid acceptance or reinforce critical checking
Intervening conditions	Personal epistemic disposition	Tolerance for uncertainty, need for certainty, confidence in one's own knowledge, willingness to question authoritative-looking responses	Influenced how quickly participants accepted, questioned, or suspended judgment

The relationships shown in Table 3 reveal that epistemic trust was produced by the interaction of several levels of influence rather than by an isolated evaluation of AI accuracy. Perceived usefulness, previous successful experiences, academic pressure, and the apparent breadth of AI capabilities acted as major causal conditions encouraging students to incorporate AI into their academic work. Repeated usefulness was especially important because satisfactory experiences accumulated into a form of experiential confidence; however, this confidence was often generalized beyond the specific tasks in which AI had previously performed adequately. Contextual conditions determined whether such confidence resulted in acceptance or verification. Students with strong prior knowledge were

Iranian Journal of Educational Sociology
E-ISSN: 2645-3460

generally better able to recognize conceptual anomalies, inappropriate terminology, oversimplification, or inconsistencies and consequently tended to use AI more selectively. In contrast, when students approached unfamiliar subjects, they frequently lacked an independent reference point from which to assess plausibility, making linguistic fluency disproportionately influential. Task stakes produced another important variation. Participants were generally more willing to rely on AI for brainstorming, language refinement, preliminary summaries, and initial explanations, whereas thesis writing, formal citation, examination preparation, specialized technical claims, and information with significant academic consequences produced greater caution. Intervening conditions explained why participants exposed to similar AI systems behaved differently. AI literacy and information literacy made students more sensitive to the distinction between a plausible answer and a justified answer. Previous experiences of hallucination frequently disrupted generalized confidence, while teacher guidance provided explicit standards for acceptable epistemic delegation. Peer culture could function in either direction: in some contexts frequent collective use normalized dependence, whereas in others the circulation of examples of AI error encouraged skepticism. These findings demonstrate that epistemic trust was situated within a broader ecology of knowledge practices and could not be reduced to the technological properties of generative AI alone.

Table 4

Processual Stages in the Formation and Recalibration of Epistemic Trust

Process stage	Dominant orientation	Typical epistemic judgment	Characteristic student behavior	Transition mechanism
Pragmatic initiation	Utility-oriented engagement	"This is useful and immediately available"	Asking AI for explanations, summaries, ideas, translations, or starting points	Repeated exposure and perceived usefulness
Heuristic credibility attribution	Appearance-based evaluation	"The answer is detailed, coherent, and probably correct"	Accepting well-structured outputs; limited checking when answers appear familiar	Accumulation of satisfactory experiences or confirmation with prior knowledge
Provisional reliance	Experience-based confidence	"It has usually worked, so I can rely on it for this type of task"	Expanding AI use; consulting it before conventional sources; delegating routine cognitive tasks	Encounter with inconsistency, consequential error, or source uncertainty
Epistemic friction	Disruption and doubt	"A convincing answer can still be wrong"	Re-reading output, questioning claims, noticing unsupported certainty, reconsidering previous reliance	Recognition of error, conflicting information, or inability to verify provenance
Active verification and calibration	Evidence-sensitive evaluation	"I can use it, but the level of checking depends on the claim and task"	Cross-checking, citation verification, prompt comparison, human consultation, risk-based boundaries	Development of verification routines and clearer understanding of AI limitations
Conditional integration	Regulated epistemic delegation	"AI is useful within defined limits, but responsibility for judgment remains mine"	Using AI selectively; differentiating low- and high-risk tasks; requiring evidence for consequential claims	Repeated reflective practice and explicit boundary formation
Divergent longer-term trajectory	Reflective augmentation or dependency	"AI supports my thinking" versus "I struggle to work without it"	Either integrating AI into independent reasoning or increasingly replacing reading, recall, and problem solving with AI	Strength or weakness of metacognitive monitoring, educational guidance, and self-imposed limits

Table 4 demonstrates that epistemic trust emerged as a recursive process rather than as a fixed attitude formed at a single moment. Most participants entered AI use through pragmatic initiation, in which usefulness preceded questions about epistemic authority. Because the systems produced immediate, fluent, and individualized responses, participants frequently moved into heuristic credibility attribution, particularly when answers were consistent with what they already knew. Repeated successful interactions produced provisional reliance, in which students increasingly assumed that AI would perform adequately within familiar task categories. The pivotal stage was epistemic friction. This occurred when a generated response was demonstrably false, a cited reference could not be located, similar prompts produced contradictory answers, or students encountered a claim that conflicted with established knowledge. Epistemic friction altered the meaning of subsequent interactions because students could no longer interpret fluency itself as sufficient evidence of reliability. For many participants, this experience initiated active verification and calibration. The resulting trust was neither complete trust nor rejection; it was differentiated according to disciplinary familiarity, task stakes, availability of verification, and potential consequences of error. Conditional integration represented

Iranian Journal of Educational Sociology
E-ISSN: 2645-3460

the most developed form of this process. At this stage, participants assigned AI specific roles while maintaining personal responsibility for final epistemic judgment. The longer-term trajectory nevertheless diverged. Some students described increasingly reflective use, in which AI stimulated questions, comparisons, alternative explanations, and more efficient inquiry. Others described a gradual reduction in independent searching, memorization, reading, and problem solving. Thus, the same initial process could culminate either in augmented epistemic agency or in dependency, depending on whether convenience remained subordinate to metacognitive monitoring and verification.

Table 5

Strategies Used by Students to Regulate and Verify Epistemic Trust

Strategy	Operational form	Primary epistemic function	Circumstances in which it was most commonly used	Identified limitation
Cross-source corroboration	Comparing AI claims with textbooks, scholarly articles, trusted websites, lecture materials, or databases	Tests whether a claim is supported independently of the AI response	Factual, theoretical, historical, and research-related claims	Requires time and access to reliable sources
Provenance checking	Locating cited references, checking authors, titles, publication details, quotations, and original documents	Evaluates whether the evidential basis of the response actually exists	Academic writing, literature review, thesis work, and citation-dependent tasks	AI-generated citation details may be partially correct and therefore difficult to detect as unreliable
Prompt triangulation	Asking the same question using different formulations, requesting counterarguments, or instructing the system to identify uncertainty	Detects instability and exposes hidden uncertainty in apparently definitive responses	Complex explanations, controversial issues, and unfamiliar topics	Consistent outputs across prompts do not independently establish truth
Claim decomposition	Breaking a long response into individual factual or conceptual claims for separate evaluation	Prevents the overall coherence of a response from concealing isolated errors	Long explanations and multi-part answers	Can make verification cognitively demanding
Prior-knowledge comparison	Comparing generated content with concepts already learned or with disciplinary expectations	Uses the student's existing knowledge as an internal plausibility check	Familiar subject areas and examination preparation	Weak when the student has insufficient prior knowledge
Human consultation	Asking instructors, supervisors, classmates, or knowledgeable professionals to assess uncertain claims	Introduces accountable human expertise into the verification process	High-stakes, specialized, ambiguous, or conflicting information	Depends on access to knowledgeable individuals
Iterative challenge	Asking AI to justify its answer, identify limitations, provide alternatives, or reconsider possible errors	Converts passive reception into dialogical scrutiny	When an initial answer appears overconfident, incomplete, or questionable	The system may produce persuasive justification for an incorrect initial claim
Risk zoning	Classifying tasks as low-, moderate-, or high-risk before determining the extent of AI reliance	Aligns verification effort with potential consequences of error	Routine writing versus thesis, publication, technical, or consequential tasks	Students may underestimate the actual risk of apparently routine tasks
Delayed acceptance	Temporarily withholding judgment until external evidence is obtained	Prevents immediate conversion of plausibility into belief	Unfamiliar or consequential claims	Difficult under severe time pressure
Self-explanation	Restating the answer independently, reconstructing reasoning, or solving a problem without AI after receiving assistance	Tests whether information has been understood rather than merely received	Concept learning, examination preparation, and problem solving	Requires motivation and may be abandoned when convenience is prioritized

The verification practices summarized in Table 5 constituted the principal behavioral mechanisms through which students converted undifferentiated confidence into calibrated epistemic trust. Cross-source corroboration was one of the most important strategies because it shifted the basis of acceptance away from the internal characteristics of an AI answer and toward independent evidence. Provenance checking was especially salient in academic tasks involving references because several participants had learned that a formally plausible citation could be inaccurate, incomplete, or nonexistent. Prompt triangulation and iterative challenge were used to examine the stability of generated information, but participants who had developed more sophisticated evaluation practices recognized that agreement across repeated AI outputs did not itself constitute external confirmation. Claim decomposition represented a more analytical strategy in which students avoided evaluating an entire fluent paragraph as one unit and instead examined discrete propositions separately. Prior-knowledge comparison was highly efficient in familiar domains, but the interviews indicated a paradox: students most needed assistance in areas where their own knowledge was weakest,

Iranian Journal of Educational Sociology
E-ISSN: 2645-3460

yet these were precisely the areas in which their capacity to identify AI error was most limited. Human consultation therefore remained particularly important for specialized and high-stakes questions. Risk zoning emerged as a notable metacognitive strategy. Reflective participants did not apply the same threshold of trust to every AI-assisted activity; they allowed more extensive delegation for ideation, stylistic revision, or preliminary organization, while requiring stronger verification for scholarly claims, citations, conceptual definitions, thesis arguments, and specialized technical information. Self-explanation was the strategy most directly connected with educational agency because it enabled students to distinguish between obtaining an answer and actually understanding it. Across the interviews, the presence of multiple verification strategies rather than reliance on a single technique was associated with the most stable form of conditional trust.

Table 6

Consequences of Different Patterns of Epistemic Trust for Learning and Epistemic Agency

Consequence domain	Constructive consequence	Mechanism producing the constructive consequence	Potential problematic consequence	Mechanism producing the problematic consequence
Access to knowledge	Faster orientation to unfamiliar topics	AI provides immediate conceptual maps, terminology, examples, and preliminary explanations	Illusion of comprehensive knowledge	Rapid answers create a sense that sufficient inquiry has already occurred
Cognitive processing	Scaffolding of difficult concepts	Reformulation, examples, simplification, and iterative explanation support comprehension	Reduced productive struggle	Students bypass effortful reasoning that contributes to durable learning
Academic inquiry	More efficient generation of questions and search directions	AI helps identify concepts, keywords, alternative perspectives, and possible connections	Displacement of primary-source engagement	AI summaries become substitutes for reading original scholarly materials
Critical thinking	Increased comparison and interrogation when AI is treated as contestable	Students question claims, seek alternatives, and test explanations	Passive acceptance of authoritative-looking language	Fluency and confidence suppress skepticism
Metacognition	Greater awareness of what is known, unknown, and in need of verification	Students use discrepancies to identify knowledge gaps	Reduced monitoring of understanding	Students equate successful task completion with genuine comprehension
Academic writing	Support for organization, clarity, language, and revision	AI functions as a responsive drafting and feedback aid	Weakening of authorship and argument ownership	Students increasingly delegate formulation of claims and argumentative structure
Epistemic autonomy	Reflective augmentation	AI remains subordinate to the learner's judgment, evidence evaluation, and responsibility	Epistemic dependency	AI becomes the default source for deciding what is true, relevant, or worth saying
Tolerance of uncertainty	More explicit recognition that claims require degrees of confidence	Exposure to AI error encourages suspension of judgment	Artificial certainty	Definitive AI language removes visible markers of uncertainty
Educational relationships	More focused questions for teachers and peers	AI assists initial exploration before human discussion	Reduced human epistemic interaction	Students replace dialogue with instructors or classmates by private AI consultation
Long-term learning habits	Development of verification routines and digital epistemic literacy	Repeated critical use establishes durable checking practices	Habitual cognitive outsourcing	Convenience becomes normalized and independent searching, recall, and problem solving decline

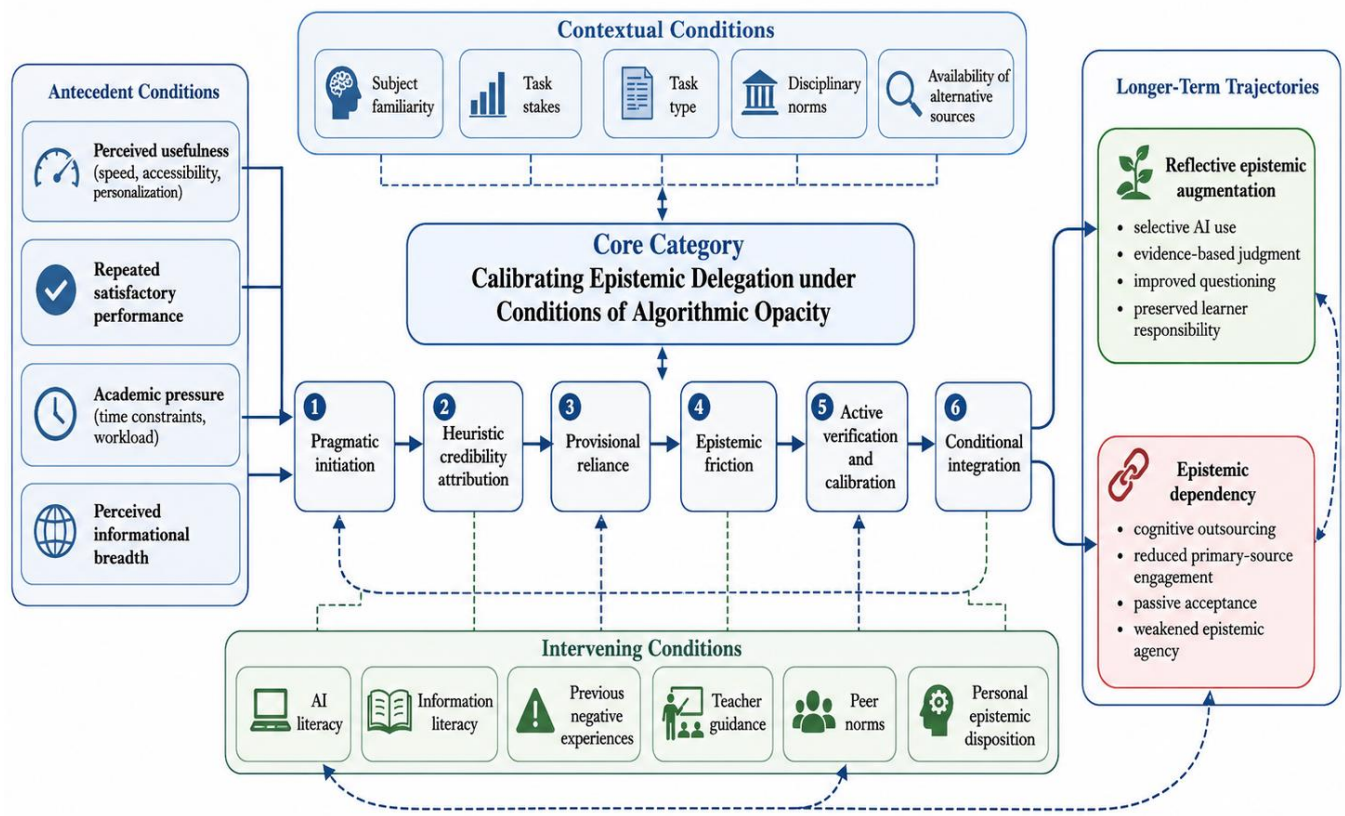
The consequences identified in Table 6 demonstrate the fundamentally ambivalent educational character of epistemic trust in AI-generated content. Participants did not describe AI as inherently beneficial or inherently harmful to learning; instead, consequences depended on the organization of trust. When AI was treated as a provisional and contestable resource, it expanded access to explanations, assisted orientation in unfamiliar domains, supported question generation, facilitated academic expression, and enabled students to compare different formulations of complex ideas. In these circumstances, AI could increase epistemic agency because students retained responsibility for deciding what constituted sufficient evidence and whether an answer deserved acceptance. The same affordances produced contrasting consequences when trust became insufficiently regulated. Immediate explanations could create an illusion of mastery, summarized information could displace engagement with original texts, and fluent prose could obscure the difference between persuasive expression and evidential justification. Participants also

Iranian Journal of Educational Sociology
E-ISSN: 2645-3460

described forms of productive intellectual effort that could gradually be bypassed when AI was consulted too early in a task. This included independently recalling information, struggling with difficult concepts, formulating an initial argument, searching for primary sources, and tolerating temporary uncertainty. The central distinction was therefore not between AI use and non-use but between reflective augmentation and cognitive outsourcing. In the former configuration, AI extended the learner's capacity while leaving epistemic responsibility with the student. In the latter, the student progressively delegated not only production of information but also judgments concerning relevance, validity, and sufficiency. This distinction formed the principal educational consequence of the emerging grounded theory.

Figure 1

Grounded Theory Model of University Students' Epistemic Trust in AI-Generated Content: Calibrating Epistemic Delegation Under Conditions of Algorithmic Opacity



Epistemic trust emerges as a cyclical, condition-dependent process shaped by utility, friction, verification, and educational context.

The integrated model represented in Figure 1 explains epistemic trust as a cyclical and condition-dependent process organized around the core category of calibrating epistemic delegation under conditions of algorithmic opacity. The model begins with pragmatic entry into AI-mediated academic activity, driven primarily by speed, accessibility, personalization, perceived competence, and academic workload. These functional advantages expose students repeatedly to AI-generated outputs and create the conditions for heuristic credibility attribution. At this point, linguistic fluency, coherence, completeness, responsiveness, and compatibility with prior knowledge function as provisional signals of credibility. Repeated satisfactory experiences may then stabilize into provisional reliance and broaden the range of tasks delegated to AI. However, the process is interrupted when epistemic friction occurs through factual error, fabricated evidence, inconsistency, inappropriate disciplinary reasoning, or inability to determine the provenance of a claim. Epistemic friction activates verification only when students possess or can access

Iranian Journal of Educational Sociology
E-ISSN: 2645-3460

sufficient metacognitive, informational, disciplinary, or social resources. Verification practices then permit recalibration of trust according to task stakes, familiarity, evidential accessibility, and possible consequences of error. The resulting form of trust is therefore conditional rather than global. Students may trust AI for specific functions while withholding authority in others. Teacher guidance, peer norms, disciplinary culture, institutional expectations, AI literacy, and previous experiences operate throughout the model as contextual and intervening influences. Over repeated cycles, the process leads toward one of two broad trajectories. Reflective epistemic augmentation develops when students preserve a distinction between assistance and authority, maintain verification routines, and regard themselves as responsible for final epistemic judgment. Epistemic dependency develops when repeated convenience lowers verification, AI becomes the default source of explanation and evaluation, and students progressively transfer responsibility for determining what is credible, relevant, and sufficient to the system.

4. Discussion and Conclusion

The present study sought to explain the process through which university students develop, evaluate, recalibrate, and regulate epistemic trust in AI-generated content from the perspective of philosophy of education. The grounded theory analysis indicated that students' epistemic trust was not a stable attitude or a simple state of acceptance versus rejection. Rather, it emerged as a dynamic, recursive, and context-dependent process organized around the core category of calibrating epistemic delegation under conditions of algorithmic opacity. Six major categories structured this process: pragmatic entry into AI-mediated knowing, heuristic attribution of epistemic credibility, epistemic friction and disruption of trust, verification and calibration of trust, educational and social regulation of epistemic trust, and consequences for epistemic agency. The findings further showed a processual movement from pragmatic initiation and heuristic credibility attribution toward provisional reliance, followed by epistemic friction, active verification and calibration, and conditional integration. Depending on the quality of metacognitive regulation, disciplinary knowledge, AI literacy, educational guidance, and verification practices, this process could culminate in either reflective epistemic augmentation or epistemic dependency. Thus, the principal finding of the study is that educationally appropriate trust in generative AI is neither unconditional confidence nor generalized suspicion, but a form of calibrated, task-sensitive, and revisable reliance in which responsibility for final epistemic judgment remains with the learner.

The first important finding concerned pragmatic entry into AI-mediated knowing. Participants generally did not begin using generative AI because they had first established that it was epistemically reliable. Instead, they were attracted by speed, ease of access, personalized explanation, responsiveness, and the reduction of time and effort required for academic tasks. In other words, utility preceded epistemic evaluation. This finding is consistent with research showing that students' attitudes toward AI-based solutions are influenced substantially by perceived usefulness and expected benefits (Le et al., 2022). It also corresponds with broader models of AI adoption in which perceived practical value, technological capacity, and contextual facilitators shape initial engagement with intelligent systems (Fulton et al., 2022). Studies of students' experiences with ChatGPT similarly show that learners value immediate explanation, availability, and the capacity of generative systems to function as virtual tutors or academic assistants (Ding et al., 2023). From the perspective of philosophy of education, however, the significance of this finding lies in the possibility that pragmatic usefulness can be gradually misinterpreted as epistemic authority. A system that repeatedly provides convenient assistance may come to be regarded not merely as useful but as knowledgeable. The distinction between "this tool helps me" and "this source deserves belief" is therefore crucial. The present findings suggest that the first step toward epistemic overreliance may occur when functional success is generalized into assumptions about truthfulness, expertise, or evidential adequacy.

The second category, heuristic attribution of epistemic credibility, demonstrated that students frequently evaluated AI-generated content through accessible surface cues, especially fluency, coherence, completeness, confidence, and consistency with their prior knowledge. These characteristics often functioned as proxies for epistemic quality when direct verification was difficult. This finding aligns with evidence that perceived anthropomorphic characteristics can affect trust in intelligent recommendation systems and influence users' confidence in the information provided (Wang et al., 2024). Generative AI is particularly powerful in this regard because it communicates through fluent, conversational language that resembles competent human explanation. The finding also corresponds with

research demonstrating that trust in AI is shaped by visible or perceived signals of reliability and legitimacy (Almeida et al., 2026). Yet the present study extends this literature by showing that the educational problem is not merely whether users perceive AI as trustworthy, but whether they mistake communicative competence for epistemic competence. This distinction is especially relevant because trustworthiness research emphasizes transparency, explainability, reliability, and accountability as dimensions that cannot be inferred from polished language alone (Zhang et al., 2024). Similarly, the literature on explainable AI highlights the importance of making the basis of algorithmic decisions or outputs more interpretable to users (Mohammed, 2023). The findings therefore suggest that students require explicit awareness that linguistic confidence is not evidence and that coherence should be treated as a cue for further examination rather than as a sufficient basis for belief.

A particularly significant finding involved epistemic friction, which emerged when students encountered factual inaccuracies, fabricated references, inconsistent answers, disciplinary errors, or uncertainty regarding the origin of generated claims. Such experiences frequently destabilized generalized trust and led participants to recognize that a highly plausible response could nevertheless be incorrect. This result is compatible with recent work describing students' movement from initial trust toward more critical reconstruction after encountering the limitations of generative AI (Zhou, 2026). The present study adds an important processual interpretation: error does not only weaken trust; under favorable educational conditions, it can become a productive epistemic event. By exposing the limits of AI-generated content, error can stimulate doubt, comparison, source checking, and renewed awareness of personal responsibility for judgment. This interpretation is also consistent with discussions of trust in AI that emphasize the need for appropriately calibrated rather than maximal trust (Çetinkaya, 2025). In educational terms, epistemic friction may therefore constitute an important condition for intellectual maturation. If students never experience or recognize AI errors, they may continue to equate confidence with correctness. When error becomes visible, however, learners may begin to differentiate the appearance of knowledge from justified knowledge. The educational challenge is to ensure that such moments lead to critical reconstruction rather than simple rejection or continued uncritical use.

The fourth major finding, verification and calibration of trust, represented the central behavioral response to epistemic uncertainty. Participants employed strategies such as cross-source corroboration, provenance checking, iterative prompting, claim decomposition, comparison with prior knowledge, consultation with instructors or peers, risk zoning, delayed acceptance, and self-explanation. These practices transformed trust from a generalized orientation into a conditional judgment based on task, stakes, available evidence, and personal competence. This finding strongly aligns with research on trustworthy automated feedback, which emphasizes that AI-supported educational guidance should be interpreted within frameworks that preserve transparency and enable users to evaluate the reliability of feedback rather than accepting it automatically (Lin et al., 2022). It also resonates with entrustment frameworks proposed for AI in professional education, where the key issue is not whether AI is globally trustworthy but which activities may appropriately be delegated to it and under what level of human oversight (Gin et al., 2024). The same logic has been applied to algorithmic assessment, where validity and trust depend on whether automated processes can be justified and monitored in relation to the consequences of decisions (Aloisi, 2023). The present findings suggest that students independently develop a similar logic of epistemic entrustment: they permit greater AI involvement in low-risk activities such as brainstorming or language refinement but require stronger verification for citations, theoretical arguments, specialized claims, or consequential academic decisions. This form of calibrated trust appears to be more educationally defensible than either complete reliance or indiscriminate distrust.

Prior knowledge emerged as a particularly important condition influencing calibration. Students were better able to evaluate AI-generated content when they already possessed enough disciplinary knowledge to identify conceptual errors, inappropriate terminology, contradictions, or implausible claims. Conversely, unfamiliar subjects increased dependence on surface credibility cues. This finding is consistent with research indicating that knowledge and prior experience shape attitudes toward and use of AI-assisted systems in specialized domains (Lin et al., 2023). It also corresponds with broader findings that public trust in AI varies according to familiarity, perceived understanding, and perceptions of risk and benefit (Syed et al., 2024). The paradox revealed by the current study is educationally important: students are often most likely to seek AI assistance when their knowledge is weakest, yet low prior knowledge simultaneously reduces their capacity to evaluate the quality of the assistance they

Iranian Journal of Educational Sociology
E-ISSN: 2645-3460

receive. Therefore, generative AI may be least epistemically safe precisely when it feels most useful. From a philosophical perspective, this reinforces the distinction between access to information and possession of knowledge. Receiving an explanation does not necessarily place a learner in a position to justify its claims. For this reason, pedagogical strategies should not assume that AI can replace foundational disciplinary knowledge; rather, prior knowledge remains a necessary condition for the responsible evaluation of AI-generated information.

The fifth category concerned the educational and social regulation of epistemic trust. Participants' trust was influenced by instructors, peer practices, disciplinary expectations, institutional rules, and assessment structures. This finding indicates that epistemic trust in AI is not constructed solely within an individual student–machine relationship but is situated within a broader educational ecology. Student-centered analyses of generative AI use in higher education likewise suggest that human–AI collaboration is shaped by institutional and pedagogical conditions and is most productive when learners remain active participants in evaluating AI contributions (Razmerita, 2024). Research on faculty and student perceptions further demonstrates that views of generative AI are strongly connected to concerns about academic integrity, acceptable use, educational benefit, and uncertainty regarding appropriate boundaries (Petricini et al., 2023). Similarly, student perceptions of ChatGPT in academic writing reveal that trust is closely intertwined with expectations concerning learning and assessment (Tossell et al., 2024). The ethical literature reinforces the importance of these contextual influences by emphasizing that trustworthy educational AI requires transparency, accountability, fairness, and explicit attention to human values (Smyrnaiou et al., 2023; Yan & Chau, 2025). The present findings therefore indicate that universities themselves participate in producing either reflective or unreflective forms of trust. Where instruction emphasizes only prohibition, students may continue using AI without developing critical evaluative skills. Where unrestricted use is normalized, dependency may be inadvertently encouraged. More constructive educational environments are those in which appropriate forms of delegation, verification, disclosure, and responsibility are explicitly taught.

The final major category addressed consequences for epistemic agency. The analysis identified two broad longer-term trajectories: reflective epistemic augmentation and epistemic dependency. Reflective augmentation occurred when students used AI to expand inquiry, obtain alternative explanations, generate questions, improve academic expression, or initiate exploration while preserving independent evaluation and verification. Epistemic dependency emerged when AI progressively replaced primary-source reading, independent reasoning, problem solving, recall, or the learner's responsibility for determining what counted as credible information. This duality is consistent with evidence that generative AI can support learning while simultaneously producing concerns about overreliance and reduced engagement with academic tasks (Ding et al., 2023; Tossell et al., 2024). Research on feedback also suggests that the educational effect of AI assistance depends substantially on perceived quality and how learners interact with feedback rather than merely on whether feedback is provided by a human or AI source (Ruwe & Kuklick, 2025). The present study extends this insight by framing the issue as one of epistemic agency. The decisive question is whether AI remains subordinate to the student's processes of judgment or becomes a substitute for them.

This distinction is directly related to emerging conceptions of trustworthy human–AI learning environments. Bidirectional human–AI alignment emphasizes that educational systems should support human goals while learners develop the competencies required to interact with AI responsibly (Shen, 2025). Ethical frameworks for learning technologies similarly emphasize that AI integration must preserve human oversight, transparency, and responsibility (Alwahaby & Cukurova, 2024). Trustworthy AI architectures for educational environments likewise foreground the need for reliability and safeguards that prevent technological capability from being confused with unrestricted legitimacy (Abdullakutty & Qayyum, 2024). The current grounded theory situates these principles within students' everyday academic experiences. Reflective epistemic augmentation occurs when learners maintain the right and responsibility to challenge AI, seek contrary evidence, recognize uncertainty, and reject generated content. Dependency occurs when these epistemic functions are progressively transferred to the system.

The core category, calibrating epistemic delegation under conditions of algorithmic opacity, integrates these findings and provides the principal theoretical contribution of the study. Generative AI introduces a distinctive educational problem because students interact with a system capable of producing convincing knowledge-like content without

Iranian Journal of Educational Sociology
E-ISSN: 2645-3460

offering the same structures of authorship, accountability, provenance, and justification associated with conventional academic sources. Concerns about algorithmic interpretability in educational systems have similarly emphasized that opacity can undermine responsible trust and necessitates mechanisms of explanation and regulation (Liu & Cheng, 2022). More broadly, current discussions of generative AI have argued that these technologies require reconsideration of established assumptions about knowledge production, research, authorship, and ethics (Machin-Mastromatteo, 2025). The present findings suggest that the appropriate educational response is not to treat AI either as an autonomous epistemic authority or merely as a neutral tool. Rather, AI should be conceptualized as an epistemically consequential assistant whose outputs must be situated within human practices of justification. Students may delegate certain cognitive operations to AI, but epistemic responsibility for deciding what should be believed, cited, submitted, or acted upon cannot be delegated in the same way.

The findings also help explain the frequently observed gap between users' expectations of AI and its actual capabilities. Research has demonstrated that perceptions and expectations surrounding AI may diverge substantially, with important implications for educational adoption and responsible use (Skenderi et al., 2024). The present theory suggests that such a gap becomes especially problematic when repeated convenience and successful low-risk interactions create generalized trust across tasks. The broader expansion of AI research and infrastructure makes the cultivation of trustworthy use increasingly important at both institutional and societal levels (Donlon, 2024). Accordingly, universities need to move beyond the question of whether students should use generative AI and focus instead on the quality of epistemic relationships students construct with it. The central educational objective should be the cultivation of learners capable of benefiting from AI without surrendering intellectual autonomy, evidential responsibility, and critical judgment.

Several limitations should be considered when interpreting the findings of this study. First, the participants were university students studying in Tehran, and the social, institutional, cultural, and technological characteristics of this setting may have influenced their experiences and interpretations of generative AI. Consequently, the grounded theory should not be assumed to represent students in all educational systems or cultural contexts. Second, participation required previous academic use of generative AI, which may have resulted in the inclusion of students who were more technologically engaged than the broader student population. Third, the findings were based primarily on self-reported interview data and therefore reflected participants' descriptions of their practices rather than direct observation of their interactions with AI systems. Students' reported verification behaviors may not always correspond exactly to what they do during actual academic tasks. Fourth, generative AI technologies are developing rapidly, and changes in model accuracy, source transparency, interface design, institutional policy, or access to integrated verification tools may alter patterns of epistemic trust over time. Finally, although theoretical sampling and constant comparison were used to develop conceptual depth, the resulting model represents an interpretive grounded theory of the participants' experiences rather than a universally deterministic sequence through which all students necessarily progress.

Future studies should test, extend, and refine the proposed grounded theory across different universities, countries, academic disciplines, and educational levels. Comparative qualitative studies could investigate whether students in fields characterized by different standards of evidence develop distinct thresholds for epistemic trust, while longitudinal research could examine how trust changes over several semesters of repeated AI use. Mixed-method research may operationalize the categories identified in this study and examine relationships among prior knowledge, AI literacy, task stakes, verification behaviors, epistemic agency, and dependency in larger samples. Experimental studies could manipulate factors such as AI confidence, source visibility, explanation quality, citation availability, or teacher guidance to determine how these variables affect students' willingness to verify generated content. Future research should also distinguish more carefully between functional trust, epistemic trust, and behavioral reliance, because students may use AI extensively without believing all of its claims or may accept claims despite relatively infrequent use. Direct observation, screen-recording, think-aloud protocols, and analysis of actual student–AI interactions would also help determine how epistemic calibration unfolds moment by moment rather than relying exclusively on retrospective accounts.

Universities should develop educational practices that treat epistemic trust in AI as an explicit component of academic and digital literacy. Students should be taught to distinguish fluent presentation from evidential reliability, verify references and factual claims, compare generated

information with primary and authoritative sources, identify task-specific risk, and recognize situations in which human expertise is required. Instructors can design assignments that require students to document how AI-generated content was evaluated, identify claims that required verification, explain why particular sources were accepted, and reflect on the limitations of AI assistance. Assessment practices should reward reasoning, source evaluation, and the ability to defend conclusions rather than only the quality of the final product. Institutions should also provide clear but flexible guidance that differentiates acceptable cognitive assistance from inappropriate substitution of academic judgment. The practical objective should not be the elimination of AI from learning, but the development of students who can use it selectively and productively while retaining responsibility for understanding, justification, authorship, and final epistemic decisions.

Authors' Contributions

Authors equally contributed to this article.

Declaration

In order to correct and improve the academic writing of our paper, we have used the language model ChatGPT.

Transparency Statement

Data are available for research purposes upon reasonable request to the corresponding author.

Acknowledgments

We hereby thank all participants for agreeing to record the interview and participate in the research.

Declaration of Interest

The authors report no conflict of interest.

Funding

According to the authors, this article has no financial support.

Ethical Considerations

All procedures performed in studies involving human participants were under the ethical standards of the institutional and, or national research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards.

References

Abdullakutty, F., & Qayyum, A. (2024). Trustworthy AI for Educational Metaverses. https://doi.org/10.36227/techrxiv.170775108.81506629/v1

Almeida, F. d., Glaum, J., Hildred, K., & Scott, I. J. (2026). Beyond Black Boxes: AI Certification Labels Increase Adoption Through Trust. *Psychology and Marketing*, *43*(9), 2464-2482. https://doi.org/10.1002/mar.70185

Aloisi, C. (2023). The Future of Standardised Assessment: Validity and Trust in Algorithms for Assessment and Scoring. *European Journal of Education*, *58*(1), 98-110. https://doi.org/10.1111/ejed.12542

Alwahaby, H., & Cukurova, M. (2024). Navigating the Ethical Landscape of Multimodal Learning Analytics: A Guiding Framework. https://doi.org/10.35542/osf.io/adxuq

Arif, A. (2023). Investigating Students' Attitudes &Amp; Trust in AI During COVID-19. *Journal of Student Research*, *12*(3). https://doi.org/10.47611/jsrhs.v12i3.5015

Çetinkaya, L. (2025). When I Say … Trust in AI. *Medical Education*, *60*(5), 490-491. https://doi.org/10.1111/medu.70118

Ding, L., Li, T., Jiang, S., & Gapud, A. A. (2023). Students' Perceptions of Using ChatGPT in a Physics Class as a Virtual Tutor. *International Journal of Educational Technology in Higher Education*, *20*(1). https://doi.org/10.1186/s41239-023-00434-1

Donlon, J. (2024). The National Artificial Intelligence Research Institutes Program and Its Significance to a Prosperous Future. *Ai Magazine*, *45*(1), 6-14. https://doi.org/10.1002/aaai.12153

Fulton, R., Fulton, D., & Kaplan, S. (2022). The Framework of Artificial Intelligence (FAI): Driving Triggers, State of the Art Over Time and Industry Adoption Influencers. *International Journal of Artificial Intelligence & Applications*, *13*(3), 85-108. https://doi.org/10.5121/ijaia.2022.13307

Gin, B. C., O'Sullivan, P., Hauer, K. E., Abdulnour, R.-E. E., Mackenzie, M. E., Cate, O. t., & Boscardin, C. (2024). Entrustment and EPAs for Artificial Intelligence (AI): A Framework to Safeguard the Use of AI in Health Professions Education. *Academic Medicine*, *100*(3), 264-272. https://doi.org/10.1097/acm.0000000000005930

Le, A.-B., Bui, M.-T., & Le, B.-D. (2022). Factors Influence Students' Attitudes Toward AI-based Innovative Solutions. *34*, 21-26. https://doi.org/10.15439/2022m9228

Lin, J., Sha, L., Li, Y., Gašević, D., & Chen, G. (2022). Establishing Trustworthy Artificial Intelligence in Automated Feedback. https://doi.org/10.35542/osf.io/5efxn

Lin, L., Tang, B., Cao, L., Yan, J., Zhao, T., Hua, F., & He, H. (2023). The Knowledge, Experience, and Attitude on Artificial Intelligence-Assisted Cephalometric Analysis: Survey of Orthodontists and Orthodontic Students. *American Journal of Orthodontics and Dentofacial Orthopedics*, *164*(4), e97-e105. https://doi.org/10.1016/j.ajodo.2023.07.006

Liu, L., & Cheng, X. (2022). Reasonable Construction and Legal Regulation of Interpretable Mechanism of Educational Artificial Intelligence Algorithms. 233-240. https://doi.org/10.2991/978-2-494069-05-3_30

Machin-Mastromatteo, J. D. (2025). Artificial Intelligence Encounters of the Generative Kind: Rethinking Knowledge, Ethics, and Research. *Information Development*, *41*(3), 549-558. https://doi.org/10.1177/02666669251358029

Iranian Journal of Educational Sociology
E-ISSN: 2645-3460

Mohammed, B. (2023). A Review on Explainable Artificial Intelligence Methods, Applications, and Challenges. *Indonesian Journal of Electrical Engineering and Informatics (Ijeei)*, *11*(4). https://doi.org/10.52549/ijeei.v11i4.5151

Petricini, T., Wu, C., & Zipf, S. T. (2023). Perceptions About Generative AI and ChatGPT Use by Faculty and College Students. https://doi.org/10.35542/osf.io/jyma4

Razmerita, L. (2024). Human-Ai Collaboration: A Student-Centered Perspective of Generative AI Use in Higher Education. *European Conference on E-Learning*, *23*(1), 320-329. https://doi.org/10.34190/ecel.23.1.3008

Ruwe, T., & Kuklick, L. (2025). Quality Counts? Examining the Role of Feedback Provider and Feedback Quality on Students' Feedback Perceptions. *British Journal of Educational Technology*, *57*(1), 272-298. https://doi.org/10.1111/bjet.70011

Shen, H. (2025). Bidirectional Human-Ai Alignment in Education for Trustworthy Learning Environments. https://doi.org/10.48550/arxiv.2512.21552

Skenderi, B., Zejnullahu, S., Skenderi, D., & Selimi, F. (2024). Understanding Gap Between Perception and Expectations for Artificial Intelligence: Implications for Sustainable Development Goals 4 and 9. *Edelweiss Applied Science and Technology*, *8*(6), 4611-4623. https://doi.org/10.55214/25768484.v8i6.2992

Smyrnaiou, Z., Liapakis, A., & Bougia, A. (2023). Ethical Use of Artificial Intelligence and New Technologies in Education 5.0. *Journal of Artificial Intelligence Machine Learning and Data Science*, *1*(4), 119-124. https://doi.org/10.51219/jaimld/anastasios-liapakis/15

Syed, W., Babelghaith, S. D., & Al-Arifi, M. N. (2024). Assessment of Saudi Public Perceptions and Opinions Towards Artificial Intelligence in Health Care. *Medicina*, *60*(6), 938. https://doi.org/10.3390/medicina60060938

Tossell, C. C., Tenhundfeld, N. L., Momen, A., Cooley, K., & Visser, E. J. d. (2024). Student Perceptions of ChatGPT Use in a College Essay Assignment: Implications for Learning, Grading, and Trust in Artificial Intelligence. *Ieee Transactions on Learning Technologies*, *17*, 1069-1081. https://doi.org/10.1109/tlt.2024.3355015

Wang, Y., Liu, W., & Yao, M. (2024). Which Recommendation System Do You Trust the Most? Exploring the Impact of Perceived Anthropomorphism on Recommendation System Trust, Choice Confidence, and Information Disclosure. *New Media & Society*, *27*(6), 3264-3292. https://doi.org/10.1177/14614448231223517

Yan, Y., & Chau, T. (2025). A Systematic Review of AI Ethics in Education. *Journal of Global Information Management*, *33*(1), 1-50. https://doi.org/10.4018/jgim.386381

Zhang, Z., Deng, W., Wang, Y., & Qi, C. (2024). Visual Analysis of Trustworthiness Studies: Based on the Web of Science Database. *Frontiers in psychology*, *15*. https://doi.org/10.3389/fpsyg.2024.1351425

Zhou, C. (2026). From Initial Trust to Critical Reconstruction: Upper Secondary Students' Engagement With Generative AI in Science Learning. *Frontiers in psychology*, *17*. https://doi.org/10.3389/fpsyg.2026.1830646

Iranian Journal of Educational Sociology
E-ISSN: 2645-3460