http://dx.doi.org/10.18623/rvd.v22.n4.3741

BETWEEN TRUST AND UNCERTAINTY: HOW STUDENTS CONSTRUCT ETHICAL BOUNDARIES IN AI-DRIVEN LEARNING

ENTRE CONFIANÇA E INCERTEZA: COMO OS ALUNOS CONSTRÓEM LIMITES ÉTICOS NA APRENDIZAGEM ORIENTADA POR IA

Article received on: 7/21/2025
Article accepted on: 10/27/2025

Omar Alobud*
*Assistant Professor, College of Science and Health Professions, King Saud bin Abdulaziz University for Health Sciences, King Abdullah International Medical Research Center (KAIMRC), Ministry of National Guard - Health Affairs, Saudi Arab
Orcid: https://orcid.org/0009-0004-7541-5873
obudo@ksau-hs.edu.sa

The authors declare that there is no conflict of interest

Abstract
As artificial intelligence becomes more embedded in academic settings, questions of trust, authorship, and ethical responsibility become increasingly urgent—especially in contexts where institutional policy is vague or absent. This qualitative study explores how undergraduate students interpret and navigate the ethical use of AI tools in learning environments lacking clear guidelines. Twelve students participated in semi-structured interviews focused on their perceptions of fairness, authorship, and moral boundaries when engaging with AI technologies. Thematic analysis revealed three key patterns shaping students' trust and use of AI: reading institutional signals to interpret what is implicitly allowed, managing emotional risks such as guilt or anxiety in the absence of policy clarity, and maintaining personal authority over their academic work despite AI involvement. Rather than relying solely on rules, students constructed their own frameworks for responsible use—often guided by emotional cues, peer discussion, and personal values. These findings suggest that ethical AI literacy requires more than technical competence; it demands shared dialogue, emotional safety, and participatory policy-making. The study calls for institutions to move beyond compliance models and engage students as co-authors of ethical practice in AI-augmented education.

Keyword: AI in Education. Academic Trust. Student Ethics. Authorship. Emotional Burden.

Resumo
À medida que a inteligência artificial se torna mais presente em ambientes acadêmicos, questões de confiança, autoria e responsabilidade ética tornam-se cada vez mais urgentes — especialmente em contextos onde as políticas institucionais são vagas ou inexistentes. Este estudo qualitativo explora como estudantes de graduação interpretam e lidam com o uso ético de ferramentas de IA em ambientes de aprendizagem que carecem de diretrizes claras. Doze estudantes participaram de entrevistas semiestruturadas focadas em suas percepções de justiça, autoria e limites morais ao interagirem com tecnologias de IA. A análise temática revelou três padrões principais que moldam a confiança e o uso da IA pelos estudantes: a leitura de sinais institucionais para interpretar o que é implicitamente permitido, o gerenciamento de riscos emocionais, como culpa ou ansiedade, na ausência de clareza nas políticas, e a manutenção da autoridade pessoal sobre seu trabalho acadêmico, apesar do envolvimento da IA. Em vez de se basearem apenas em regras, os estudantes construíram suas próprias estruturas para o uso responsável — frequentemente guiados por pistas emocionais, discussões com colegas e valores pessoais. Essas descobertas sugerem que a alfabetização ética em IA requer mais do que competência técnica; ela exige diálogo compartilhado, segurança emocional e formulação participativa de políticas. O estudo defende que as instituições vão além dos modelos de conformidade e envolvam os estudantes como coautores da prática ética na educação aumentada por IA.

Palavras-chave: *IA na Educação. Confiança Acadêmica. Ética Estudantil. Autoria. Sobrecarga Emocional.*

1 INTRODUCTION

Artificial intelligence language models have rapidly moved from experimental tools to everyday companions in academic life. As platforms like ChatGPT and GPT-4 become widely accessible, many students now rely on them to generate ideas, clarify confusing concepts, or draft early versions of assignments. The appeal is obvious: AI can streamline learning, offer instant feedback, and adapt explanations to individual needs (Kasneci et al., 2023). Educational institutions are also exploring these tools for use in course design, content development, and teaching support (Ding et al., 2023; Liu et al., 2023). As a result, higher education is undergoing a quiet shift in how knowledge is accessed, produced, and assessed. Yet with this shift comes a growing awareness that the use of AI in learning is not purely technical. Especially in academic environments, where questions of authorship, fairness, and evaluation are central, the presence of AI prompts deeper concerns—not about what it can do, but whether it should be trusted.

To understand these concerns, it is necessary to move beyond conversations about functionality and focus instead on the experience of trust. Much of the discourse around AI emphasizes accuracy, efficiency, and scale, yet these factors alone do not determine whether a student will actually use or depend on the tool. Trust, unlike performance, is an internal and psychological judgment. It involves a willingness to rely on a system in situations where outcomes are uncertain, and consequences may be personal or serious. Research in psychology helps explain this process. According to Kahneman and Egan's (2011) dual-process theory, individuals make decisions using both fast, intuitive reactions (System 1) and slower, deliberative reasoning (System 2). In the context of AI tools, students may experience immediate emotional reactions—comfort, suspicion, anxiety—before consciously evaluating the tool's usefulness or appropriateness. Lazarus's (1991) model further shows that emotions are closely tied to appraisals of personal relevance and risk. These theories suggest that trust in AI is not simply about its capabilities, but about how students interpret its role in their academic lives—how they feel about using it, what

risks they perceive, and whether they believe the tool supports or undermines their learning integrity.

These psychological processes are rarely isolated; they are embedded within institutional and cultural contexts that also shape trust. When students engage with AI tools in academic settings, they are not just assessing the system—they are reading signals from their instructors, institutions, and peers. Trust in the tool often depends on whether it is endorsed by the university, included in coursework, or discussed openly in class. In the absence of clear guidance, students may feel uncertain about what is allowed, what counts as misuse, or who is accountable for the outcome of AI-assisted work. These uncertainties can generate emotional stress, particularly when the stakes involve plagiarism policies, assignment evaluation, or grading fairness (Cotton et al., 2024). Other students worry about equity—questioning whether AI tools reflect Western norms, marginalize nonstandard writing, or reinforce educational biases (Mutimukwe et al., 2022; Lai et al., 2023). In these cases, trust is not just a question of whether the system works—it becomes a question of whether the educational environment feels safe, fair, and supportive of responsible experimentation. Even when the AI performs well, students may hesitate to use it if they feel that its ethical boundaries, legal risks, or institutional policies are unclear.

Despite the growing integration of AI into education, few studies have directly explored how students make sense of trust in these systems. Much of the existing research focuses on important but distinct concerns—such as privacy, data leakage, or algorithmic harm. For example, Das et al. (2024) and Smith et al. (2023) highlight the technical vulnerabilities of large language models and call for more robust data protection protocols. Liu et al. (2023) emphasize the institutional challenge of integrating AI ethically, while reviews like Qin et al. (2024) examine broad ethical frameworks and governance models. While valuable, these studies rarely center the user's psychological process—how trust is built, withheld, or negotiated in everyday learning decisions. Yet this lens is crucial. If students do not trust the system—or the environment surrounding its use—they may avoid it altogether, use it in risky ways, or experience emotional conflict despite its benefits. Understanding trust as a lived, emotional, and cognitive experience can offer important insights for improving both the design of AI tools and the structures in which they are used.

To address this gap, the present study explores how students assess and experience trust in AI language models when using them for academic purposes. Rather than evaluating system performance, this study focuses on the human process of trust: what makes students feel confident in the tool, what causes hesitation, and how institutional factors affect their decisions. Drawing on in-depth interviews with university students, the research investigates the emotional and cognitive dimensions of trust, as well as the institutional cues that shape it. Specifically, it asks: "How do students assess and experience trust in AI language models used for educational purposes, and what factors influence their willingness to rely on these tools?" By framing trust as a student-centered construct, this study contributes to ongoing conversations about ethical AI integration in education—not by adding more rules, but by understanding how students make sense of the systems they are being asked to trust.

2 METHOD

This study employed a qualitative, exploratory design to understand how university students experience and assess trust in AI language models used for academic tasks. Given the complex and subjective nature of trust, a qualitative approach was chosen to allow for in-depth exploration of participants' personal perspectives and decision-making processes. Twelve university students were recruited through purposive sampling based on one inclusion criterion: prior use of an AI language model (e.g., ChatGPT, GPT-4) in their academic work. Participants came from a variety of disciplines and levels of study to ensure diverse perspectives. All had used AI tools for learning-related tasks such as summarizing articles, generating writing ideas, editing text, or solving problems.

Data were collected through semi-structured, one-on-one interviews. Each interview lasted approximately 30 to 45 minutes and was conducted in a private, quiet setting. The interview guide consisted of open-ended questions aimed at eliciting participants' reflections on how they decide to trust or distrust AI tools, what factors influence their comfort or caution, and how institutional policies or classroom signals shape their behavior. Follow-up prompts were used flexibly to explore emerging insights. All interviews were audio-recorded with participant consent and transcribed verbatim for analysis. Thematic analysis was used to interpret the data. Transcripts were reviewed iteratively and coded inductively, focusing on meaning patterns related to emotional

responses, legal awareness, institutional influence, and personal decision-making. Codes were refined through constant comparison, and related codes were grouped to form overarching themes that captured shared experiences and differences across participants. To preserve credibility, analysis involved repeated cross-checking of themes against raw data and peer debriefing within the research team.

3 RESULTS

The findings revealed that students do not approach AI language models as neutral or purely functional tools. Instead, their trust is shaped by how they interpret academic expectations, how they emotionally respond to using the tool, and how much control they feel they can retain. Through their reflections, it became clear that trust was not a fixed belief in the system's ability, but a series of decisions made in context—sometimes cautious, sometimes confident, often shaped by the perceived legal and institutional environment. Three key patterns emerged from the interviews: reading institutional signals, managing emotional risks, and maintaining personal authority over the work.

Many students described the need to "read the room" when deciding whether or how to use AI tools. In the absence of clear university rules or professor instructions, they relied on subtle cues—phrases like "original work" or "independent thinking"—to infer what was acceptable. This uncertainty made some feel uneasy, even if their actual use of AI was limited to brainstorming or early drafts. One student shared that they used AI silently because they feared judgment, even though no one had told them it was wrong. This need to interpret unspoken rules became central to how trust in the system was formed—not just in the tool, but in the environment that surrounded it.

Beyond rules, many participants described an emotional weight tied to their use of AI. Even when the tool was used ethically, it sometimes produced feelings of guilt or discomfort. Students worried that others might question their integrity or misunderstand their intentions. Participants said they felt nervous submitting work, even after rewriting it in their own words, because they still saw AI as "crossing a line." These reactions weren't about the tool's output—they were about internal doubts and unclear external signals. In these moments, trust became fragile, shaped as much by anxiety as by reasoning. Still, students consistently emphasized the importance of staying in control. Even those who appreciated the tool's convenience were careful not to rely on it too

heavily. They described rewriting, editing, or discarding AI-generated suggestions to make sure the final work felt like their own. Trust, in this sense, was conditional. It wasn't a matter of "believing in" the tool—it was about believing in their ability to guide it. As one student put it, "I trust myself to decide when it helps and when it doesn't." This autonomy appeared to be a foundation of ethical use, helping students navigate both practical and moral uncertainties.

The findings show that trust in AI is not a simple yes-or-no judgment. It is built through ongoing negotiation—between policy and perception, between confidence and doubt. Students assess the tool, but they also assess the environment and themselves. Their trust depends not just on how well AI performs, but on how clearly academic institutions define its place, and how much space students feel they have to use it responsibly.

4 DISCUSSION

The study paints a clear but layered picture of how students build and adjust trust in AI tools in their academic lives. Instead of simply accepting or rejecting these systems, they moved through an evolving process shaped by what their university signaled, how they felt about using the tools, and their own sense of control. Trust wasn't a single decision. It shifted depending on risk, ethical comfort, and how clear—or unclear—the guidance from the institution was. From their accounts, three key themes stood out: reading institutional signals, handling emotional risks, and protecting their own authorship. Together, these themes reveal the human side of AI use in education when policies are vague or missing.

Students often spoke about the confusion they felt trying to guess what their institutions considered acceptable. In the absence of clear policies or instructor instructions, they relied on hints—like course syllabi or the tone in class—to make sense of what was allowed. These signals quietly shaped how and when they used AI, even when their use was cautious. As Pavlik (2023) notes, when institutions fail to set boundaries, the ethical weight falls on students, leaving them in a moral gray zone. This reflects concerns raised by UNESCO (2019), which called for structured guidance to support ethical AI use. Many students used AI quietly, tested it in private, or avoided it altogether. Their choices were not just technical—they were ethical negotiations. Jobin

et al. (2019) underline that trust in AI is built through institutional responsiveness, not just user awareness. NSW Government (2023) likewise stresses the need for transparent communication and governance in education. In this study, when that clarity was missing, students were left to "read between the lines."

The emotional weight behind these decisions was just as important. Even when students felt they were acting responsibly, they often felt uneasy or unsure. Their discomfort wasn't about technology itself but about unclear boundaries. Qin et al. (2024) point out that this kind of discomfort often comes from a lack of ethical clarity and emotional safety. Students described fear of being judged, guilt, or feeling like they had crossed an invisible line. Choung et al. (2022) found that trust and emotional comfort go hand in hand—when institutions don't support students, trust weakens. Adams et al. (2022) also showed that ethical unease grows when technology outpaces policy. Li et al. (2024) argue that AI ethics education must help students manage these emotional tensions, something participants here were doing alone. This shows that emotions aren't side issues—they shape how trust in AI is built or broken.

Another important insight emerged around how students tried to keep ownership of their work. They were not blindly trusting AI. Instead, they rewrote, fact-checked, or discarded AI output to keep their academic voice intact. For them, trust meant using AI selectively, not depending on it fully. Liu et al. (2023) emphasize that building self-regulation is key in AI-rich learning. Qin et al. (2024) describe this as "accountability by practice," where ethical choices come from personal commitment, not external rules. Pratama et al. (2023) argue that AI in education should leave room for critical engagement. Lai et al. (2023) add that keeping decision-making autonomy supports student well-being. Redecker (2017) also highlights the need to teach digital authorship explicitly. What students did here reflects these principles—they drew boundaries, reclaimed judgment, and exercised agency without being told to.

These findings shift the weight of responsibility away from students alone. Yes, students showed agency, but it was often a reaction to unclear policies and inconsistent teaching practices. UNESCO (2021) calls for shared ethical dialogue where students are treated as moral agents, not just rule followers. In its absence, students build their own ethical maps—often guessing or relying on peers. Kieslich et al. (2022) note that public trust in AI depends on feeling morally grounded and supported. European Parliament (2021) stresses that governance must empower users, not just protect against misuse.

Institutions, then, have an ethical duty not only to set AI rules but to create spaces where those rules are discussed and lived. Otherwise, the burden of navigating uncertainty remains on students, even when they are trying to do the right thing.

This study, like all qualitative work, has its limits. The sample was small and context-specific—twelve students shaped by their university culture and their own interpretations. As Berendt (2019) notes, AI ethics is both personal and contextual. What feels right in one setting may not in another. Future research should bring in faculty, administrators, and policymakers to see how ambiguity is created and perceived across levels. Li et al. (2024) stress that ethics education must involve all stakeholders. Wang et al. (2024) argue for ongoing evaluation of AI governance, not one-time policies. These insights point to a clear message: students should not carry the weight of uncertainty alone. Institutions must be active partners in shaping what ethical AI use looks like in real learning environments.

5 CONCLUSION

This study shows that students' trust in AI tools isn't fixed or purely about technology. It's something they build and adjust through negotiation—interpreting vague rules, weighing emotional risks, and drawing personal lines around authorship. Instead of treating trust as faith in system accuracy, the students framed it as a careful stance shaped by what institutions do—or fail to do. When formal guidance is missing, students don't act recklessly. They move cautiously, creating their own ethical frameworks to fill the silence. This makes clear that trust is not just technical; it's social, relational, and deeply shaped by context.

More importantly, the emotional cost of that ambiguity lands on students. Many used AI responsibly but still felt uneasy, unsure whether their actions would be seen as acceptable. That anxiety isn't incidental—it reflects a lack of clear communication from institutions. Emotional discomfort becomes a signal of systemic gaps, not individual failure. Students aren't only navigating how to use a tool; they're navigating how their choices will be judged in a shifting moral space. Real ethical literacy must go beyond technical know-how. It must also create emotional safety, giving students a shared language and clear ground to stand on.

Lastly, this study highlights that students' agency is an underused strength. Far from being passive users or rule-breakers, participants showed thoughtful judgment, editing AI outputs, and seeking clarity. Institutions could harness this by involving students directly in shaping AI policy. Rather than issuing rigid top-down rules, they can build trust through open dialogue and shared authorship. A participatory approach—grounded in empathy, transparency, and mutual respect—would shift education from reactive control to proactive trust-building. In doing so, AI governance in learning spaces can become something co-created, not simply enforced.

REFERENCES

Adams, C., Pente, P., Lemermeyer, G., Turville, J., & Rockwell, G. (2022). Artificial intelligence and teachers' new ethical obligations. *The International Review of Information Ethics, 31*(1). https://doi.org/10.29173/irie483

Ahuja, A. S., Polascik, B. W., Doddapaneni, D., Byrnes, E. S., & Sridhar, J. (2023). The digital metaverse: Applications in artificial intelligence, medical education, and integrative health. *Integrative Medicine Research, 12*(1), 100917. https://doi.org/10.1016/j.imr.2022.100917

Berendt, B. (2019). AI for the common good?! Pitfalls, challenges, and ethics pen-testing. *Paladyn, Journal of Behavioral Robotics, 10*(1), 44–65. https://doi.org/10.1515/pjbr-2019-0004

Bullock, J. B., Pauketat, J. V. T., Huang, H., Wang, Y.-F., & Anthis, J. R. (2025). Public opinion and the rise of digital minds: Perceived risk, trust, and regulation support. *arXiv preprint*. https://doi.org/10.48550/arXiv.2504.21849

Choung, H., David, P., & Ross, A. (2022). Trust in AI and its role in the acceptance of AI technologies. *International Journal of Human-Computer Interaction, 38*(6), 1–15. https://doi.org/10.1080/10447318.2022.2050543

Cotton, D. R., Cotton, P. A., & Shipway, J. R. (2024). Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. *Innovations in Education and Teaching International, 61*(2), 228–239. https://doi.org/10.1080/14703297.2023.2190148

Das, B. C., Amini, M. H., & Wu, Y. (2024). Security and privacy challenges of large language models: A survey. *arXiv preprint* arXiv:2402.00888. https://doi.org/10.48550/arXiv.2402.00888

Ding, J., Li, B., Xu, C., Qiao, Y., & Zhang, L. (2023). Diagnosing crop diseases based on domain-adaptive pre-training BERT of electronic medical records. *Applied Intelligence, 53*(12), 15979–15992. https://doi.org/10.1007/s10489-022-04346-x

EP Committee. (2021). *European Parliament resolution of 19 May 2021 on artificial intelligence in education, culture and the audiovisual sector (2020/2017 (INI))* [Resolution]. European Parliament. https://www.europarl.europa.eu/doceo/document/TA-9-2021-0238_EN.html

Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence, 1*(9), 389–399. https://doi.org/10.1038/s42256-019-0088-2

Kahneman, D., & Egan, P. (2011). *Thinking, fast and slow.* Farrar, Straus and Giroux. https://writemac.com/wp-content/uploads/2021/07/Thinking-Fast-and-Slow.pdf

Kasneci, E., Seßler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., … Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences, 103*, 102274. https://doi.org/10.1016/j.lindif.2023.102274

Kieslich, K., Keller, B., & Starke, C. (2022). Artificial intelligence ethics by design: Evaluating public perception on the importance of ethical design principles of artificial intelligence. *Big Data & Society, 9*(1), 20539517221092956. https://doi.org/10.1177/20539517221092956

Lai, T., Xie, C., Ruan, M., Wang, Z., Lu, H., & Fu, S. (2023). Influence of artificial intelligence in education on adolescents' social adaptability: The mediatory role of social support. *PLoS ONE, 18*(3), e0283170. https://doi.org/10.1371/journal.pone.0283170

Lazarus, R. S. (1991). *Emotion and adaptation.* Oxford University Press. https://books.google.com

Li, Y., Zhu, Y., & Fan, X. (2024). Exploration and enlightenment of adolescent artificial intelligence ethics education: A case study of MIT. *Modern Distance Education, 1–13.* https://chn.oversea.cnki.net/kcms/detail/detail.aspx?filename=YUAN202401001&dbcode=CJFQ&dbname=CJFDLAST2024

Liu, Y., Han, T., Ma, S., Zhang, J., Yang, Y., Tian, J., … Ge, B. (2023). Summary of ChatGPT-related research and perspective towards the future of large language models. *Meta-Radiology*, 100017. https://doi.org/10.1016/j.metrad.2023.100017

Mutimukwe, C., Viberg, O., Oberg, L. M., & Cerratto-Pargman, T. (2022). Students' privacy concerns in learning analytics: Model development. *British Journal of Educational Technology, 53*(4), 932–951. https://doi.org/10.1111/bjet.13234

NSW Government. (2023). *Australian framework for generative artificial intelligence in schools: Consultation paper.* NSW Government. https://education.nsw.gov.au/content/dam/main-education/about-us/strategies-and-reports/consultation-items/AI_Consultation_Paper.pdf

Pavlik, J. V. (2023). Collaborating with ChatGPT: Considering the implications of generative artificial intelligence for journalism and media education. *Journalism & Mass Communication Educator, 78*(1), 84–93. https://doi.org/10.1177/10776958221149577

Pratama, M. P., Sampelolo, R., & Lura, H. (2023). Revolutionizing education: Harnessing the power of artificial intelligence for personalized learning. *Klasikal: Journal of Education, Language Teaching and Science, 5*(2), 350–357. https://doi.org/10.52208/klasikal.v5i2.877

Qin, A., Jingmei, Y., Xiaoshu, X., Yunfeng, Z., & Huanhuan, Z. (2024). Decoding AI ethics from users' lens in education: A systematic review. *Heliyon, 10*(20), e39357. https://doi.org/10.1016/j.heliyon.2024.e39357

Redecker, C. (2017). *European framework for the digital competence of educators: DigCompEdu (No. JRC107466).* Joint Research Centre (Seville site). https://ideas.repec.org/p/ipt/iptwpa/jrc107466.html

Smith, V., Shamsabadi, A. S., Ashurst, C., & Weller, A. (2023). Identifying and mitigating privacy risks stemming from language models: A survey. *arXiv preprint* arXiv:2310.01424. https://doi.org/10.48550/arXiv.2310.01424

UNESCO. (2019). *Beijing consensus on artificial intelligence and education.* UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000368303

UNESCO. (2021). *Recommendation on the ethics of artificial intelligence.* UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000381137

Wang, N., Wang, X., & Su, Y. S. (2024). Critical analysis of the technological affordances, challenges and future directions of generative AI in education: A systematic review. *Asia Pacific Journal of Education*, 1–17. https://doi.org/10.1080/02188791.2024.2305156

Authors' Contribution

Both authors contributed equally to the development of this article.

Data availability

All datasets relevant to this study's findings are fully available within the article.

How to cite this article (APA)

Alobud, O. (2025). BETWEEN TRUST AND UNCERTAINTY: HOW STUDENTS CONSTRUCT ETHICAL BOUNDARIES IN AI-DRIVEN LEARNING. *Veredas Do Direito*, *22*(4), e223741. https://doi.org/10.18623/rvd.v22.n4.3741