Received: 21 October 2024 | Revised: 17 January 2025 | Accepted: 10 February 2025

DOI: 10.1002/jee.70000

GUEST EDITORIAL

The role of learner trust in generative artificially intelligent learning environments

Large language models (LLMs) and specifically generative pre-trained transformer (GPT) models have dramatically increased the availability, use, and performance of generative artificial intelligence (AI) tools. Inevitably, there have been pushes and steps toward integrating generative AI into educational settings (Baidoo-Anu & Ansah, 2023; Su & Yang, 2023), and researchers have argued that these technologies hold the potential for offering deeply personalized learning (Jaldi et al., 2025; Mannekote et al., 2024). Although generative AI is still evolving, these technologies are frequently adopted as educational institutions strive to scale or operate effectively while improving their educational offerings and perceived value (e.g., Swaak, 2024). Given the continued integration of generative AI into educational settings, researchers should consider the myriad factors that will influence the successful implementation and use of these tools. We focus on learners' trust as an underappreciated contextual factor and a necessary condition for continued and appropriate use of generative AI in education.

Trust results from social experiences, and can be a critical component of technological interventions in education. For example, the media equation theory suggests we treat computers similar to how we treat people (Reeves & Nass, 1996), and social agency theory suggests that social cues in learning environments activate social schema that make the learner more likely to learn the material (Mayer et al., 2003). The cognitive–affective–social theory of learning in digital environments explicitly lays out how social cues can influence the learning process (Schneider et al., 2022). Given the social underpinnings of computer-based learning experiences, understanding how learners' trust in generative AI tools affects their use and learning outcomes is apropos. There is extensive literature on trust in various human and automation settings like driving, the workplace, and healthcare (e.g., Asan et al., 2020; Dirks & de Jong, 2022; Kraus et al., 2020). Unfortunately, research around learners' trust and computer-based systems is more limited, particularly research exploring the complex relationships and interactions involved in learning environments that include generative AI.

For the engineering educator, the need to operationalize and identify key problems of trust in generative AI is particularly important. It is likely that engineering students will soon be designing critical infrastructure within the domain of AI, or relying on AI in some capacity. Engineering educators should facilitate students developing a critical lens toward generative AI, rather than unilaterally accepting AI outputs in various tasks without first validating those outputs. Engineering educators should also avoid promoting the surprising capabilities of AI performance, especially in comparison to human performance. Such comparisons are often limited to closed-world environments (e.g., classic board games) or very narrow scopes of task (e.g., text summarization), and ignore how challenges to AI performance might arise upstream or downstream (e.g., changes to the underlying data, model, or application space). As such, we want learners to build critical heuristics informing their trust in generative AI. That way when they use generative AI, they are able to understand the capabilities and shortcomings of the systems and what they are most appropriately used for (Qadir, 2023). Understanding the underpinnings of a trusting relationship with technology in general, and generative AI specifically, is therefore an important topic for engineering educators to consider.

Although our research group is not situated in the context of generative AI, we have conducted several studies investigating how different educational interventions can influence trust and learning, and how trust and learning constructs intersect. These studies highlight how trust research can be conducted in the context of educational research, which we hope will inspire novel studies on trust in generative AI tools in educational settings. For example, we examined how the type of voice used by a virtual character (either human, modern text-to-speech, or older text-to-speech) influenced learners' trust, learning, and perceptions of the system (Chiou et al., 2020). While no significant differences in learning outcomes were found, the type of voice did influence the learners' trust and credibility in the system.

The results of this study clearly highlighted what had been shown in previous research regarding the similarities between modern text-to-speech and learning outcomes, yet the results around trust and credibility were novel and intriguing. Building further, we explored these constructs in relation to refutational texts (Schroeder et al., 2023), which have been shown to widely benefit learning (for a meta-analysis, see Schroeder & Kucera, 2022). Our results showed that trust and credibility were measurably distinct from one another; a finding we feel must be considered in future studies of trust in relation to generative AI. This finding departs from other established work that operationalizes trust as a component of credibility (Tseng & Fogg, 1999), or work that focuses on factors affecting behavioral intent to use technology (Venkatesh et al., 2003) rather than on clarifying the complex interdependencies that can underpin trusting attitudes toward technology (Chiou & Lee, 2023; Lee & See, 2004). Furthermore, perceptions of trust in the message did not mediate the relationship between prior knowledge and learning outcomes; however, they did predict learning outcomes almost as well as prior knowledge. These findings further highlight the role of trust and related constructs in education, such as credibility (McCroskey et al., 2004), and the importance of building an understanding of how they operate and influence one another.

Research like this is increasingly important as the adoption of generative AI in formal education settings is still emerging (e.g., Giannakos et al., 2024; Wu, 2023). AI in education involves various groups including educators, researchers, education technology developers, and legislators, and each with their own set of goals and needs. The complexity of these group relationships and their interactions with AI in education makes the examination of trust within that system even more pertinent. For example, schools' goals and needs may include having a cost-effective tool that produces quantifiable differences in student achievement. Teachers' goals and needs include having an easy-to-use tool that can alleviate their ever-increasing workload while aligning with the material and their pedagogical approach. Students' goals and needs include learning material, interacting with a clear and easy interface, and easing their own workload, which differs from teachers' workload (Jochim & Lenz-Kesekamp, 2023). Stances toward technology, including continued use, learning gains, and perceptions of the technology depend on addressing different group needs and goals.

Research has shown that trust in technology is a critical factor in appropriate use and successful adoption in complex task environments like education. Goldshtein et al. (under review) examined automated writing evaluation (AWE) contexts as a case study for understanding how trusting can unfold in automated learning systems. AWE tools are widely used to support standardized assessments, writing instruction and assessment, and to improve writing proficiency. Unlike most human–automation systems, AWE includes at least three participants: the student, teacher, and the automated tool. Each participant has a set of goals, needs, and requirements for trusting the tool, some unique, while others overlapping. The researchers adopt a relational trusting perspective (Chiou & Lee, 2023) which emphasizes sociotechnical interdependencies between multiple involved participants, their goals, and their goal environments. The relational approach conceptualizes trust as a dynamic and evolving state that must be cultivated and sustained through relationships with technology. Automated systems are perceived to be more trustworthy when they are credible (Matthews et al., 2020; Rubin et al., 2020) and responsive (Natarajan et al., 2022). Perceptions of credibility and responsivity are nurtured by people and systems' interactions and the strategies employed by the different entities that are part of the system.

As generative AI use rises in educational settings, particularly in engineering education, we see a dire need for research to understand how we can design and implement these tools in ways that support relational trust with students, teachers, and other involved groups. Transdisciplinary work combining insights and practices from researchers of education, trust, human-computer interaction, computer science, and engineering will be critical for effectively using generative AI in engineering education with accessibility and equity in mind, while addressing various student and workforce needs. There are a plethora of questions related to learners' trust in generative AI. Questions related to learners' trust in generative AI can be grounded in theories about trust (e.g., Chiou & Lee, 2023) and in theories of learning that account for social interactions (e.g., Schneider et al., 2022). For example, "*How does the pedagogical approach used to design generative AI-based lessons influence learners*" *trust in generative AI, and how does their trust influence learning outcomes?,'* and questions like "*How does generative AI's way of communication with learners influence their trust in the system, and to what extent does this trust influence their learning outcomes?*" may be worthwhile to explore in an engineering education context. Similarly, there are equally important questions that must be addressed for other stakeholder groups. For example, engineering educators may wish to use generative AI tools to aid in creating examples for their courses. In this case, a critical question may be, "*To what extent does the educator's trust in the generative AI output depend on their AI literacy and prompting strategy?*" While these are only a few examples, this research space is broad. In particular, investigating the intersections of pedagogical design and learning outcomes, along with learner and instructor trust, could be particularly fruitful to better understand how engineering educators can design more effective instruction when students are using generative AI. This line of research could also help encourage students to build appropriate levels of trust in the technology as they learn about its effectiveness in different contexts. Further, it is critical that we examine within ourselves as educators how, when, and why we should trust various generative AI tools in pedagogical contexts.

As we stand facing a pivotal moment with generative AI, these tools' ability to address the most impactful problems in education requires responsivity as a central component of their development and validation. This way, generative AI tools can continue to earn trust by attending to specific needs, goals, and concerns within education and into the next-generation engineering workforce. Given the importance of engineering in society, understanding the relationships between educators, engineering students, practitioners, and generative AI has never been more necessary.

Maria Goldshtein [1]
Noah L. Schroeder [2]
Erin K. Chiou [3]

[1] *Learning Engineering Institute, Arizona State University Arizona State University, Tempe, Arizona, USA*
[2] *Department of Computer & Information Science & Engineering, University of Florida, Gainesville, Florida, USA*
[3] *Human Systems Engineering, Arizona State University, Mesa, Arizona, USA*

Correspondence
Maria Goldshtein, Arizona State University Arizona State University, Learning Engineering Institute, Tempe, AZ, USA.
Email: maria.goldshtein@asu.edu

ORCID

Maria Goldshtein https://orcid.org/0000-0001-7440-9245

REFERENCES

Asan, O., Bayrak, A. E., & Choudhury, A. (2020). Artificial intelligence and human trust in healthcare: Focus on clinicians. *Journal of Medical Internet Research*, *22*(6), e15154. https://doi.org/10.2196/15154

Baidoo-Anu, D., & Ansah, L. O. (2023). Education in the era of generative artificial intelligence (AI): Understanding the potential benefits of ChatGPT in promoting teaching and learning. *Journal of AI*, *7*(1), 52–62. https://doi.org/10.61969/jai.1337500

Chiou, E. K., & Lee, J. D. (2023). Trusting automation: Designing for responsivity and resilience. *Human Factors*, *65*(1), 137–165. https://doi.org/10.1177/00187208211009995

Chiou, E. K., Schroeder, N. L., & Craig, S. D. (2020). How we trust, perceive, and learn from virtual humans: The influence of voice quality. *Computers & Education*, *146*, 103756.

Dirks, K. T., & de Jong, B. (2022). Trust within the workplace: A review of two waves of research and a glimpse of the third. *Annual Review of Organizational Psychology and Organizational Behavior*, *9*(1), 247–276. https://doi.org/10.1146/annurev-orgpsych-012420-083025

Giannakos, M., Azevedo, R., Brusilovsky, P., Cukurova, M., Dimitriadis, Y., Hernandez-Leo, D., Järvelä, S., Mavrikis, M., & Rienties, B. (2024). The promise and challenges of generative AI in education. *Behaviour & Information Technology*, *43*(16), 1–27. https://doi.org/10.1080/0144929X.2024.2394886

Goldshtein, M., Chiou, E. K., & Roscoe, R. D. (under review). *Improving automated writing evaluation through relational trust*. Manuscript submitted for publication.

Jaldi, C. D., Ilkou, E., Schroeder, N., & Shimizu, C. (2025). Education in the era of Neurosymbolic AI. *Journal of Web Semantics*, *85*(12), 100857. https://doi.org/10.1016/j.websem.2024.100857

Jochim, J., & Lenz-Kesekamp, V. K. (2023). Unlocking the power of ChatGPT: A framework for applying generative AI in education. *ECNU Review of Education*, *6*(3), 355–366. https://doi.org/10.1177/20965311231168423

Kraus, J., Scholz, D., Stiegemeier, D., & Baumann, M. (2020). The more you know: Trust dynamics and calibration in highly automated driving and the effects of take-overs, system malfunction, and system transparency. *Human Factors*, *62*(5), 718–736. https://doi.org/10.1177/0018720819853686

Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors*, *46*(1), 50–80. https://doi.org/10.1518/hfes.46.1.50_30392

Mannekote, A., Davies, A., Pinto, J. D., Zhang, S., Olds, D., Schroeder, N. L., Lehman, B., Zapata-Rivera, D., & Zhai, C. (2024). Large language models for whole-learner support: Opportunities and challenges. *Frontiers in Artificial Intelligence*, *7*, 1–7. https://doi.org/10.3389/frai.2024.1460364

Matthews, G., Lin, J., Panganiban, A. R., & Long, M. D. (2020). Individual differences in Trust in Autonomous Robots: Implications for transparency. *IEEE Transactions on Human-Machine Systems*, *50*(3), 234–244. https://doi.org/10.1109/THMS.2019.2947592

Mayer, R. E., Sobko, K., & Mautone, P. D. (2003). Social cues in multimedia learning: Role of speaker's voice. *Journal of Educational Psychology*, *95*(2), 419–425. https://doi.org/10.1037/0022-0663.95.2.419

McCroskey, J. C., Valencic, K. M., & Richmond, V. P. (2004). Toward a general model of instructional communication. *Communication Quarterly*, *52*(3), 197–210. https://doi.org/10.1080/01463370409370192

Natarajan, M., Akash, K., & Misu, T. (2022). *Toward adaptive driving styles for automated driving with users' trust and preferences*. Paper presented at the ACM/IEEE International Conference on Human-Robot Interaction https://doi.org/10.1109/HRI53351.2022.9889313

Qadir, J. (2023). *Engineering education in the era of ChatGPT: Promise and pitfalls of generative AI for education*. Paper presented at the IEEE Global Engineering Education Conference https://doi.org/10.1109/EDUCON54358.2023.10125121

Reeves, B., & Nass, C. (1996). *The media equation: How people treat computers, television, and new media like real people* (Vol. 10, pp. 19–36). Cambridge.

Rubin, R. B., Palmgreen, P., & Sypher, H. E. (2020). Source credibility scale–McCroskey. In R. B., Rubin, P. Palmgreen, & H. E. Sypher (Eds.), *Communication research measures* (pp. 332–339). Routledge.

Schneider, S., Beege, M., Nebel, S., Schnaubert, L., & Rey, G. D. (2022). The cognitive-affective-social theory of learning in digital environments (CASTLE). *Educational Psychology Review*, *34*(1), 1–38. https://doi.org/10.1007/s10648-021-09626-5

Schroeder, N. L., Chiou, E. K., Siegle, R. F., & Craig, S. D. (2023). Trusting and learning from virtual humans that correct common misconceptions. *Journal of Educational Computing Research*, *61*(4), 790–816. https://doi.org/10.1177/07356331221139859

Schroeder, N. L., & Kucera, A. C. (2022). Refutation text facilitates learning: A meta-analysis of between-subjects experiments. *Educational Psychology Review*, *34*(2), 957–987. https://doi.org/10.1007/s10648-021-09656-z

Su, J., & Yang, W. (2023). Unlocking the power of ChatGPT: A framework for applying generative AI in education. *ECNU Review of Education*, *6*(3), 355–366. https://doi.org/10.1177/20965311231168423

Swaak, T. (2024). *Arizona state and openAI are now partners. What does that mean?* https://www.chronicle.com/article/arizona-state-and-openai-are-now-partners-what-does-that-mean

Tseng, S., & Fogg, B. J. (1999). Credibility and computing technology. *Communications of the ACM*, *42*(5), 39–44.

Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS Quarterly*, *27*(3), 425–478. https://doi.org/10.2307/30036540

Wu, Y. (2023). Integrating generative AI in education: How ChatGPT brings challenges for future learning and teaching. *Journal of Advocacy, Research and Education*, *2*(4), 6–10. https://doi.org/10.56397/JARE.2023.07.02

AUTHOR BIOGRAPHIES

Maria Goldshtein is a research scientist at the Learning Engineering Institute at Arizona State University. Her research focuses on examining students' varying backgrounds and interactions with the higher educational system in general, and as it relates to educational technology in particular.

Noah L. Schroeder is a research scientist with the Learn Dialogue Group at the University of Florida. His research is focused on designing effective educational technologies and research synthesis.

Erin K. Chiou is an associate professor of human systems engineering at Arizona State University. Her research focuses on human-agent interactions in complex work systems, including how trust influences system performance.