# Albert or Alberta? Investigating Socratic Dialogue and Modality in AI-Assisted Engineering Problem Solving

Augusto Grimaldi Ter Sarkisian[1], Cameron Miller[2], Neeraja Poonjolai[1], Candelaria Aramburu[1], and Jad Atweh[1,3,*]
[*]*Corresponding Author: jad.atweh@ufl.edu*

[1] *Department of Industrial & Systems Engineering*
*University of Florida*
*Gainesville, Florida, USA*

[2] *Department of Psychology*
*University of Florida*
*Gainesville, Florida, USA*

[3] *Department of Engineering Education*
*University of Florida*
*Gainesville, Florida, USA*

***Abstract*—Artificial intelligence (AI)-based tools are increasingly embedded in engineering education. While current systems provide direct answers, such approaches may encourage passive engagement and limit transfer of learning. This study investigates how AI chatbot interaction style (direct response vs. Socratic questioning) and communication modality (text vs. audio) influence engineering students' learning and transfer of problem-solving skills. Sixty undergraduate Industrial and Systems Engineering students participated in a between-subjects experiment across four conditions. Over two phases, participants solved Operations Research problems with their assigned AI in Phase 1 and independently in Phase 2. Results revealed that students interacting with the Socratic model (Alberta) completed independent problems faster, despite taking longer when solving problems with AI support. Communication modality moderated these effects: text-based Socratic interactions produced the strongest performance, whereas audio-based Socratic interactions often increased completion time without consistent gains in perceived learning or performance. Qualitative findings further revealed that students valued the Socratic model's ability to help them understand concepts rather than simply supplying answers. This work provides empirical evidence to inform the design of classroom-integrated AI tools that support learning and more effective human-AI collaboration in engineering problem solving.**

***Keywords*—Engineering Education, AI Chatbots, Socratic Dialogue, AI Modality, Learning**

## I. Introduction

As Artificial Intelligence (AI)-based conversational systems are becoming a key component in higher education, students in engineering disciplines are further seeking procedural support during problem solving [1]. Large language model (LLM)-based tools now provide on-demand solutions and step-by-step guidance across topics ranging from linear algebra to operations research [2]. While such tools offer novel access to fast information, their increasing usage among students raises an important pedagogical question: Do AI chatbots simply accelerate answer acquisition, or do they support learning and transfer of knowledge?

Various organizations have proposed standards and recommendations for technology integration in today's dynamic educational system. One such framework from the Partnership for 21st Century Learning (P21) envisions a future for classrooms in which transformative technology use develops real-world competencies in students, fostering four key skills: critical thinking, collaboration, communication, and creativity [3]. Under the right conditions, building interactions with AI chatbots into the curriculum has the potential to promote these real-world competencies in a way that redefines what is learned in the classroom and how students learn it.

Widely available AI chatbots prioritize responsiveness and completeness, often delivering direct explanations or fully worked solutions [4], [5]. This may encourage passive engagement and limit opportunities for engineering students to think critically [6]. Research suggests that learning is strengthened when students actively explain and justify decisions which have been consistently shown to improve conceptual understanding and knowledge transfer [2]. Alternative approaches to AI-supported learning emphasize guided reasoning rather than direct answer delivery. In particular, Socratic-style interactions, centered on questioning and prompting, have been shown to encourage deeper cognitive processing and reflection [6]. Additionally, advances in conversational interfaces have introduced multiple modes of interaction, such as text-based and voice-based communication, each of which may shape how students engage with problem-solving tasks [7], [8].

Despite growing interest in AI-supported learning, there remains a need to better understand how different forms of AI interaction influence student engagement and learning outcomes in engineering contexts. This study examines how chatbot guidance style and communication modality affect students' perceived learning and their ability to apply problem-solving skills independently.

## II. Background

LLM-based tools have fundamentally altered how students approach academic support in engineering education. Socratic-style AI has increasingly been explored as a way to move beyond answer-giving and instead bolster learners' critical thinking through questioning and guided reflection [6], [9]. Across several recent studies, Socratic chatbots are positioned as a possible alternative to traditional AI tutors that provide direct solutions. For example, Duelen et al. [6] implemented a Socratic chatbot to help citizens interrogate news content and found that systematic questioning can stimulate users' critical reflection about disinformation signals. In K-12 science classrooms, Kao et al. [10] showed that an AI-powered Socratic version of an argument-driven inquiry environment produced greater gains in critical thinking than a non-AI version. Together, these findings suggest that embedding Socratic dialogue into AI systems can positively influence learners' reasoning processes and critical thinking across diverse domains. Across these studies, Socratic AI is typically examined outside of higher-level engineering coursework and is often evaluated using immediate or self-report outcomes rather than independent performance on novel problems [5]. Consequently, it remains unclear whether Socratic guidance, when implemented in AI chatbots, actually leads to superior learning and transfer of problem-solving skills compared with direct-answer systems in engineering education contexts.

In addition to interaction style, the modality through which AI communicates may shape cognitive engagements. Nowadays, advances in speech and conversational interfaces have enabled voice-based AI assistants that simulate human dialogue. Empirical research on educational chatbots underscores that modality is present but understudied as a driver of learning outcomes. Experimental work on online learning has shown that audio-based presentations can, under certain conditions, outperform text by reducing extraneous cognitive load and improving performance on academic tasks [4], [7]. In engineering problem solving, where symbolic nation and equations are central, text-based interaction may be preferred by students [8], [11]. On the other hand, audio interaction may increase conversational flow between the human and the machine. Despite growing interest in multimodal AI systems, little is known about how communication modality interacts with guidance style to influence learning outcomes.

Finally, much of the emerging research on AI in education evaluates immediate task performance during assisted problem solving [12]. However, engineering competence requires students to perform independently, without real-time AI support. Higher education studies note that, although AI tools are now widely available to students, relatively few investigations collect detailed learning data that connects patterns of AI use to conceptual understanding and attainment of learning outcomes over time [13]. As a result, empirical investigations that examine how AI interaction style influences subsequent unassisted problem solving remain underexplored [2], [14].

The present study addresses these three gaps by examining how AI guidance style (direct answer vs. Socratic dialogue) and communication modality (text vs. audio) jointly shape engineering students' perceived learning and transfer of problem-solving skills. We further examine how these effects influence students' independent performance without AI support.

## III. Methdos

### A. Participants

Sixty undergraduate Industrial and Systems Engineering (ISE) students (30 males, 29 females, and 1 undisclosed gender) at the University of Florida were recruited for the study. To be eligible to participate in the study, participants were required to have previously completed the Operations Research I (OR1) course to ensure familiarity with the problem domain. The study was conducted over Zoom, with each session lasting approximately 45-90 minutes. The participants received class credit for their participation in the study. This study was approved by the University of Florida's Institutional Review Board (IRB protocol #ET00048087).

### B. Experimental Design

This study employed a 2×2 between-subjects design, with interaction style (direct response vs. Socratic questioning) and communication modality (text vs. audio) as the independent variables. Participants were randomly assigned to one of four conditions using a stratified randomization procedure. We developed two custom AI chatbots within ChatGPT 5.1. The first chatbot, Albert, was designed to reflect a direct-response interaction style, providing answer-oriented feedback during problem solving. The second chatbot, Alberta, was designed to implement a Socratic questioning approach, prompting students to reason through the problem using guided questions and hints rather than direct solutions.

Participants progressed through the problem in sequential components, and for each response, Albert evaluated the answer as correct or incorrect. If the response was incorrect, the model provided the correct solution for that component; if correct, participants were prompted to proceed to the next part. The model restricted feedback to the current component and did not provide information about subsequent steps.

In contrast, the Alberta model implemented a Socratic, question-driven approach intended to promote reasoning and reflection. Responses that did not fully solve the component, including expressions of uncertainty, were treated as incorrect attempts. In these cases, Alberta provided targeted hints rather than direct solutions. After three unsuccessful attempts, the model revealed the correct answer for that part and followed with a probing question designed to reinforce understanding. For correct responses, the model also generated follow-up questions encouraging participants to explain or justify their reasoning. Participants interacted with their assigned model using either text-based or voice-based input. Performance metrics included task accuracy, completion time, interaction logs, and post-task self-reports of perceived learning and satisfaction.

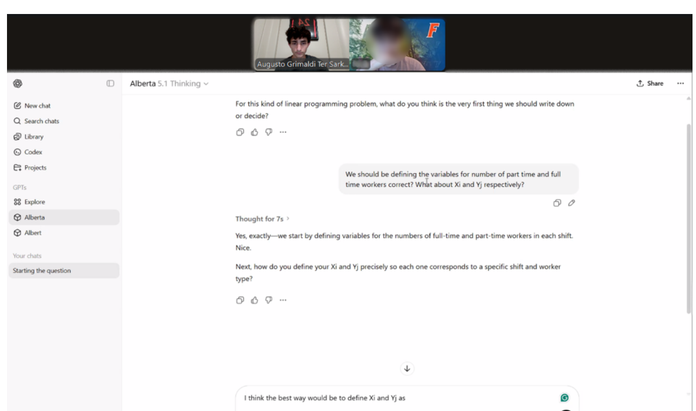


Fig. 1. Participant completing Phase 1 of study using the Alberta text model

## C. Task

Participants completed two mixed-integer linear programming problems aligned with the content of OR1. The problems were designed to be equivalent while differing in contextual details to minimize memorization effects. Each problem required participants to define decision variables, formulate an objective function, and construct constraints reflecting operational policies, resource limitations, and production requirements. Both problems included a binary decision component associated with a fixed cost condition.

The study consisted of two phases. In Phase 1, participants solved the first problem with assistance from their assigned AI model and modality. In Phase 2, participants solved a second, structurally similar problem independently without access to AI, allowing assessment of retained understanding and transfer of problem-solving skills. Participants were allowed to use a calculator, and no time limit was imposed.

## D. Experimental Procedure

Upon arrival, participants were assigned a unique anonymous identifier, provided informed consent, and completed a pre-experiment questionnaire. They were then briefed on the study and instructed to solve an engineering optimization problem, while performing calculations on paper or a tablet. Participants were informed that there was no time limit.

Participants were provided with login credentials for a ChatGPT account configured with the experimental models. Based on their assigned condition, they were instructed which model (Albert or Alberta) and modality (text or voice) to use. The first problem was then delivered via Zoom chat, and the experimenter recorded the time elapsed from problem presentation to the participant's indication that they had finished. After completing the first problem, participants were asked to complete a post–Phase 1 questionnaire.

Participants were then instructed to log out of ChatGPT and set aside all materials from the first problem. The experimenter introduced Phase 2, during which participants solved a second problem independently without AI assistance. Completion time was again recorded. Upon finishing, participants completed a post–Phase 2 questionnaire. Finally, participants scanned their work and emailed it to the experimenter.

# IV. Results

Seven two-way ANOVAs were conducted to examine the effects of chatbot Model (Albert vs. Alberta) and Modality (text vs. voice) on Phases 1 and 2 dependent measures. To control for Type I error across multiple comparisons, a Bonferroni correction was applied and the adjusted alpha level was set to $\alpha = .05/7 = .0071$.

## A. Phase 1: AI-Assisted Performance

The two-way ANOVA of Phase 1 completion time (in minutes; Figure 2) revealed main effect of Model, $F(1, 56) = 5.38$, $p = .024$, $\eta_p^2 = 0.088$, such that participants working with Albert ($M = 29.22$, $SD = 12.61$) completed the task faster than those working with Alberta ($M = 36.43$, $SD = 12.25$). There was also a main effect of Modality, $F(1, 56) = 5.55$, $p = .022$, $\eta_p^2 = 0.09$, with voice interaction ($M = 29.16$, $SD = 10.55$) yielding shorter completion times than text interaction ($M = 36.49$, $SD = 14.03$). The Model × Modality interaction was not significant ($p = .69$).

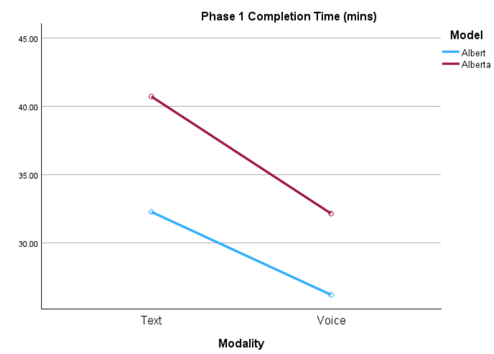


Fig. 2. Phase 1 completion times by Chatbot Model and Modality.

The Phase 1 accuracy ANOVA revealed that the main effect of Model was not significant, $F(1, 56) = 3.17$, $p = .08$, although accuracy tended to be higher for Alberta than for Albert (Figure 3). There was a main effect of Modality, $F(1, 56) = 4.72$, $p = .03$, $\eta_p^2 = 0.078$, with text interaction ($M = 8.95$, $SD = 1.09$) yielding higher accuracy than voice interaction ($M = 8.3$, $SD = 1.25$). The Model × Modality interaction was not significant ($p = .5$).

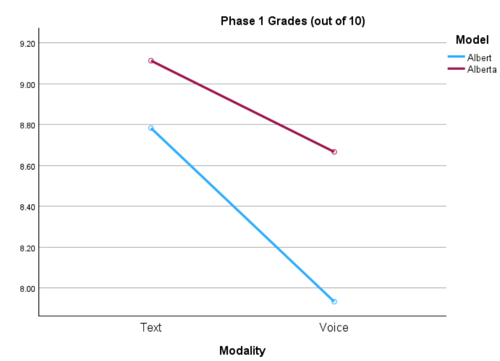


Fig. 3. Phase 1 accuracy by Chatbot Model and Modality.

For Phase 1 satisfaction, a two-way ANOVA showed that the main effect of Modality was not significant, $F(1, 56) = 1.19$, $p = .28$ (Figure 4), indicating similar satisfaction ratings for text and voice. However, there was a significant main effect of Model, $F(1, 56) = 12.39$, $p < .001$, $\eta_p^2 = 0.18$, such that participants reported higher satisfaction with the Alberta model ($M = 4.03$, $SD = 0.93$) than with the Albert model ($M = 3.07$, $SD = 1.17$). The Model × Modality interaction was not significant ($p = .72$).

A two-way ANOVA examined the effects of on perceived learning from the AI-assisted phase, measured with the item "I learned information that I did not know before solving this question" (1–5 Likert scale; Figure 5). The main effect of Modality was not significant, $F(1, 56) = 0.07$, $p = .82$,

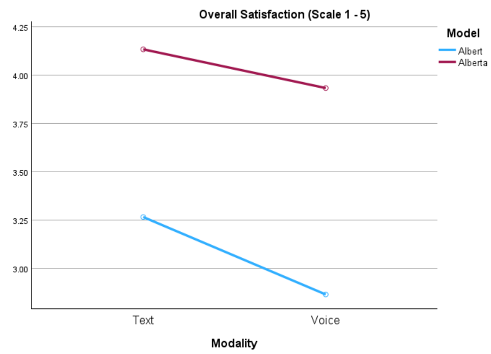


Fig. 4. Phase 1 satisfaction ratings by Chatbot Model and Modality.

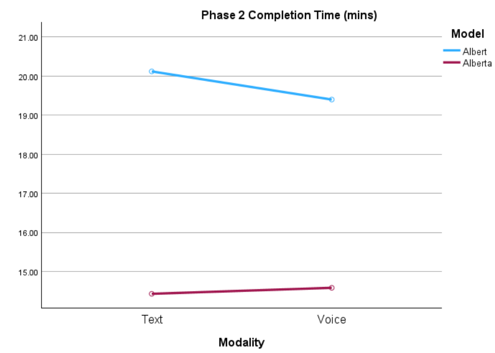


Fig. 6. Phase 2 completion times by Chatbot Model and Modality.

with similar perceived learning for text. However, there was a significant main effect of Model, $F(1, 56) = 19.55$, $p < .001$, $\eta_p^2 = 0.26$, such that participants reported learning more with the Alberta model ($M = 3.87$, $SD = 1.01$) than with the Albert model ($M = 2.6$, $SD = 1.19$). The Model × Modality interaction was not significant ($p = .25$).

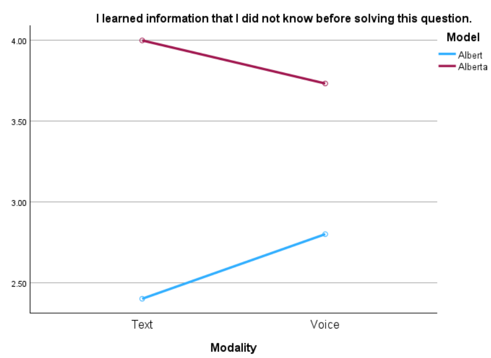


Fig. 5. Phase 1 perceived learning by Chatbot Model and Modality.

### B. Phase 2: Independent Performance

For Phase 2 completion time, time to solve the problem without AI support (Figure 6), the main effect of Modality was non-significant, $F(1, 56) = 0.02$, $p = 0.89$. However, there was a main effect of Model, $F(1, 56) = 6.99$, $p = .011$, $\eta_p^2 = 0.11$, such that participants who had previously worked with Alberta ($M = 14.5$, $SD = 6.11$) solved the independent problem faster than those who had worked with Albert ($M = 19.76$, $SD = 8.79$). The Model × Modality interaction was not significant ($p = .83$).

A two-way between-subjects ANOVA was conducted to examine the effects of AI Model (Albert vs. Alberta) and interaction Modality (text vs. voice) on participants' independent problem-solving accuracy in Phase 2 (Figure 7). The interaction between Model and Modality was not statistically significant, $F(1, 56) = 0.034$, $p = .85$. There was a significant main effect of Model, $F(1, 56) = 35.34$, $p < .001$, $\eta_p^2 = 0.39$, such that participants who interacted with the Alberta model ($M = 9.05$, $SD = 0.74$) demonstrated significantly higher independent accuracy compared to those who interacted with the Albert model ($M = 7.71$, $SD = 1.21$). A significant main effect of Modality was also observed, $F(1, 56) = 19.96$, $p < .001$, $\eta_p^2 = .26$, with participants in the text condition ($M = 8.88$, $SD = 1.09$) outperforming those in the voice condition ($M = 7.88$, $SD = 1.1$) in Phase 2 accuracy.

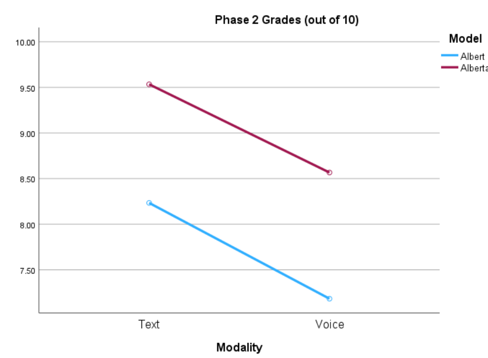


Fig. 7. Phase 2 independent problem-solving accuracy by AI Model and Modality.

Descriptive statistics revealed that participants in the Alberta condition reported higher recall overall ($M = 4.33$, $SD = 0.66$) compared to those in the Albert condition ($M = 3.33$, $SD = 1.3$). Regarding Modality, voice ($M = 3.93$, $SD = 0.94$) and text ($M = 3.73$, $SD = 1.31$) conditions produced similar overall recall ratings. The main effect of the two-way ANOVA was statistically significant, $F(1, 56) = 16.98$, $p < .001$, $\eta_p^2 = 0.23$, indicating that the Alberta model was associated with higher recall than Albert.

The main effect of Modality was not statistically significant, $F(1, 56) = 0.68$, $p = .41$. There was a statistically significant Model × Modality interaction, $F(1, 56) = 12.76$, $p < .001$, $\eta_p^2 = 0.19$. Simple effects analyses with pairwise comparisons indicated that in the Text Modality, participants who interacted with Alberta reported significantly higher recall ($M = 4.67$, $SD = 0.49$) than those who interacted with Albert ($M = 2.8$, $SD = 1.21$), with a mean difference of 1.87, $p < .001$, 95% CI [1.18, 2.55]. In contrast, in the Voice Modality, recall did not differ significantly between Alberta ($M = 4$, $SD = 0.65$) and Albert ($M = 3.87$, $SD = 1.19$), $p = .7$, 95% CI [-0.55, 0.82] (Figure 8).

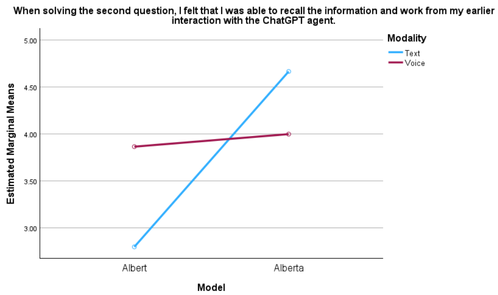


Fig. 8. Recall ratings as a function of AI Model (Albert vs. Alberta) and Modality (text vs. voice).

### C. Qualitative Analysis

A thematic analysis of student responses revealed trends within both modalities and interaction styles. For modality, 33.33% ($n = 10$) of participants using the audio format distinctly stated that it slowed down their ability to complete the task relative to the more common text format. Some participants additionally described the AI voice model as distracting at times when it misinterpreted speech and felt that the model was more agreeable than expected. Many negative responses with audio came from the Albert Voice group, with 80% of these responses ($n = 8$) coming from participants who received this condition. Participants in the Alberta Voice condition reported more positive, beneficial results ($n = 7$, 46.67%) from the audio format than the Albert Voice group, with one Alberta Voice participant stating: *"It made it easier because it felt like I was having a conversation with someone and I had to be more focused"* (P25).

Participants interacting with the Alberta model frequently described the experience as more beneficial for learning, noting that the model's questioning approach encouraged them to think more deeply about the problem rather than passively receiving answers. One participant stated that Alberta *"forced you to think instead of the AI agent giving you the answer"* (P32). Others appreciated the learning benefits of this structure, describing it as *"actively learning as I was working on the problem"* (P19) and *"not just copying down information"* (P23). Some participants reported negative effects of the model's probing questions, which increased the time required to solve the problem.

For Phase 2 responses, where participants reflected on their independent task while comparing it to Phase 1, many iterated that the knowledge check allowed them to apply their learning and demonstrate problem-solving skills. One participant shared that *"Working independently, it is easy to get off the right track early and it snowballs each step"* (P43) When reflecting on their AI collaboration, many participants reported seeing beneficial results from completing a similar question alongside an AI chatbot.

## V. Discussion & Conclusion

This study investigated how AI chatbot interaction style (Albert direct response vs. Alberta Socratic questioning) and communication modality (text vs. audio) influence engineering students' learning and transfer of problem-solving skills. The findings reveal that both interaction style and modality significantly shaped learning outcomes, though in distinct ways across different phases of the task.

### A. Time Costs of Socratic Interaction

Completion time differences in Phase 1 indicated that both the Socratic interaction style and text modality added time for task completion relative to their counterparts. The Alberta model was expected to take longer due to the additional probing questions that required students to articulate their reasoning before advancing. However, the audio modality produced a decrease in task time, likely due to its hands-off benefit [7]. Participants who worked with the audio modality reported being better able to work on the question without physically going back and forth between their AI chatbot and their work, allowing for more fluid multitasking.

There were no significant accuracy changes based on interaction style, but participants using the audio modality models showed a decline in performance. This was likely caused by ChatGPT's audio feature not having the ability to reference previous responses and display numerical values visually. Participants' inability to read, reread, or carefully process the AI's entries at their own pace may have reduced the AI partner's effectiveness [7].

### B. Socratic Advantages in Independent Problem-Solving

After transitioning to the independent phase (Phase 2), the pattern reversed in ways that supported our central hypothesis. Modality held no significant effect on completion time, but participants who had previously used Alberta had shorter completion times compared to those who used Albert. Phase 2 grade accuracy further supported our hypothesis that Alberta participants would significantly outperform Albert participants. Students who used the Socratic model reported significantly higher satisfaction, recall, and perceived learning scores than those who used the direct model. This data suggested that even with longer initial task time, students could recognize the advantage in collaborating with a model which adopted the Socratic approach. This aligns with prior research showing that students value pedagogical approaches that promote deeper understanding over immediate convenience [1], [12].

This shift demonstrated that the Socratic model promoted problem-solving skills and transfer of knowledge at a greater rate than the direct model. The additional time spent with Alberta in Phase 1 was not wasted, but rather reflected active cognitive engagement as students reasoned through each step, explained their logic, and responded to probing questions. This mirrors the concept of productive friction, in which introducing struggle during initial learning improves retention and transfer [6], [9].

Additionally, students who used the Alberta model in the text condition reported much higher recall than those who used Albert in the same condition. This difference suggests that the benefits of Socratic questioning may depend on how information is delivered. Text appears to strengthen the impact of guided questioning by allowing students to track steps, revisit reasoning, and process information more deliberately, which supports deeper learning and retention. This is consistent with cognitive load theory, which suggests that learners benefit from being able to control the pace and review of complex information [15].

### C. The Role of Modality in Engineering Education

In the case of modality in collaborative AI use for engineering education, findings showed that audio modality may be less effective for transfer of knowledge relative to text format, particularly in the context of complex, multi-step engineering problems [7]. Feedback from participants suggested that processes containing various numbers, equations, and subscripts, which are commonplace in engineering, are more efficiently supported when the user can visually reference information. The inability to see and review technical content appeared to be a limitation of the current audio interface.

### D. Implications for Design, Pedagogy, and Practice

The findings of this study bear direct implications for designers, educators, and students working with AI in educational contexts. For *designers of AI systems*, the results highlight that interaction style is a consequential design decision that shapes learning outcomes. Socratic models are more effective for learning because question-driven interaction patterns require students to articulate reasoning and engage with underlying concepts rather than passively receiving answers. This reflects established principles of active learning and generative learning theory [5]. For *educators*, these findings suggest that instructors should favor Socratic AI models for contexts where understanding and transfer are the primary goals, and reserve direct-response tools for tasks where efficiency is most important and learning is not the primary aim. Educators might also guide students toward text-based interactions when working with complex, notation-heavy material, while allowing audio for more conceptual or discussion-based activities.

## VI. Limitations & Future Work

Several limitations should be acknowledged when interpreting these findings. First, our sample consisted of ISE students at a single institution. This may limit the generalizability of the results to other engineering disciplines, educational levels, and institutional contexts. Second, the study examined performance on one type of engineering problem, and it remains to be seen whether the benefits of Socratic AI interaction extend to other problem types with different cognitive demands, such as open-ended design tasks or lab-based problem-solving.

Future work should explore how individual differences, such as prior knowledge, may moderate the effects of interaction style and modality. Long-term retention and transfer to dissimilar problems were not assessed and constitute important directions for future research. Additionally, this study focused on individual learning; future research should examine how Socratic AI interactions function in collaborative team settings, such as group projects, where students must negotiate shared understanding.

Collectively, this study demonstrates that the Socratic method in AI is effective for student learning and transfer of knowledge in engineering education. While direct-response AI may expedite immediate task completion, Socratic questioning builds the problem-solving skills that enable students to work independently [16]. As AI becomes increasingly integrated into educational environments, this research underscores that how we design AI interactions matters as much as the technology itself. By prioritizing guided inquiry over immediate answers, we can transform AI from an information delivery tool into a partner that cultivates the independent problem-solving skills essential for engineering practice.

## References


[1] A. T. Neumann, Y. Yin, S. Sowe, S. Decker, and M. Jarke, "An LLM-driven chatbot in higher education for databases and information systems," *IEEE Transactions on Education*, vol. 68, no. 1, 2024.

[2] M. Gerlich, "AI tools in society: Impacts on cognitive offloading and the future of critical thinking," *Societies*, vol. 15, no. 1, p. 6, 2025.

[3] R. Vivekanandan and M. Pierre-Louis, "21st century skills: What potential role for the Global Partnership for Education? A landscape review," Global Partnership for Education, Tech. Rep., 2020.

[4] O. Al-Shamaileh, O. Mubin, and N. Nizamuddin, "Talking tech in education: A systematic review of chatbot evaluation, methods, communication modalities, and impact," *International Journal of Human–Computer Interaction*, vol. 42, no. 5, pp. 2883–2919, 2026.

[5] A. Blasco and V. Charisi, "The impact of large language models on students: A randomised study of socratic vs. non-socratic AI and the role of step-by-step reasoning," *SSRN Electronic Journal*, 2024.

[6] A. Duelen, I. Jennes, and W. Van den Broeck, "Socratic AI against disinformation: Improving critical thinking to recognize disinformation using socratic AI," in *Proceedings of the 2024 ACM International Conference on Interactive Media Experiences*. ACM, Jun. 2024.

[7] J. D. de Oliveira Neto, W. D. Huang, and N. C. de Azevedo Melli, "Online learning: audio or text?" *Educational Technology Research and Development*, vol. 63, no. 4, pp. 555–573, 2015.

[8] A. Bewersdorff, C. Hartmann, M. Hornberger, K. Seßler, M. Bannert, E. Kasneci, and C. Nerdel, "Taking the next step with generative artificial intelligence: The transformative role of multimodal large language models in science education," *Learning and Indiv. Differences*, 2025.

[9] L. Favero, J. A. Pérez-Ortiz, T. Käser, and N. Oliver, "Enhancing critical thinking in education by means of a socratic chatbot," in *International Workshop on AI in Education and Educational Research*. Cham: Springer Nature Switzerland, Oct. 2024, pp. 17–32.

[10] S. Kao, P. Grant, and S. Woltering, "Socratic AI in K–12 science classrooms: Effects on critical thinking, motivation, and self-regulation in a randomized controlled trial," 2025.

[11] M. Laun and F. Wolff, "Chatbots in education: Hype or help? a meta-analysis," *Learning and Individual Differences*, vol. 119, 2025.

[12] B. Jose, J. Cherian, A. M. Verghis, S. M. Varghise, M. S, and S. Joseph, "The cognitive paradox of AI in education: Between enhancement and erosion," *Frontiers in Psychology*, vol. 16, p. 1550621, 2025.

[13] J. L. Pallant, J. Blijlevens, A. Campbell, and R. Jopp, "Mastering knowledge: The impact of generative AI on student learning outcomes," *Studies in Higher Education*, pp. 1–22, 2025.

[14] M. K. Hasan, "How AI quietly undermines the joy and effort of learning: a call for rebalancing education in the digital age," *Annals of Medicine and Surgery*, vol. 87, no. 8, pp. 4693–4694, 2025.

[15] J. Sweller and P. Chandler, "Evidence for cognitive load theory," *Cognition and Instruction*, vol. 8, no. 4, pp. 351–362, 1991.

[16] J. A. Atweh, N. M. Moacdieh, and S. L. Riggs, "Identifying individual-, team-, and organizational-level factors that affect team performance in complex domains based on recent literature," in *Proceedings of the HFES Annual Meeting*, 2022.