Engineering Educators Bringing the World Together
2025 ASEE Annual Conference & Exposition
Palais des congrès de Montréal, Montréal, QC • June 22–25, 2025
Paper ID #46760

Student, Faculty, and Industry Perceptions about Generative AI's Capabilities to Perform Specific Tasks

Dr. Lauren Singelmann, Minnesota State University, Mankato

Lauren Singelmann is an assistant professor at Minnesota State University, Mankato through Iron Range Engineering. She earned her PhD in Electrical and Computer Engineering and STEM Education.

Dr. Jack Elliott, Minnesota State University, Mankato

Jack Elliott is an assistant professor at Iron Range Engineering, a part of Minnesota State University Mankato. Dr. Elliott received his PhD in Engineering Education and M.S. in Mechanical Engineering from Utah State University as an NSF Graduate Research Fellow. His research includes student social support networks in engineering education, experimental fluid dynamics, and developing low-cost technology-based tools for improving fluid dynamics education.

Dr. Yuezhou Wang, Minnesota State University, Mankato

Dr. Yuezhou Wang is an associate professor in both Iron Range Engineering and Twin Cities Engineering programs. He received his B.S. in Mechanical Engineering from Shanghai Jiaotong University, China (2008) and Ph.D. in Materials Science and Engineering from University of Minnesota, Twin Cities (2017). His leading teaching competencies are in areas of materials science, structural analysis, finite element modeling and dynamic systems. He has a broad range of research interests. His technical research focuses on multiscale modeling on mechanical behavior of nanofibers and carbon nanotube materials. In the area of pedagogical research, he is interested in using learning analytics tools to understand and assess engineering students' motivation entrepreneurially minded learning.

Prof. Jacob John Swanson, Minnesota State University, Mankato

Jacob Swanson is the department chair and Professor of Engineering in the Department of Integrated Engineering at Minnesota State University Mankato

Student, Faculty, and Industry Perceptions about Generative AI's Capabilities to Perform Specific Tasks

Introduction

Generative AI (GAI) tools like ChatGPT and Copilot can quickly prepare polished, five paragraph essays and clever limericks about any given topic, but can they multiply seven two-digit numbers? Or answer a question from the Fundamentals of Engineering exam? Or tell you what the image in a "connect-the-dots" puzzle is? GAI tools are designed to be able to produce human-like language responses to given prompts, but performance varies depending on the nature of each task. To further complicate the evaluation of GAI performance, each tool (e.g. ChatGPT, Copilot, Gemini) has its own process for generating responses, and these processes can evolve rapidly – with success varying across tools and across time. Although ChatGPT can successfully complete different types of tasks, current models still show errors in logic, factual information, arithmetic, grammar, reasoning, coding, and even the model's own self-awareness [1]. Assessing the performance of these tools is an ongoing task, and one that engineering students, faculty, and industry professionals must engage with when deciding how to use the responses they get from a GAI tool.

This exploratory study aims to showcase student, faculty, and industry perceptions about the capabilities of GAI to perform various tasks, as well as how they approach testing this performance. The methods, results, and discussion sections offer various insights to the engineering education community; the methods describe an activity that helps build GAI literacy for students, faculty, and industry professionals; the results report on perceptions about GAI within and across these groups; and the discussion provides a variety of takeaways about the current state of GAI and how to improve GAI literacy among engineering students, faculty, and industry professionals.

Background

Many GAI tools like ChatGPT, Copilot, Gemini, and Claude are built using a transformer architecture that uses training data to be able to create human-like responses to various prompts. Transformers are a type of deep learning model that aim to represent language through numerical parameters [2]. The parameters not only quantitatively describe meanings of individual words, but they also describe how certain words used together change context. For example, the word "dog" takes on different contexts in the phrases "my pet dog," "my hotdog," or "my favorite band Three Dog Night," and the self-attention mechanism of the transformer can quantitatively describe how the context of the word "dog" changes in each phrase. By using billions of parameters to create an approximation of language, these tools can generate very human-like responses.

ChatGPT is reported to be the fastest growing online platform yet – gathering 100 million users in two months [3]. These users are not only engaging with these tools for fun, but also for work and school; a 2024 Pew Research study found that 20% of working adults had used ChatGPT for a task at work [4], and a 2024 Digital Education Council found that 86% of higher education students surveyed used AI to complete tasks at school – with 54% of respondents using AI at least weekly [5].

As these tools continue to become more widespread, AI literacy has been identified as a key workplace skill by the World Economic Forum [6]. For engineers, AI literacy is arguably even more important as engineers will play an increased role as not only AI consumers, but also AI contributors [7]. Therefore, within engineering education, there is a growing call for development of curriculum that promotes AI literacy. Some educators have even developed frameworks for teaching machine learning (ML) and AI to engineering students such as the Conceptual Framework for Teaching Machine Learning that breaks machine learning concepts into data, tasks, algorithms, and assessment [8] and the Machine Learning Education Framework that describes relevant knowledge, skills, and attitudes that should be taught about ML/AI [7].

One of the key skills identified in the Machine Learning Education Framework [7] is "ML Problem Scoping", or determining if ML/AI is an appropriate tool for the problem – a skill that many students struggle with. Although ChatGPT and other GAI tools have many strengths, they also have (sometimes unexpected) weaknesses. A variety of publications document some of these weaknesses both generally (e.g. Borji's "Categorical Archive of ChatGPT Failures" [1]) and within the context of engineering education (e.g. Nikolic et al.'s assessment of ChatGPT on various engineering assignments [9]). Specifically, although these tools perform well in many contexts, they can also "hallucinate" information, perform computational errors, and integrate technical information poorly. Because of these performance challenges, Kasneci et al. identify "strong fact-checking strategies" as one of the key features of good AI-embedded education [10].

Although a variety of work has been done to "fact-check" and assess the performance of GAI tools broadly, less work has been done to explore people's perceptions of that performance, particularly in the context of teaching and practicing engineering. Understanding perceptions of GAI performance is an important step in creating impactful learning experiences to improve GAI literacy for students, faculty, and industry professionals. Therefore, this paper aims to fill this gap by describing the development of a GAI literacy activity, reporting on perceptions of GAI performance from a pilot group, and using these findings to make recommendations about how to develop and implement GAI literacy activities for a variety of audiences.

Methods

To accomplish our study purpose and align our methods with the study goals, we developed the following Research Questions (RQs):

> **RQ1:** How can a performance evaluation activity be implemented to help students, faculty, and industry professionals grow in their GAI literacy?
> **RQ2:** What findings were discovered when implementing this activity with pilot groups of students, faculty, and industry professionals?

Participants

This study involved participants from three groups: students, faculty, and industry professionals. The student participants (n = 23) were enrolled in the Department of Integrated Engineering, which emphasizes project-based and cooperative learning. The program allows students to customize their education across disciplines such as mechanical, electrical, biomedical, and computer engineering. At the time of the study, all student participants were taking a 1-credit course titled *Machine Learning for Engineers*, offered in April 2024. The faculty participants (n=18) were faculty from a variety of disciplines who were enrolled in a professional development course about the use of generative AI during September 2024. For both the students and instructors, the activities were completed early in the course as part of the course activities and assignments. Secondary data was collected from these assignments for this study following an approved Institutional Review Board (IRB) protocol. The industry professionals (n=14) were recruited via an email that was distributed to existing industry contacts following an approved IRB protocol. Their participation was voluntary and occurred during December 2024.

Description of the GAI Literacy Activity

Data collection activities for this study included an activity with two components: 1) a survey that asked participants to rate their confidence that ChatGPT would be able to perform a given task, and 2) an evaluation phase where participants assessed ChatGPT's performance in completing these tasks. These activities were designed to improve GAI literacy while also building the 3 Cs of the entrepreneurial mindset (connection-building, curiosity, and creating value) [11].

The survey portion of the activity included ten different tasks for participants to evaluate. For each task, participants were asked to rate their confidence that ChatGPT would be able to complete the task on a scale from 1-5 with 1 being "not confident at all" and 5 being "very confident". Participants were instructed to not complete any testing during this phase. Further, these tasks were designed to be relevant to the participant group, so each of the three groups had slightly different questions as shown in Appendix 1.

The evaluation phase then included survey prompts that asked participants to test how well ChatGPT was able to complete each of the ten tasks. The students and faculty completed this activity in groups during a course or professional development meeting time, and the industry professionals completed this phase individually as the second part of the survey. For each of the tasks, participants evaluated the success as “successful”, “conditionally successful”, or “not successful” and were asked to provide their rationale for their evaluation.

Quantitative Analysis

To allow the research team to consider the results of the quantitative Likert survey responses within and across groups, we began the quantitative analysis by grouping like items in the survey activities. For example, the second item (Q2) generally asked participants to identify their confidence in ChatGPT to provide useful citations. To contextualize this general question: “How confident are you that ChatGPT can give you citations that you can use for your work?” for each participant, we adapted the discussion of work to their specific tasks. Specifically, Q2 was specified as: “How confident are you that ChatGPT can,” “... give you citations that you can use for your DLA (final project),” for the students; “give you appropriate academic citations that you could use for a paper about the ethics of AI,” for the faculty; and “give you correct and relevant citations you can use for your work,” for the industry partners: Consolidations across all survey items are depicted in Appendix 1. Not all survey items were applied in all groups (e.g., Q13 was only given to the student participants).

To compare participants’ subsequent evaluation of the GAI’s accuracy in completing the given task within and across groups, we assigned numerical values to the “successful”, “conditionally successful”, or “not successful” responses. Specifically, “successful” evaluations were assigned a five, “conditionally successful” responses were assigned a 3, and “not successful” responses were assigned a 1. This transformation allowed the data to be plotted and compared to the confidence level responses.

Qualitative Analysis

To identify qualitative trends in the data across survey items, we began by conducting independent inductive descriptive coding of the full set of qualitative survey data [12]. Specifically, researchers read participants’ responses across the data and recorded a) emergent descriptive codes, and b) excerpts which demonstrated those descriptive codes. After this independent process, the researchers then independently identified potential themes across their own descriptive codes.

Once each member of the research team had identified their own descriptive codes and emergent themes, the research team met to collaboratively consolidate and refine key themes in the context of the RQs. Finally, a single set of themes and codes within those themes was identified. While we did not iteratively collect data to ensure saturation, we believe this process of identifying key themes is adequate for this exploratory study.

Results
We distributed the survey invitations to industry members, students, and faculty members in accordance with an approved IRB protocol. From these groups, we received valid responses from 14 industry members, 23 students, and 18 faculty members yielding a total of 55 participants. Results are both quantitative and qualitative in nature, and we begin by presenting the quantitative survey results.

Quantitative Survey Results
To identify trends within and across the three participant groups' responses, we present the quantitative survey results through box and whisker plots where the solid horizontal line represents the median, boxes represent the first to third quartile (Q1 to Q3), and the whiskers represent 1.5 times the Interquartile Range (IQR). To maintain clarity, no outliers are presented (outlier are values less than Q1-1.5*IQR or greater than Q3+1.5*IQR). Figure 1 demonstrates through this method participants' confidence in their selected GAI resource to accurately answer each question. Figure 2 demonstrates through this method participants' evaluation of their selected GAI accurately answering each question according to participants' own prompting and evaluation of the response.

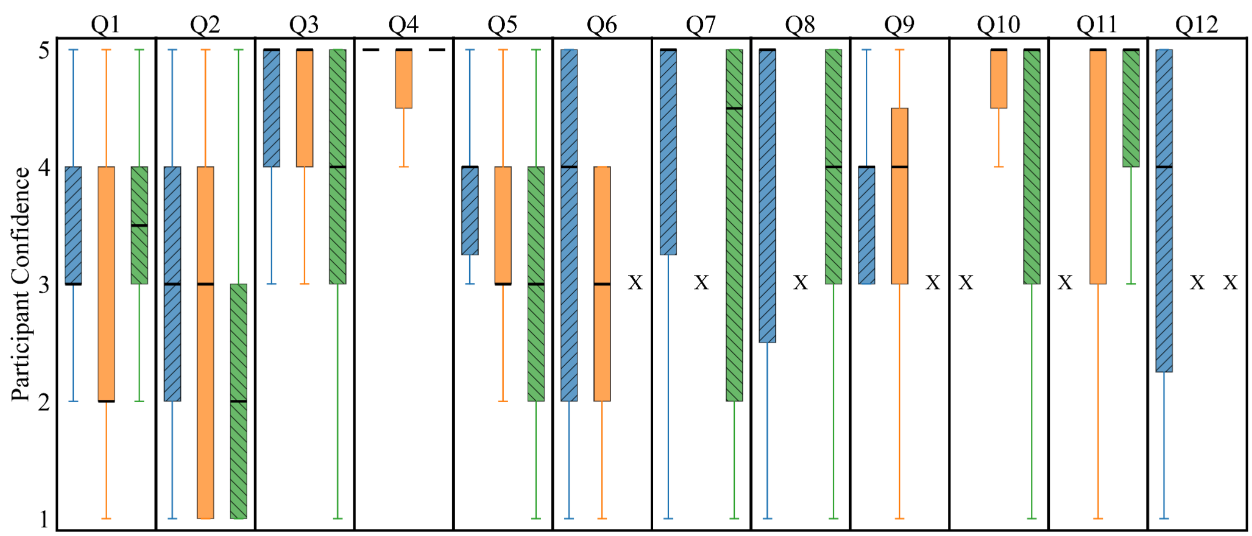

Figure 1. Participants' confidence in GAI accurately accomplishing a given task accurately from 1 (not successful), 3 (conditionally successful) and 5 (successful) by industry (blue, upper left to lower right cross hatch), student (orange), and faculty (green, lower left to upper right cross hatch) groupings. Questions without respective data are indicated by an "X."

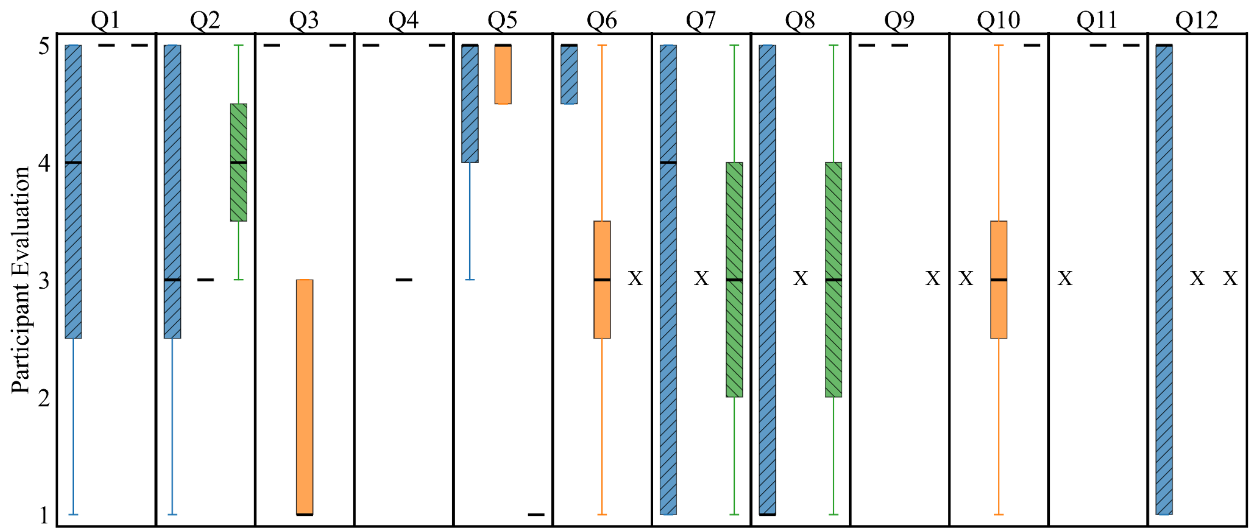

Figure 2. Participants' evaluation of GAI accurately accomplishing a given task accurately from 1 (not successful), 3 (conditionally successful) and 5 (successful) by industry (blue, upper left to lower right cross hatch), student (orange), and faculty (green, lower left to upper right cross hatch) groupings. Questions without respective data are indicated by an "X."

Apart from independent descriptions of the participants' expectations and evaluations, a key area of interest is those items which have an evaluation which did or did not align well with participants' confidence. To identify these differences, Figure 3 demonstrates the participants' median evaluation in the y-coordinate against the participants' median prior confidence in the x-coordinate according to participant groups. To add clarity, the locations of those items which have an absolute difference in the median confidence and evaluation values are indicated by the item number (i.e., Q#) rather adjacent to the circular point.

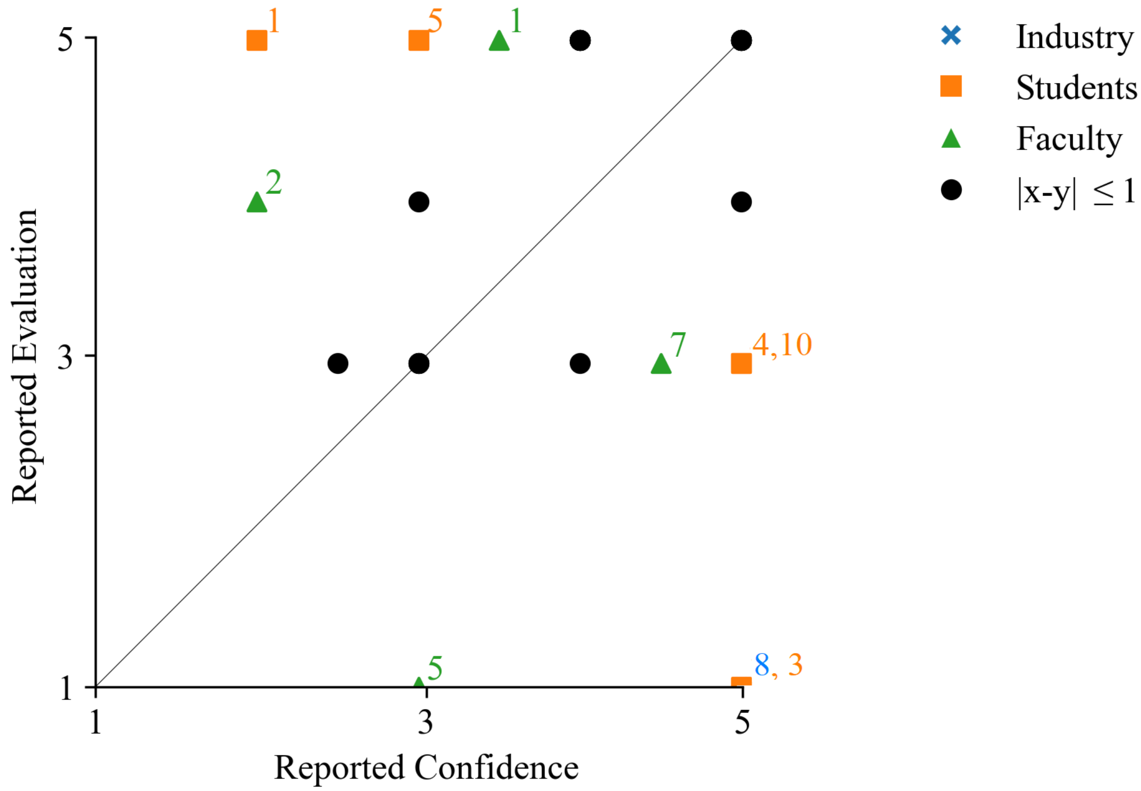

Figure 3. Industry (blue), student (orange), and faculty (green) participants' median evaluation of GAI accurately accomplishing a given task vs. participants' median confidence in GAI to accurately accomplish a given task. Questions with an absolute evaluation and confidence difference greater than one (1-5 scale) are indicated by the item number next to the point. All other observations are black and their numbers are omitted for clarity.

<u>*Qualitative Survey Results*</u>
After analyzing the artifacts created during the activities, we identified six main themes.

1. Prompt development strategies and reflection - Overall, participants recognized the importance of good prompting; one student participant noted that "having bad or vague inputs or questions may cause it to respond with bad answers since it just responds with the statistically most likely thing," and an industry participant noted that "ChatGPT is sensitive to how you input the question and needed data." However, ability varied among participants to develop prompts that meaningfully evaluated the given task.

2. Response evaluation strategies - Participants also varied in their strategies for evaluating and verifying responses. Overall, there seemed to be a general difference in the willingness to fact-check through a primary source across different groups. Particularly, the student group had a number of responses which they rated as correct, but did not include evaluation of the answer (i.e., prime minister of UK, citations, and close-ended problems). For example, for one group, ChatGPT responded “As of my last update in January 2022, the Prime Minister of the United Kingdom was Boris Johnson. However, please verify with a current and reliable source to confirm the current Prime Minister,” and the student participants subsequently marked the response as correct – even though the current Prime Minister was Rishi Sunak. Similarly, one student group tested ChatGPT’s ability to solve an integral by simply asking it, “Can you take the integral of a function?” and trusting it when it responded that it could. Some students did participate in more thorough fact checking, but a lack of fact checking was more widespread in the student group than in the other groups. Industry professionals, on the other hand, seemed more likely to verify their evaluation using primary sources such as Wolfram Alpha to check the integral and Google to fact check information.

3. Hallucinations, especially in the context of citations - Overall, participants from all groups were more skeptical about the citations ChatGPT was providing – with a handful of students and industry professionals noting personal experiences in their evaluations. For example, one student reported that it “Gave me some general citations that existed, but from personal experience we know that it gives some inaccurate information when asked for specifics.” An industry professional reported that “We have had some issues where the citations were either weak or incomplete for what we would feel comfortable using. However, this has improved significantly even since earlier this year.” Another industry professional wrote that “It looked good but I don't have time to double check (no one would ever question it even if it was wrong).” The increased skepticism for citations may be because this is a task that participants would likely be conducting, and many may be aware of consequences of incorrect citations.

4. Perception of task complexity and its impact on answer quality - Many respondents rationalized results by explaining that the performance was dependent on how complex the task is. One student wrote “It works for simple situations” in response to the integral question. Similarly, an industry professional wrote “For simple problems with simple data sets it would excel. For complicated problems with a complex data set it will falter.” However, some realized the results did not align with these initial perceptions about complexity and performance. As one industry professional stated “On the multiplication: I thought that this would be a simple exercise, but it was not. I tried entering the numbers in a variety of formats, but the response returned was highly inaccurate.” However, in some cases, participant confidence was surprisingly low for “simple” tasks. For example, the confidence level that ChatGPT would be able to explain what [PROGRAM] (the engineering program that the students are in) was relatively low – even though the response can be found with a simple Google search.

5. Comparing the GAI tool to a search engine - Many participants were surprised that the GAI was not always able to answer questions about things that occurred after 2021. One student respondent noted that they “thought it auto updated, but information is outdated - no data from the last 2 years. As fast as stuff moves now, it could be missing quite a bit.”

6. Varying levels of performance of the tool - Not all differences in evaluation could be described just by the way the user prompted the tool or what version of the tool was being used. For example, when prompting ChatGPT to tell them the number of Ls in the word “pickleball”, it sometimes responded correctly with “three” and sometimes responded incorrectly with “two”.

Discussion

The results and analysis help us respond to our original research questions. RQ1 was “How can a performance evaluation activity be implemented to help students, faculty, and industry professionals grow in their GAI literacy?” The Methods section of this paper presents information about how the activity was implemented, and the results and analysis led to additional takeaways about activity implementation. These takeaways are presented later in the Discussion. RQ2 was “What findings were discovered when implementing this activity with pilot groups of students, faculty, and industry professionals?” The Results section of this paper presents a variety of quantitative and qualitative observations about the results of a pilot implementation with the three groups, and analysis of these results can provide insights to any user of a GAI tool. These insights are also presented later in the Discussion.

Takeaways for those looking to implement the GAI Literacy Activity

Based on the findings from the pilot implementation of the activities, we present three main takeaways for those looking to implement this GAI literacy activity with students, faculty, and/or industry professionals: choosing prompts, creating opportunities for sharing, and following up with feedback.

Choosing prompts: Although the prompts chosen for this pilot study can be a great starting point for those looking to implement a similar activity, we encourage those implementing this activity to focus more on the principles of how good prompts should be chosen. We used the framing of the Entrepreneurial Mindset 3 Cs to identify prompts: curiosity, connections, and creating value [11]. First, tasks should promote curiosity by choosing tasks that may lead to unexpected results or findings. For example, multiplying seven two-digit numbers feels like an easy task, but GAI tools often (surprisingly) do not generate the correct answer. These questions spark curiosity among participants about why their initial expectations do not align with the results. Second, tasks should allow students to identify connections between their results and other key learning outcomes about GAI. For example, if you want participants to better understand how training data impacts the results of an AI tool, choose tasks that highlight how GAI responds to things it has not been trained on (e.g. current events). Finally, tasks should help participants think about where GAI tools can either create new value or not create new value for them in their roles, so tasks can be chosen that relate to work that the participants do (e.g. writing, calculations, writing code, etc.)

To refine those tasks, we encourage preemptively testing the prompts to find things that may be unexpected. Because the performance is always changing, social media can also be a fun and helpful resource for finding “ChatGPT fails”. The “Ls in pickleball” prompt was inspired by a popular circulating post about ChatGPT not being able to answer how many “Rs” are in the word “strawberry”.

Creating opportunities for sharing: Because the performance of these tools on these tasks can vary (even if the prompting and tool was the same), we recommend offering time for participants to review the results from other groups. For example, some participants got a correct answer on “number of Ls in Pickleball”, and others got an incorrect answer. For those who got the correct answer, it may be easy to just move on thinking that it was an easy task for the tool to complete. By engaging with other group’s results, participants can see how complex it is to determine if a GAI tool can or cannot do something. Rather than viewing it as a binary assessment for each task (successful or not successful), participants can adjust their thinking to see assessment as an evolving process.

Following up with feedback: As the facilitator of the activity, it is important to follow up with feedback for the participants and provide meaning to some of the findings. Many of the participants in the pilot iteration of the activity tried to provide their own meaning to the results they were seeing – with varying levels of correctness. In addition to making meaning out of the results, it is also important to work together to identify best practices for continuing to fact-check and assess responses from GAI tools. Because these tools are evolving and changing, this is not meant to be a one-and-done exercise; rather, it is meant to build habits and understanding that can be continued. The facilitator plays an important role in framing how the activity should be used and interpreted.

Takeaways for those using or teaching about GAI more broadly

For those learning about AI or machine learning applications, it can be helpful to frame the application in the context of the machine learning framework [8]. This framework has four components: data, task, algorithm, and assessment. By describing GAI in the context of this framework, we offer takeaways that not only lead to better awareness about the affordances and limitations of GAI tools, but also how the technical features of the tool can explain some of these affordances and limitations.

Data: Data refers to the type of information that an AI solution works with – in the case of GAI solutions, this information is natural language. All GAI tools have some sort of training data and in this case a library of text was used to train the model. ChatGPT, for example, is trained with publicly available information that was available before 2021 [13]. ChatGPT can still answer questions related to things after 2021, but the model does this by accessing the internet for questions that suggest internet access may be necessary. For example, at the time of this writing, ChatGPT doesn't have training data that includes who the current Prime Minister of the United Kingdom is. However, ChatGPT is trained to access the internet to find this information when asked. In addition, OpenAI states that "information that our users or human trainers and researchers provide or generate" can be used when generating a ChatGPT response. To responsibly use a GAI tool, it is important to have an understanding of what training data was used to train the tool, as well as how the tool uses other plug-ins to do things like access the internet or run Python code to give a better answer [13].

Task: An AI solution's task is the thing that is being trained to do. In the case of ChatGPT and other GAI tools, it is trying to predict the next word it should write when given all of the context of the prompting before it. Ultimately, ChatGPT is a *language* model, meaning it is trained to create human-like language. ChatGPT does not "think" or "understand" in the same way a human does, especially when considering computational or problem-solving tasks. Instead, it is using context and its approximation of language to predict what word comes next. Although this approximation can deliver amazing results in some contexts, it should be noted that it is designed to complete language tasks, not necessarily computational or reasoning tasks. By understanding the intended use of these tools, we can better understand when it is appropriate to use them.

Algorithm: The algorithm is the computational process followed to complete the task. GAI tools use transformers, a type of neural network that uses parameters to numerically describe the meaning of words, as well as the meaning of words used together. These parameters then allow the algorithm to predict which word will come next. However, these algorithms are not deterministic. Randomness is built in through a concept called "temperature", or how often the model should choose a word which is not the top predicted next word. This is why the same person asking the same question at the same time to the same tool might still get a different response. Recognizing this level of variability not only across tools, but within the same tool, is a key thing to consider when using GAI tools for education. A key consideration is that the answer generated by these tools may change for the educator and students, even on the same day.

Assessment: The final component of the framework is assessment, or how we are measuring if an AI algorithm is successful. Because assessment of GAI responses is extremely subjective, it can be complicated to determine when a GAI response is "good." Not only is a "good" response accurate, but it is also resistant to bias, coherent, and relevant. Ultimately, GAI tools are instructed to help the user, and a key challenge is balancing these sometimes competing constraints. For example, sometimes incorrect information is presented confidently in an attempt to provide the user with what they are looking for. Some attempts have been made to write system prompts (i.e. the instructions that these tools use for how to respond) that give caveats about performance related to some types of tasks, these caveats have not been added for every type of error, and they sometimes do not appear even for known errors. Although GAI tools as a whole are being refined through assessment, that does not mean that any given response has been verified or assessed. Therefore, fact-checking responses from GAI tools is an imperative part of the process – especially when using these tools to create reports, lessons, or or content that will be shared with others.

Limitations and Future Work

Because this work was a first exploratory attempt at implementing an activity for teaching GAI, there are limitations and opportunities for future work to determine the generalizability of the findings. One key limitation is the evolving nature of ChatGPT and the challenge of constraining what version someone is using – even throughout a session. In addition, we did not administer a pre-survey to determine participants' existing familiarity with generative AI. Since someone who uses ChatGPT daily will inevitably approach it with different expectations than someone new to such technologies, the absence of baseline data introduces pre-existing bias. Another important consideration is the rapidly evolving nature of AI tools, which changed significantly between 2023 and 2024 when the study took place. Even minor updates can alter capabilities and user experiences, meaning that participants in one cohort could have been interacting with a substantially different version of ChatGPT or Copilot. Finally, the range of tasks used in our study (e.g., coding challenges, doing math or generating research citations) only partially reflects the broad spectrum of AI applications in academia and industry. As a result, our current findings might not fully capture how attitudes toward AI tools could shift if participants were asked to solve more creative, high-stakes, or interdisciplinary tasks.

Moving forward, future work should incorporate a pre-survey on participants' AI usage, track the specific versions of the tools tested, and expand the types of prompts into more meaningful task-driven ones. By further extending this work, we can develop more nuanced guidelines and educational strategies for teaching AI literacy, and ensure all GAI users have a good ethical standing when using AI tools.

Conclusion

This work presented an activity for improving GAI literacy, the results of a pilot group of students, faculty, and industry professionals completing the activity, and a list of takeaways for those interested in implementing the activity and for those who use GAI tools. These tools are rapidly evolving and changing – meaning the fact-checking and assessment of responses will be an ongoing process for anyone using the tools. Activities like the one presented can provide structure for engineering students, faculty, and industry professionals as they engage in this process.

References

[1] A. Borji, “A categorical archive of chatgpt failures”, *arXiv*, 2023.

[2] W. X. Zhao et al., “A Survey of Large Language Models”, *arXiv*. 2024.

[3] M. Haefele, “Has the AI rally gone too far?,” *UBS CIO Daily Updates*. https://www.ubs.com/global/en/wealthmanagement/insights/chief-investment-office/house-view/daily/2023/latest-25052023.html (accessed 2025).

[4] C. McClain, “Americans’ use of ChatGPT is ticking up, but few trust its election information,” *Pew Research Center*, Mar. 24, 2024. https://www.pewresearch.org/short-reads/2024/03/26/americans-use-of-chatgpt-is-ticking-up-but-few-trust-its-election-information/

[5] Digital Education Council, “Digital Education Council Global AI Student Survey 2024,” *DEC Global Survey,* 2024. https://www.digitaleducationcouncil.com/post/digital-education-council-global-ai-student-survey-2024

[6] World Economic Forum, “The role of AI in Education 4.0,” *World Economic Forum Reports,* 2024.

[7] N. Lao, “Reorienting machine learning education towards tinkerers and ML-engaged citizens”, Massachusetts Institute of Technology Cambridge, MA, USA, 2020.

[8] L. Singelmann and J. Covarrubias, “A Framework For Teaching Machine Learning For Engineers”, *SEFI 2023 Annual Conference Proceedings,* 2023.

[9] S. Nikolic et al., “ChatGPT versus engineering education assessment: a multidisciplinary and multi-institutional benchmarking and analysis of this generative artificial intelligence tool to investigate assessment integrity”, *European Journal of Engineering Education*, vol. 48, no. 4, pp. 559–614, 2023.

[10] E. Kasneci et al., “ChatGPT for good? On opportunities and challenges of large language models for education”, *Learning and individual differences*, vol. 103, p. 102274, 2023.

[11] J. S. London, J. M. Bekki, S. R. Brunhaver, A. R. Carberry, and A. F. McKenna, “A Framework for Entrepreneurial Mindsets and Behaviors in Undergraduate Engineering Students: Operationalizing the Kern Family Foundation’s" 3Cs"”, *Advances in engineering education*, vol. 7, no. 1, p. n1, 2018.

[12] J. Saldaña, “Coding techniques for quantitative and mixed data”, *The Routledge reviewer’s guide to mixed methods analysis*, pp. 151–160, 2021.

[13] OpenAI, “How ChatGPT and our foundation models are developed”, OpenAI Help Center,” 2024. https://help.openai.com/en/articles/7842364-how-chatgpt-and-our-foundation-models-are-developed

Appendix 1. Tasks Assessed

Question	Code	Faculty Specific	Industry Specific	Student Specific
Tell you the mission, vision, and core values of [Insert Institution]?	Q1	Tell you the mission, vision, and core values of [INSTITUTION]?	Tell you the mission, vision, and values of your company?	Explain what [PROGRAM] is?
Give you citations that you can use for your work?	Q2	Give you appropriate academic citations that you could use for a paper about the ethics of AI?	Give you correct and relevant citations you can use for your work?	Give you citations that you can use for your DLA?
How confident are you that ChatGPT can tell you when [Famous person] died?	Q3	Tell you when Jimmy Buffett died?	Tell you when James Earl Jones died?	Tell you when Jimmy Buffett died?
Tell you who the President of the United States is?	Q4	Tell you who the President of the United States is?	Tell you who the President of the United States is?	How confident are you that ChatGPT can tell you who the President of the United States is?
How confident are you that ChatGPT can write working Python code that can complete a task?	Q5	Write working Python code that counts how many Ls are in the word PICKLEBALL	Write working Python code that can complete a task?	Write working Python code that completes a given task?
Answer an FE practice question correctly?	Q6	n/a	Answer a question on the Fundamentals of Engineering exam?	Answer an FE practice question correctly?
Tell you how many L's are in the word PICKLEBALL?	Q7	Tell you how many L's are in the word PICKLEBALL?	Tell you how many Ls are in the word PICKLEBALL?	n/a
Multiply a list of 7 different 2-digit numbers?	Q8	Multiply a list of 7 different 2-digit numbers?	Multiply a list of seven 2-digit numbers?	n/a
Take the integral of a function?	Q9	n/a	Take the integral of a function?	Take the integral of a function?
Tell you who the Prime Minister of the UK is?	Q10	Tell you who the Prime Minister of the UK is?	n/a	Tell you who the Prime Minister of the UK is?
Tell you what day the 2024 solar eclipse was?	Q11	Tell you what day the 2024 solar eclipse was?	n/a	Tell you what day the 2024 solar eclipse was?
How confident are you that ChatGPT can tell you the weather?	Q12	n/a	Tell you [recent event]?	n/a
Tell you what a connect-the-dots image is of	Q13	Tell you what a connect-the-dots image is of	n/a	n/a