Work-In-Progress: Students' Prompting Strategies When Solving an Engineering Design Task

David Reeping
Engineering and Computing Education
University of Cincinnati
Cincinnati, OH
https://orcid.org/0000-0002-0803-7532

Aarohi Shah
Engineering and Computing Education
University of Cincinnati
Cincinnati, OH
shah3ao@mail.uc.edu

***Abstract*— This work-in-progress research paper investigates how students prompt generative AI while tackling an engineering design challenge. As tools like ChatGPT become common in education, we must understand how to incorporate them into our teaching and guide students on their proper use. Although the existing literature focuses on understanding what students use generative AI tools for, less work has been done to examine students' prompting strategies when engaging with these systems. We leveraged a new brainstorming assignment in a first-year engineering course at a large Midwest public university, where students ideated collaboratively using ChatGPT for the design of their term project – a semiautonomous robot (n = 97 teams, 589 prompts). Most prompts (~50%) ranged between 55 and 95 characters, or approximately 8 to 19 words. Our initial qualitative findings suggest that their approaches center on seeking information, much like how they would use a search engine. Others directly ask the chatbot to provide alternatives for their design without providing the appropriate criteria or constraints. A subset of prompts had an evaluative component, asking ChatGPT to weigh ideas against one another. We found that 11% of prompts included instructing ChatGPT to produce the "best" solution instead of generating multiple ideas, suggesting students focus on using the chatbot in a more convergent design process - unlike divergent thinking in brainstorming.**

Keywords—first year experience, ideation, prompting, generative AI, ChatGPT

I. Introduction

With the rapid adoption of generative AI platforms like ChatGPT (undergirded by large language models, LLMs), there is a growing need to effectively integrate them into educational settings and instruct students on their appropriate use [1], [2], [3]. Incorporating these tools in engineering education specifically can transform how we approach teaching many of the core concepts in the curriculum including, mathematics, programming, and – of particular note in this paper – the design process. LLM-driven technology can augment traditional design practices by enabling students to brainstorm, evaluate, and refine ideas more efficiently using tools like ChatGPT as a co-designer. Despite these promising possibilities, there remains a gap in understanding students' actual prompting strategies when tackling engineering design tasks and how these strategies affect their problem-solving approaches.

II. Research Aims

The central research question guiding this study is: *What prompting strategies do students use when interacting with ChatGPT for engineering design tasks?* By investigating how students interact with ChatGPT while working on a specific design problem – ideating for a term project to develop a semiautonomous robot – we aim to uncover specific prompting behaviors that can inform better instructional strategies for integrating generative AI into engineering education.

III. Background

Despite the avalanche of work attempting to weave generative AI tools like ChatGPT into various aspects of education and detailing concern about students' improper use of the tool, there is surprisingly little work detailing how students prompt these systems. Much of the current literature about large language models in education concerns perceptions about the technology and surveying students on their use cases, of which there are too many to list (e.g., [4], [5], [6], [7]). Studies examining students' interactions that go beyond self-reporting strategies, perceptions, or preferences are far more scarce. From the limited evidence thus far, it is becoming apparent that much can be improved with respect to students' prompting strategies.

For example, Zamfirescu-Pereira et al. [8] tasked participants, some of whom had no prompt design, programming, or machine learning experience, to modify a chatbot to act like a professional chef guiding the chatbot's user through a recipe using only prompts (i.e., no-code) and evaluate their prompts systematically. Participants were biased toward engaging with the LLM using prompts that overgeneralized failures when a single prompt did not work, incorrectly assuming that the chatbot could not perform the request. The participants also focused on using direct instruction without providing examples to contextualize what they meant by their instruction. Moreover, the authors reported that the participants had difficulties evaluating their prompts, even ignoring a prompt testing feature in the interface. These participants were primarily graduate students and professionals. So, how would undergraduate students fare considering the graduate students and professionals in [5] struggled?

One example of a study involving undergraduate students was Denny et al. [9], who proposed a new type of programming exercise called prompt problems administered through a webbased tool called *Promptly*. The premise of a prompt problem involves providing the student with a stimulus (e.g., a screenshot of “Enter your name: Bob” Hello Bob) and an incomplete prompt such as “Write a Python program that.” The goal of the prompt problem is to have the student extract the salient details of the task so that they can communicate the desired output to ChatGPT, which, compared to [5], is a much simpler task than guiding the redesign of a chatbot. Still, to the authors’ surprise, graduate students in their pilot study struggled with the prompt problems. Moreover, they observed students in introductory computer science courses writing inefficient prompts, repeating the problem statement verbatim to *Promptly*, and misinterpreting the goal of the problem. These pitfalls align to some extent with Krupp [10] in the context of collaborative problem-solving for Physics with ChatGPT; specifically, they similarly noted undergraduate students using a copy-and-paste approach to prompting for most of the interactions (42% of the time). However, in a positive sense, some students also modularized the problem into chunks to feed into the LLM and built off of ChatGPT’s response by asking clarifying questions or correcting mistakes. Both of these studies reiterate a common narrative of potential overreliance on LLM-based tools; if the user does not have the necessary content knowledge to evaluate the output, then the polished nature of the LLM’s may be *correct enough* to convince the user of its legitimacy [11].

The prospects of collaboration with LLM-based tools are not entirely negative. For example, Han et al. [12] curated a dataset of interactions from 213 university students in an English as a Second Language (ESL) course. In the context of essay writing, the authors reported that the students’ essays improved across all factors they considered when students interacted with ChatGPT through a platform called Revising an Essay with ChatGPT on an Interactive Platform for EFL learners (*RECIPE)*. As perhaps one of the few examples of public distribution for student use of ChatGPT-like tools, their dataset of interactions with ChatGPT (n = 4330 utterances, 1913 of which are student utterances) is available online and supplemented by intent labels, meaning each interaction is coded with one of 13 predefined categories such as request for revision and request for generation. These efforts motivate our study, which involves forming a dataset in the engineering design context to explore how students collaborate with LLMs to address engineering design tasks.

IV. Research Design

To elicit student prompts for an engineering design problem, we implemented a new brainstorming assignment in the introduction to engineering course at the University of Cincinnati, a large public Midwest institution. We analyzed the assignments using a qualitative perspective most aligned with Charmaz’s constructivist grounded theory [13]. A grounded theory approach was chosen because there are no widely accepted theories of prompting to note with respect to prompt structure or intent, especially not for co-designing. We do acknowledge there are techniques such as chain-of-thought prompting, which involves providing examples of the desired reasoning in the initial prompt [14], and *according to* prompting, which intends to reduce fabrication tendencies in LLMs when querying for factual information [15]. One possible framework for classifying prompts exists called TELER (**T**urn, **E**xpression, **Le**vel of Details, and **R**ole), which categorizes prompts into seven levels based on the composition of the instructions (e.g., breaking the goal of the prompt into subtasks, asking the LLM to evaluate its output, and examples of the desired output) [16]. The first level of prompting (Level 0) is simply a user inputting data without any directions. In contrast, a prompt in the final level (Level 6) includes a description of the high-level goal, a bulleted list of subtasks, examples for the LLM to use for its output, information gathered using different retrieval-based methods, and a statement asking the LLM to explain its output. However, this taxonomy was not helpful for our purposes because of the length of the prompts in our dataset, as we will discuss in our preliminary findings. Thus, in light of the lack of available apriori frameworks, we built the patterns from the data themselves.

A. Data Collection

In Fall 2023, all students (N = 1570) in an introduction to engineering course were assigned a semester-long robotics project. The project tasked teams of three or four students to design and build a semiautonomous robot capable of performing a series of tasks, including line following, maneuvering over obstacles, picking up a container they could design filled with three gradations of weights, and transporting the container to a new location. The students were provided with a LEGO Mindstorms kit and could not ask for additional parts. There were several checkpoints throughout the semester, including an initial demonstration of walking, short video status updates, and a new brainstorming assignment we will focus on here.

The brainstorming assignment intended to have students purposefully develop several ideas for one of their robot’s functions, such as how it walks to avoid obstacles. The first step had students pick a brainstorming method (i.e., hybrid [17], 6-3-5 brainwriting [18], six thinking hats [19], or wrong theory protocol [20]) and implement it with their team to develop a set of ideas. They would then use a morphological chart, a table of different ways to accomplish a product’s functionality, to organize their ideas and develop at least one concept per team member from the chart. Next, they were tasked with creating a set of at least five prompts and interacting with ChatGPT for at least 30 minutes. After their conversation, they would integrate new or refined ideas from ChatGPT into their morphological chart and develop a set of concepts that mesh their original ideas with those created using generative AI. Students answered a series of reflection questions based on their experiences with the first two steps.

For this data collection effort, we retrieved the entire assignment from consenting teams via a file transfer from instructors of each course section after downloading all relevant files from our institution’s learning management system, Canvas. The possible data for analysis include brainstorming evidence (e.g., drawings, transcripts of brainstorming, notes), planned prompts with ChatGPT, conversation with ChatGPT, morphological charts for their traditional and generative brainstorming, and reflections on their experience with the brainstorming sessions (n = 97 teams, 589 prompts). There were 447 teams in total and 97 consenting teams; thus, our response rate was 22%.

B. Analysis

For this work-in-progress paper, we analyzed the set of prompts students planned before engaging with ChatGPT. At this stage, we were most interested in the general trends in how students would think through the prompts they would send to ChatGPT. Therefore, we extracted the prompts from each assignment and collected them in a common spreadsheet. Each prompt set was associated with a study team ID to cluster prompts that built upon one another within each team.

We analyzed the prompts using the two-stage coding approach described by Saldaña [21]. Four research team members were tasked with coding the prompt data by examining common elements that could form salient codes, such as the use of different question words or the goal of the prompt, and comparing each prompt against one another sequentially – i.e., the constant comparative method [22]. The team met weekly to discuss the coding process, mimicking the process outlined by Richard and Hemphill [23]. In the second stage, we generalized the common elements into themes by grouping the individual codes during our regular meetings. As part of our coding process, we also examined the length of the prompts to determine if we could break prompts into categories based on the number of characters used to engage with ChatGPT. We present our preliminary findings in this paper. We plan to expand our analysis to extract additional details like prompting structure and align these insights with students' conversations with ChatGPT.

C. Limitations

We should note that, like previous research [9], [12], these prompts are limited to students from a single institution. Aside from the submissions from consenting teams potentially being different from non-consenting teams, the process used to admit students to the study could also present some selection bias. For a team's assignment to be used in this research, we needed permission from each team member. We could not offer additional incentives to avoid coercion by team members, nor could extra credit be provided because of the coordinated course structure. This additional complexity in getting unanimous consent likely impacted the number of teams we could recruit despite our response rate being a respectable 22%. We also acknowledge that for this particular analysis, the prompts are what the team planned to submit to ChatGPT. In some cases, the actual prompt inputted to ChatGPT differed – these differences are considered in upcoming analyses. Moreover, because the task was unmoderated, there were details of their engagement we could not confirm; for example, we could not be sure that students interacted with ChatGPT for the desired amount of time. We plan to control this factor in future work by moderating the design task with a subset of students.

V. Preliminary Results and Discussion

Considering the intent of the assignment was to practice structured brainstorming, most of the prompts involved direct instructions for ChatGPT to generate ideas. However, we saw considerable variation in how these requests were structured. We found three types of prompting strategies after coding the prompts in our dataset. These included idea generation, evaluation, and information seeking. The three prompt types are summarized in Table I.

TABLE I. Summary of Prompting Strategies

Prompt Type	Example
Idea Generation	What are **some good ideas** for simple bipedal leg movement for a Lego Mindstorm robot?
Evaluation	**Compare** the use of four legs on an EV3 Lego robot to it having six legs.
Information Seeking	Using a Lego EV3, **how can you determine** the weight of an object picked up by the robot?

Although this assignment aimed to have students using ChatGPT for a divergent thinking task – coming up with several ideas in a structured fashion – we should note some teams elected to prompt ChatGPT for the *best idea* instead. About 11% of the planned prompts referred to the best idea for a given function or application; for example, "*best ways for a robot to walk without using wheels or tracks*" and "*what is the best physical design for an EV3 robot to lift up an object?"* This suggests that a minority of the teams intended to use ChatGPT to offload decision-making onto the generative AI system instead of weighing its suggestions against the design criteria and constraints.

To further understand our data, we examined the length of the prompts. Upon analyzing the prompt lengths used by students before interacting with ChatGPT, we observed that the lengths primarily ranged from 55 to 95 characters, or approximately 8 to 19 words in a single sentence. In the TELER framework, the vast majority of prompts would be considered Level 1 (i.e., "simple one-sentence directive expressing the high-level goal" [16, p. 3]) – there was almost no variance in prompt type when viewed from the perspective of TELER's prompt levels. However, when examining the prompts' intent, we see some differences. We will begin with the most frequent theme, idea generation.

A. Idea Generation

As noted, the most common prompt type we observed instructed ChatGPT to generate ideas. For example, a prompt could be as simple as "*potential designs to pick up material bins*" or "*generate designs for robot feet to prevent drifting and provide traction."* Despite the teams all working on the same project, which might suggest the queries would homogenize, the kind of ideas they wanted to generate were diverse. To illustrate, one team wanted to scope out possible functions the robot needed to perform: "*What are some major functions of a robot that is designed to pick up debris after a natural disaster?*" However, another team wanted to troubleshoot a specific issue with their robot's current design, "*The robot is struggling to move with the added weight of the bin. What are some ways we can fix this?"*

A key weakness of most prompts is that they contain little contextual information to help the LLM generate ideas relevant to the team's project. The most salient contextual information is what materials the teams have access to. Each team was given a kit containing various Lego pieces for their project near the beginning of the semester; they were not required to use all the pieces but could not ask for supplemental parts. Therefore, ChatGPT could give them an excellent idea in theory, but it was outside the scope of what could be built using their equipment. However, other teams were savvier and provided additional context to constrain the possibilities that ChatGPT could generate; for example, one team planned the following prompt:

> "Generate 10 different to create locomotion from rotational motion on land *that does not move using wheels using two motors* and *order them by ease of implementation* considering programming aspects as well as building *if it were done with LEGO EV3 components*."

Notice that, unlike the prompt, "*Give me good ideas for a Lego Mindstorms claw*," the team recognized the need to provide constraints for their design, one of which was that their robot needed to walk over obstacles and not roll. This constraint implied they could not use wheels for locomotion. They also tasked ChatGPT with ordering its ideas based on how easy each would be to implement in terms of programming while noting the materials were related to the Lego EV3 set. In the most extreme example of providing ChatGPT with design criteria and constraints, one team went as far as uploading a document provided by the instructors describing the project.

B. Evaluation

Another type of prompting strategy was asking ChatGPT to make some judgment on the team's behalf. These prompts tended to be structured as a comparison; for example, one team planned the following prompt, "*Compare and contrast a 'forklift' method and a 'claw machine' method for a Lego robot lifting bins of materials.*" Other teams combined the process of idea generation with evaluation by asking for a posthoc comparison, such as:

> *"Ideate the method for robot to hold the object after picking it up by different equipment as well as using different force to hold it. Then choose the best idea from those."*

Again, like the prompts in idea generation, a challenge here is not giving ChatGPT enough information to make an appropriate comparison. Take the prompts "*What is an easier shape for a claw, forklift, or conveyer to grab?*" and "*What are the most effective shapes to open and close on objects?"* for instance. Without proper framing, the model is unlikely to answer this question satisfactorily; in particular, there is no indication of what is being grabbed.

Alternatively, the few longer prompts provided much more context using a different approach. Among various prompting strategies in the literature and popular media, assigning a role to a large language model is one of the most popular (e.g., "you are an expert programmer in C#" or "you are a professional writing coach") [24]. Only one team used this strategy in the following prompt style (which also was the longest in our dataset) to create a list of pros and cons for their locomotion design:

> *"Imagine you are a college student at the University of Cincinnati studying Mechanical Engineering. You are taking the ENED 1100 course which is an intro to engineering and design course. For this course you are tasked with creating a Lego Mindstorm Robot using LabVIEW and Python code that will be able to navigate through obstacles and 6 different line types (Straight Dashed, Straight Solid, Straight Dotted, Curved Dashed, Curved Solid, Curved Dotted) by using motors and light sensors. You are in a group of four with three other engineering students taking the same course. With all this in mind, list the pros and cons of each type of locomotion design."*

C. Information Seeking

The final type of prompting style we observed involved students using ChatGPT to find specific information related to the technicalities of implementing ideas for their design. In fact, two prompts made specific reference to ChatGPT searching the internet despite the tool not being able to do so at the time (e.g., *"With all of this new information, are there any other suggestions you can find on the internet that would fit our requirements and improve our design?"*).

Unlike prompts aimed at generating creative ideas or making evaluations, these prompts were designed to elicit factual and direct responses akin to queries posed on a standard internet search engine. For instance, basic prompts observed in the dataset were: "*How dense is plastic/weight of plastic?*" and "*How does a light sensor for the robot work?"* Other prompts were tailored more specifically to how elements of the design would work: "*Using a Lego EV3, how can you determine the weight of an object picked up by the robot?*" A notable prompt that showcased how ChatGPT may be used in a brainstorming workflow requested a synthesis of ideas and a summary of key features for each, illustrated by the following prompt:

> "*have several ideas for my robot's functions, but I need a summarized list of the key features and capabilities for each, such as walking in a straight line, speed, turning, and picking up trash. Can you help me with brief descriptions of the top three ideas, including their unique selling points?*"

VI. Discussion and Future Work

In contrast to the work by the likes of Zamfirescu-Pereira et al. [8] and Denny et al. [9], we explored prompts the students planned to ask ChatGPT as opposed to what they inputted into the system; our next step is to analyze their actual conversations and determine how they used the conversational nature of the tool to dig deeper into their queries. Still, the prompts students planned to give to ChatGPT appear to align with what we have seen in the literature, such as using the tool to explain concepts, as we saw with our information-seeking theme [25].

From this preliminary analysis, we observed a crosscutting theme that the students in our sample often used ChatGPT in an elementary fashion. There's little use of the few "best practices," such as assigning the LLM a role [24] or using chain of thought prompting [14]. Like [8], most of the prompts were direct instruction with no examples for the model to build from. Occasionally, prompts played into the fear that students would use the tool to circumnavigate critical thinking by asking for the best or correct answer, such as one of the prompts that essentially asked ChatGPT to complete the main task for the term project "*Design a robot made from Lego parts that moves without wheels, picks up an object, and can follow a line on the ground with a light sensor*." As noted throughout our discussion, the teams' prompts frequently do not provide enough detail to provide feasible solutions for their project. Therefore, students' understanding of criteria and constraints when co-designing with generative AI may be a potential area for future research.

VII. Conclusion

This work-in-progress paper leveraged actual student prompts to help develop our understanding of how students may use ChatGPT to complete engineering design tasks. As we continue analyzing the data in this assignment, we expect to extract more detailed intersections of their prompting with the engineering design process and identify other areas to guide future curriculum development for what it means to use generative AI in the design process.

Acknowledgments

The authors would like to thank the students who allowed their work to be used in this research. We also thank undergraduate research assistants, Tanishq Jadhav and Dhruv Singh, for their contributions to the analysis.

References

[1] J. Qadir, "Engineering Education in the Era of ChatGPT: Promise and Pitfalls of Generative AI for Education," in *2023 IEEE Global Engineering Education Conference (EDUCON)*, May 2023, pp. 1–9. doi: 10.1109/EDUCON54358.2023.10125121.

[2] M. Menekse, "Envisioning the future of learning and teaching engineering in the artificial intelligence era: Opportunities and challenges," *J. Eng. Educ.*, vol. 112, no. 3, pp. 578–582, 2023, doi: 10.1002/jee.20539.

[3] S. Stankovski, G. Ostojić, S. Tegeltija, M. Stanojević, M. Babić, and X. Zhang, "Generative AI Applications and Tools in Engineering Education," in *2024 23rd International Symposium INFOTEH-JAHORINA (INFOTEH)*, Mar. 2024, pp. 1–4. doi: 10.1109/INFOTEH60418.2024.10495941.

[4] A. Shoufan, "Exploring Students' Perceptions of ChatGPT: Thematic Analysis and Follow-Up Survey," *IEEE Access*, vol. 11, pp. 38805–38818, 2023, doi: 10.1109/ACCESS.2023.3268224.

[5] T. T. A. Ngo, "The Perception by University Students of the Use of ChatGPT in Education," *Int. J. Emerg. Technol. Learn. Online*, vol. 18, no. 17, pp. 4–19, 2023, doi: 10.3991/ijet.v18i17.39019.

[6] H. Singh, M.-H. Tayarani-Najaran, and M. Yaqoob, "Exploring Computer Science Students' Perception of ChatGPT in Higher Education: A Descriptive and Correlation Study," *Educ. Sci.*, vol. 13, no. 9, Art. no. 9, Sep. 2023, doi: 10.3390/educsci13090924.

[7] C. C. Tossell, N. L. Tenhundfeld, A. Momen, K. Cooley, and E. J. de Visser, "Student Perceptions of ChatGPT Use in a College Essay Assignment: Implications for Learning, Grading, and Trust in Artificial Intelligence," *IEEE Trans. Learn. Technol.*, vol. 17, pp. 1069–1081, 2024, doi: 10.1109/TLT.2024.3355015.

[8] J. D. Zamfirescu-Pereira, R. Y. Wong, B. Hartmann, and Q. Yang, "Why Johnny Can't Prompt: How Non-AI Experts Try (and Fail) to Design LLM Prompts," in *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, in CHI '23. New York, NY, USA: Association for Computing Machinery, Apr. 2023, pp. 1–21. doi: 10.1145/3544548.3581388.

[9] P. Denny *et al.*, "Prompt Problems: A New Programming Exercise for the Generative AI Era," in *Proceedings of the 55th ACM Technical Symposium on Computer Science Education V. 1*, in SIGCSE 2024. New York, NY, USA: Association for Computing Machinery, Mar. 2024, pp. 296–302. doi: 10.1145/3626252.3630909.

[10] L. Krupp *et al.*, "Unreflected Acceptance -- Investigating the Negative Consequences of ChatGPT-Assisted Problem Solving in Physics Education," Aug. 21, 2023, *arXiv*: arXiv:2309.03087. doi: 10.48550/arXiv.2309.03087.

[11] A. Azaria, R. Azoulay, and S. Reches, "ChatGPT is a Remarkable Tool—For Experts," *Data Intell.*, vol. 6, no. 1, pp. 240–296, Feb. 2024, doi: 10.1162/dint_a_00235.

[12] J. Han *et al.*, "Exploring Student-ChatGPT Dialogue in EFL Writing Education," presented at the NeurIPS'23 Workshop on Generative AI for Education, 2023.

[13] K. Charmaz, *Constructing Grounded Theory*, Second edition. Los Angeles, Calif.: Sage, 2014.

[14] J. Wei *et al.*, "Chain-of-Thought Prompting Elicits Reasoning in Large Language Models," Jan. 10, 2023, *arXiv*: arXiv:2201.11903. doi: 10.48550/arXiv.2201.11903.

[15] O. Weller, M. Marone, N. Weir, D. Lawrie, D. Khashabi, and B. Van Durme, "'According to ...' Prompting Language Models Improves Quoting from Pre-Training Data," May 22, 2023, *arXiv*: arXiv:2305.13252. doi: 10.48550/arXiv.2305.13252.

[16] S. K. K. Santu and D. Feng, "TELeR: A General Taxonomy of LLM Prompts for Benchmarking Complex Tasks," Oct. 24, 2023, *arXiv*: arXiv:2305.11430. doi: 10.48550/arXiv.2305.11430.

[17] K. Girotra, C. Terwiesch, and K. T. Ulrich, "Idea Generation and the Quality of the Best Idea," *Manag. Sci.*, vol. 56, no. 4, pp. 591–605, Apr. 2010, doi: 10.1287/mnsc.1090.1144.

[18] B. Rohrbach, "Kreativ nach Regen-Methode 635, eine neue Technik zum Losung von Problemen," *Absatz-Wirtsch.*, vol. 12, pp. 73–75, 1969.

[19] E. de Bono, *Six Thinking Hats*, Revised and Updated ed. edition. Boston: Back Bay Books, 1999.

[20] V. Svihla and L. Kachelmeier, "The wrong theory protocol: A design thinking tool to enhance creative ideation," in *Proceedings of the Sixth International Conference on Design Creativity (ICDC 2020)*, 2020, pp. 223–230. doi: 10.35199/ICDC.2020.28.

[21] J. Saldaña, *The coding manual for qualitative researchers*, 3rd ed. Sage, 2016.

[22] B. G. Glaser, "The Constant Comparative Method of Qualitative Analysis*," *Soc. Probl.*, vol. 12, no. 4, pp. 436–445, Apr. 1965, doi: 10.2307/798843.

[23] K. A. R. Richards and M. A. Hemphill, "A Practical Guide to Collaborative Qualitative Data Analysis," *J. Teach. Phys. Educ.*, vol. 37, no. 2, pp. 225–231, doi: 10.1123/jtpe.2017-0084.

[24] I. R. Indran, P. Paranthaman, N. Gupta, and N. Mustafa, "Twelve tips to leverage AI for efficient and effective medical question generation: A guide for educators using Chat GPT," *Med. Teach.*, vol. 0, no. 0, pp. 1–6, 2023, doi: 10.1080/0142159X.2023.2294703.

[25] M. P. Rogers, H. M. Hillberg, and C. L. Groves, "Attitudes Towards the Use (and Misuse) of ChatGPT: A Preliminary Study," in *Proceedings of the 55th ACM Technical Symposium on Computer Science Education V. 1*, in SIGCSE 2024. New York, NY, USA: Association for Computing Machinery, Mar. 2024, pp. 1147–1153. doi: 10.1145/3626252.3630784.