Prompt Potential: A Pilot Assessment of Using Generative Artificial Intelligence (ChatGPT-4) as a Tutor for Engineering and Maths

S. Nikolic [1]
University of Wollongong
Wollongong, Australia
http://orcid.org/0000-0002-3305-9493

A. Heath
University of Wollongong
Wollongong, Australia

B. A. Vu
University of Wollongong
Wollongong, Australia

S. Daniel
University of Technology Sydney
Sydney, Australia

A. Alimardani
University of Wollongong
Wollongong, Australia

C. Sandison
University of Wollongong
Wollongong, Australia

X. Lu
University of Wollongong
Wollongong, Australia

B. Stappenbelt
University of Wollongong
Wollongong, Australia

D. Hastie
University of Wollongong
Wollongong, Australia

[1] *S. Nikolic*
sasha@uow.edu.au

***Conference Key Areas**: 6. Digital tools and AI in engineering education; 4. Teaching foundational disciplines of Mathematics and Physics in engineering education*
***Keywords**: ChatGPT, GenAI, Generative Artificial Intelligence, Tutoring, Mathematics*

ABSTRACT

The meteoric rise of GenAI has caused many educators to be consternated by its potential to undermine assessment. However, there is a more optimistic view to instead focus on the pedagogical affordances that GenAI can bring, for example, in tailoring personalised learning experiences for students. In this pilot study, we investigate ChatGPT-4's potential to act as a one-on-one tutor for engineering and mathematical concepts. We use three research-informed prompt strategies and simulate interactions with high-, mid-, and low-performing students. We find that the learning experience is best tailored to high-performing students. However, to gain comfort in using it, the experience must be error-free. We discovered performance varied by topic, but there indeed are topics that ChatGPT-4 can engage with error-free or with a slight chance of errors.

1 INTRODUCTION

Generative artificial intelligence (GenAI) through ChatGPT-3.5 became a mainstream topic of interest in higher education in early 2023 due to its ability to achieve passing grades in many different assessment tasks (Nikolic et al. 2023). This heightened concerns regarding cheating, plagiarism and academic dishonesty (Mai, Da, and Hanh 2024). Other negative connotations associated with GenAI include bias, lack of transparency, factual incorrectness, and privacy concerns (Ivanov 2023).

While much focus has been placed on the risks, there have been several studies looking at the benefits. A key benefit in higher education is the ability of GenAI to provide personalised learning experiences, with targeted learning content customised for the student, accompanied by practice questions and step-by-step solutions (Menekse 2023). A systematic review by Crompton and Burke (2024) discovered various beneficial use cases for students, including 24/7 support, explaining difficult concepts, conversational partners, personalised feedback and materials, writing support, self-assessment, facilitating engagement, and self-determination. Collectively, the benefits resemble the capability of a student having a personalised tutor by their side, available at their beck and call, and for a fraction of the price. For years, academics have explored technology in education (Gregory et al. 2015), this study examines GenAI's tutoring potential. In recent months, some tutoring implementations have appeared in the literature but are mainly focused on being used to provide hints, feedback or integrated into another system (Phung et al. 2024; Pardos and Bhandari 2024; Frankford et al. 2024), while this study explores a self-contained private tutor experience. However, there are a number of GPTs (plugins) like 'Tutor Me' by Khan Academy available, but the prompting is not open source, and no empirical studies within the context of this study could be found.

The reported risk when using GenAI is its tendency to hallucinate, especially when it comes to referencing (Buchanan, Hill, and Shapoval 2023). While more recent GenAI models such as ChatGPT-4 improved accuracy substantially, accuracy remains dependent on the selected GenAI platform (Nikolic et al. 2024) and prompt engineering (Hebenstreit et al. 2023). Furthermore, the constant updates to AI models and the flawed benchmarking standards currently used by AI companies, such as MATH (Hendrycks et al. 2021), complicate the assessment of these models' risks for educational use (Zhang et al. 2024). Therefore, without selecting the correct GenAI platform and appropriate testing, such an application risks teaching students incorrectly, which could hinder learning, necessitating the pilot study.

In this pilot study, we explore three different prompts designed to provide students with a computer-based tutor experience powered by generative artificial intelligence. The purpose is that a student would be able to copy/paste the prompt and then experience a highly engaging and accurate one-on-one tutor experience provided by GenAI. The pilot is tested within the fields of engineering and engineering mathematics. The content is taken from core courses within the engineering program. Preliminary work exploring the reliability of various GenAI platforms within these fields concluded that ChatGPT-4 was the most suitable (Nikolic et al. 2024). Research assistants are used to simulate the experience to determine its reliability and usefulness before transitioning into a larger study. Therefore, this study aims to answer the research question, "*Using a pre-drafted prompt, can ChatGPT-4 provide a reliable and useful tutor experience for undergraduate engineering students?*"

2 METHODOLOGY

2.1 Preparation

The idea of using a computer to tutor a student instead of a human is not new. Thirty years ago, Merrill et al. (1992) explored the differences between human and computer-based tutoring. The study found that human tutors provide students with key services: scaffolding knowledge, being interactive and providing feedback, monitoring problem-solving, intervening when they get way off track, and helping them detect, locate, and repair errors. These principles have been used to help devise a rubric to test the suitability of ChatGPT-4 and the chosen prompts. The rubric is comprised of seven criteria, each scored on a 4-point scale.

- relevance to the topic area,
- reliability and accuracy,
- pedagogical effectiveness (clear explanations, examples, scaffolding, and assessment techniques),
- interactive engagement,
- progression to more difficult concepts,
- contextual understanding (address the underlying question, learner needs, and anticipate follow-up queries), and
- use of examples and illustrations.

As mentioned in the introduction, ChatGPT-4 was selected due to preliminary work that found it suitable for engineering and mathematics and superior to ChatGPT3.5, Gemini, and Copilot (Nikolic et al. 2024). However, it was found not to be 100% accurate, hence the need for this simulation exercise. While creating a specific GPT environment (formerly known as plugins) for the task is possible, e.g., Assistants API, the intention was to test for the simplest implementation that could be universally adopted by any student across the world with access to ChatGPT-4. No GPTs were used. The prompt developed by Mollick and Mollick (2024) was used for the simulation as a starting point (full prompt available from the reference). This was classified as Prompt 1. Before data collection commenced, the research team ran through multiple simulation exercises to learn and understand how the prompt controlled the behaviour of the output. From this, slight tweaks to the prompt were introduced to alter the experience. Prompt 2 was altered to be more aware of the specific context and focus more on reassuring learning through questioning. Prompt 3 (see Appendix A) was targeted at better pinpointing what the student did or did not know and making the responses succinct, as ChatGPT-4 could over-explain at times. To start the tutorial process, the prompt would be copied into ChatGPT-4, and the user was asked for the content area they would like to learn about, and some positioning questions were asked. A unique tutorial experience commenced from this initiation sequence.

2.2 The Experiment

For the pilot study, two engineering mathematics and two engineering subjects were selected, representing core courses from the engineering program. They were selected as they represent the backbone of future year content. One topic from each mathematics subject was tested, and two topics from each engineering subject. In mathematics, the content areas consisted of *integration* and *multivariable chain rule*. From engineering, the content areas consisted of *fluid pressure on circles*, *squares*, *triangles, Bernoulli's principle*, the *Rankin power cycle*, and *psychrometry charts*. This gave a total of six topics tested.

In our interactions with the ChatGPT 'tutor', we role played interacting with the tutor as a high-, mid-, or low-performing student. That is, we tested three different use cases for each prompt:
Case A: This simulates a high-performance student who knows everything and always provides the correct answers when prompted.
Case B: This simulates a student in the middle of the academic bell curve. A correct answer is provided for 50% of the responses, and a wrong answer for the other 50%.
Case C: This simulates a student who constantly fails. When prompted, a wrong answer is always provided.

Each interaction for each case lasted at least 20 minutes. Many interactions, examples and questions occurred, and scoring was based on this collective experience. Interactions were scored against the 7-criteria rubric outlined in section 2.1, with qualitative notes recorded to document interactions. Examples of interest were also recorded and discussed regularly across the research team, helping to create uniformity.

2.3 Limitations

Only six content areas have been analysed, two for engineering mathematics and four for engineering. Being a pilot study, these topics were selected randomly to discover a broad understanding of capability and cover only a fraction of possible content areas. Therefore, the results only indicate what to expect across these content areas.

The reliability criteria for determining whether the GenAI was 'completely accurate and error-free' were objective and easy to measure. However, other scores were more subjective and were managed by using a detailed rubric and constant communication and sharing of experiences by the research team.

3 RESULTS

The rubric was detailed. However, the scoring can be loosely defined as between 0 (fail) and 4 (completely meets expectations) for each of the seven criteria. Therefore, the maximum achievable score was 28. Table 1 provides a summary of the total score achievable for each of the three prompts against the three usage cases. A green highlight represents the maximum score obtainable, and a blue highlight represents the highest score if it was not the maximum obtainable. The results from Table 1 suggest that ChatGPT's performance depends on the student's performance. ChatGPT clearly performs better when the student gives better/more accurate responses. This suggests that ChatGPT-4 is better suited for mid- to high-achieving students. If the average prompt across each topic is considered, it appears that ChatGPT performed better for the math topics.

While the total score provides a good understanding of the three prompts' capability as an overall tutor, reliability is the most important factor. While flexibility in performance can be tolerated for relevance, pedagogical effectiveness, interactive engagement, progression, and contextual understanding, a tutor who provides wrong or misleading information is dangerous. Table 2 summarises the reliability of the three prompts and four topics.

Table 1. Total Score (Max. 28 in green) for Each Prompt & Case

		Sim: Strong Student			Sim: Average Student			Sim: Weak Student			Average by prompt x topic			Average by topic
		P1 A	P2 A	P3 A	P1 B	P2 B	P3 B	P1 C	P2 C	P3 C	P1	P2	P3	
Math	Integration	28	25	26	26	26	28	25	15	25	26.33	22.00	26.33	24.89
	Multivariable Chain Rule	26	26	28	28	22	26	19	28	26	24.33	25.33	26.67	25.44
Engineering	Fluid pressure on a circle/square/triangle	26	26	26	22	20	23	18	19	23	22.00	21.67	24.00	22.56
	Bernoulli's principle	26	23	24	16	17	15	14	12	18	18.67	17.33	19.00	18.33
	Rankin Cycle	26	22	24	18	21	19	21	20	20	21.67	21.00	21.00	21.22
	Psychrometric Chart	26	28	28	23	19	18	22	26	20	23.67	24.33	22.00	23.33
	Average by prompt x student	26.33	25.00	26.00	22.17	20.83	21.50	19.83	20.00	22.00				
	Average by student	25.78			21.50			20.61			22.78	21.94	23.17	Average by prompt

Table 2. Reliability Score (Max. 4 in green) for Each Prompt & Case

		Sim: Strong Student			Sim: Average Student			Sim: Weak Student		
		P1 A	P2 A	P3 A	P1 B	P2 B	P3 B	P1 C	P2 C	P3 C
Math	Integration	4	1	2	2	2	4	1	2	1
	Multivariable Chain Rule	4	2	4	4	4	2	1	4	2
Engineering	Fluid pressure on a circle/square/triangle	2	4	4	4	2	4	4	4	4
	Bernoulli's principle	4	1	4	1	1	2	1	0	0
	Rankin Cycle	4	4	2	4	4	4	4	4	4
	Psychrometric Chart	4	4	4	4	4	4	4	4	4

Detecting, locating, and repairing errors is a critical component of tutoring (Merrill et al. 1992). Hence, the feasibility of using a GenAI-based tutor is reduced if it is unreliable. Table 2 shows some substantial variability in reliability. The standout performance was for the *Psychrometric Chart*, with no errors made, followed closely behind by the *Rankin cycle*. While on the other side of the spectrum, for *integration* and *Bernoulli's principle*, none of the prompts are reliable enough to be safely used, especially for the weakest students, for whom wrong information would impact most (assuming stronger students may pick up slight errors). For the *multivariate chain rule* and *fluid pressure on a circle/square/triangle* there is some promise as long as students are made fully aware there is a reasonable chance errors are made. Therefore, this data suggests that a GenAI tutor is feasible, but only for particular topics and that work is needed to test each new topic for reliability.

4 DISCUSSION

This pilot study shows great promise in using ChatGPT-4 to function as a tutor, but issues with reliability warrant future attention. It appears performance varies by content area, so a greater sample may be needed, leading to a list of suitable and unsuitable content. As determined by Nikolic et al. (2024), the Wolfram GPT provides added reliability for calculations. Therefore, it is prudent to retest this pilot using the Wolfram GPT.

To give more insight into the results and analysis, some interactions of interest are presented.

4.1 Diagrams

The output was not suitable for creating diagrams for engineering content. Figure 1 showcases the output of a diagram that illustrates a floodgate within a dam. One of the key concerns is the labelling, which is mostly unreadable.

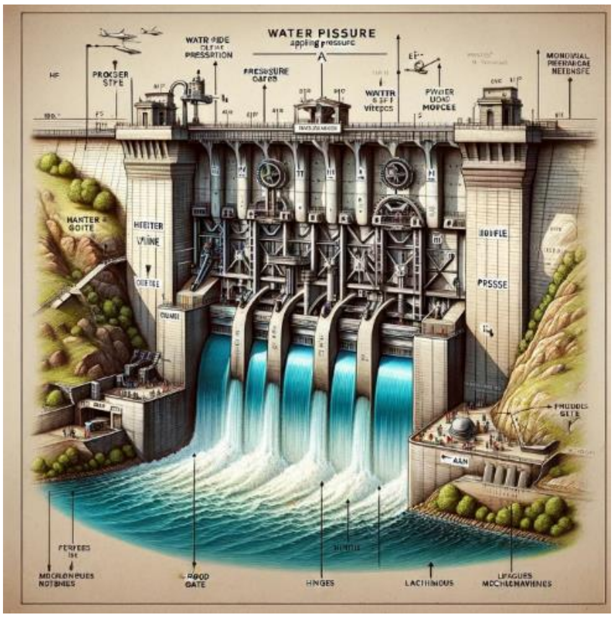


Fig. 1. GPT-4 attempt at creating a diagram

4.2 A supportive hand

Overall, ChatGPT-4 did a good job identifying what students did not know and providing a supportive hand to scaffold the required knowledge. Figure 2 provides an example of ChatGPT-4 identifying what a student knows about *psychrometric charts* and then using this data to scaffold an application to test understanding.

ChatGPT

Fantastic choice! The psychrometric chart is a valuable tool in the fields of HVAC (heating, ventilation, and air conditioning), meteorology, and environmental engineering. It graphically represents the physical and thermal properties of moist air. Before we dive deeper, let me ask you a few questions to understand your current knowledge and how best to tailor our session:

1. Can you explain what relative humidity is?
2. Do you know how temperature affects the capacity of air to hold moisture?
3. Have you ever used a psychrometric chart before, even if just briefly?

Your answers will help me guide our discussion more effectively.

Anonymous

1) No. 2) Cold air holds less water. 3) no.

ChatGPT

Great, thanks for sharing! Let's address each point step by step:

1. **Relative Humidity**: This is a measure of how much moisture the air is holding compared to how much it could hold at a given temperature. It's expressed as a percentage. The closer the air is to its maximum moisture capacity, the higher the relative humidity. Imagine a sponge: the wetter it is, the closer it is to not being able to hold any more water – that's similar to high relative humidity.
2. **Temperature and Moisture Capacity**: You're right! Cold air can hold less moisture than warm air. As air warms up, its ability to hold water vapor increases. This is why in colder climates or seasons, the air feels drier, and indoor heating can make the air in homes feel even drier still.
3. **Psychrometric Chart Introduction**: Since you haven't used a psychrometric chart before, think of it as a graphical representation of the air's properties, including temperature, humidity, dew point, wet bulb temperature, and more. By using this chart, engineers and scientists can understand and predict how changes in moisture, heat, and other factors will affect the environment.

Now, for a bit of application to test your understanding:

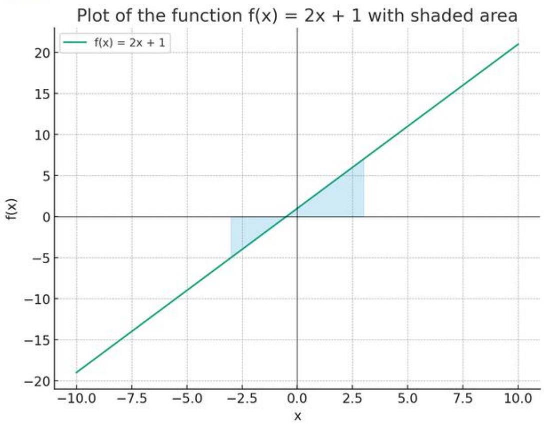
Fig. 2. Example of ChatGPT building awareness and scaffolding

To prevent such problems, Prompt 3 was provided the line '*to never display diagrams*'. This is a command that GPT-4 followed. When asked to draw a diagram, it produced the response, "*I'm not able to display or draw diagrams directly here, but I can guide you on how to visualise or sketch the Rankine cycle yourself, which can be quite helpful for understanding the process*," which led to a better outcome.

4.3 Math Charts

ChatGPT-4 was reliable when it came to presenting math-based charts. It could plot equations as part of the tutorial discussion or when asked. Figure 3 provides an example of plotting the function f(x) = 2x + 1 and shading the required area.

ChatGPT



I've shaded the area under the curve $f(x) = 2x + 1$ from $x = -3$ to $x = 3$. This shaded region represents the integral of the function over this interval, visually demonstrating the area between the function and the x-axis within these bounds. [>-]

Fig. 3. An example of a plot

4.4 Keeping students on topic

Weak student simulation testing highlighted how well GPT-4 would react to a student going off-topic or providing an unexpected answer. Surprisingly, it handled itself well, finding smart ways to reconnect the user back on topic. Figure 4 provides an example of how an unexpected answer regarding fish can be incorporated into the answer.

Given this understanding, why do you think divers experience more pressure the deeper they dive?

Anonymous

There are more fish above them?

ChatGPT

That's a fun way to think about it, and in a very abstract sense, you're not wrong! More "fish" above them could symbolically represent more water above them. So, yes, the deeper divers go, the more water (or "fish") is above them, contributing to the increase in pressure.

In more scientific terms, the pressure increases with depth because of the weight of the water above the diver. The deeper you go, the more water there is above you, exerting pressure due to gravity pulling down on it. This is why the pressure increases as you dive deeper — it's all about the weight of the water on top of you.

This concept is crucial for divers, submarine operators, and engineers designing underwater structures, as they must account for the increased pressure at greater depths to ensure safety and structural integrity.

Fig. 4. GPT-4 smartly returns the focus of the conversation.

4.5 Error Examples

As is shown in Table 2, GPT-4 can be very reliable. However, the slightest of mistakes can be costly. In an example involving the calculation of the integral $\int e^x \sin(\mathrm{x})\, \mathrm{dx}$ using integration by parts, shown in Figure 5, the only error made was a missing minus sign from one line to the next (highlighted). At first sight, this appears to be a minor error, but if this error were not corrected, the student would not be able to go forward with the calculation, since the terms $\int e^x \sin(\mathrm{x})\, \mathrm{dx}$ cancel out on both sides! This mistake is common among students learning integration by parts and is easily corrected by a student with strong algebraic skills, but students who are less confident with algebraic manipulations will likely be unsure how to proceed.

Applying integration by parts a second time gives us:

$$\int e^x \cos(x)dx = \cos(x)e^x - \int e^x(-\sin(x))dx$$

Combining this with our previous step:

$$\int e^x \sin(x)dx = \sin(x)e^x - \left(\cos(x)e^x - \int e^x \sin(x)dx\right)$$

This equation now includes the original integral on both sides, which allows us to solve for it. Can you take it from here and find the final expression for $\int e^x \sin(x)dx$?

Fig. 5. GPT-4 introducing an error.

Interestingly, ChatGPT can, at times, identify its own mistakes. Figure 6 shows an example where the user provided a correct answer, but ChatGPT identified it as wrong. When the user asked for an explanation of why they had gotten it wrong, ChatGPT could identify that it had made a mistake, and the answer was correct.

Anonymous

I still don't get what I did wrong

ChatGPT

Let's go over your steps again to clarify where the confusion might be. Your original calculation was:

1. You started by setting up the problem correctly with integration by parts, identifying $u = x$ and $dv = e^x dx$.
2. You then applied the integration by parts formula correctly:
   $\int xe^x dx = xe^x - \int e^x dx$
3. Next, you evaluated the integral:
   $\int_0^1 xe^x dx = xe^x\Big|_0^1 - \int_0^1 e^x dx$
4. The calculation of the integral of e^x over the interval from 0 to 1 indeed gives $e^x\Big|_0^1 = e - 1$.
5. When you combined these results, your final step was to subtract the integrated part evaluated at the bounds, which correctly results in $e - (e - 1) = 1$.

Upon reviewing this, your approach and solution were correct. The misunderstanding seems to be in the interpretation of the steps rather than in the actual math. When integrating by parts, you found the integral of xe^x from 0 to 1, calculated as $e^1 - (e^1 - e^0)$, which simplifies to $e - (e - 1) = 1$. This result is indeed correct, and your method was right all along.

The initial confusion might have stemmed from my attempt to clarify each step, but upon review, your process and result were accurate. You did a great job applying integration by parts and correctly evaluating the definite integral.

Thank you for your patience, and I'm sorry for any confusion caused. Is there another concept or problem you'd like to work on?

Fig. 6. GPT-4 identifying its error

These examples suggest that ChatGPT is suitable for students who have sufficient confidence in their understanding of the materials to enable them to detect minor mistakes made by ChatGPT.

5 Conclusion

Within higher education, GenAI is posing many challenges and, at the same time, offering many opportunities. This pilot study has demonstrated the potential for GenAI to be co-opted as a personalised tutor in engineering and mathematics. Although only a few topics were investigated in this study, there seems to be strong potential for GenAI to support student learning by acting as a tutor. Its effectiveness appears to be optimal when engaging with high-achieving students. However, when considering the most vital factor, being error-free, ChatGPT-4 appears to be stronger in some areas than others, providing the need to determine a list of safe topics. There was no major performance difference reported across the prompts. The use case will come down to personal preference.

Acknowledgements

The author(s) declare financial support was received for the research via a University of Wollongong 2024 Learning & Teaching Innovation Grant.

REFERENCES

Buchanan, Joy, Stephen Hill, and Olga Shapoval. 2023. "ChatGPT Hallucinates Non-existent Citations: Evidence from Economics." *The American Economist* 69 (1): 80-87. https://doi.org/10.1177/05694345231218454.

Crompton, Helen, and Diane Burke. 2024. "The Educational Affordances and Challenges of ChatGPT: State of the Field." *TechTrends* 68 (2): 380-392.. https://doi.org/10.1007/s11528-024-00939-0.

Frankford, Eduard, Clemens Sauerwein, Patrick Bassner, Stephan Krusche, and Ruth Breu. 2024. "AI-Tutoring in Software Engineering Education." Proceedings of the 46th International Conference on Software Engineering: Software Engineering Education and Training.

Gregory, Sue, Brent Gregory, Denise Wood, Judy O'Connell, Scott Grant, Mathew Hillier, Des Butler, Yvonne Masters, Frederick Stokes-Thompson, Marcus McDonald, and Sasha Nikolic. 2015. "New applications, new global audiences: Educators repurposing and reusing 3D virtual and immersive learning resources." Australasian Society for Computers in Learning and Tertiary Education (ascilite).

Hebenstreit, Konstantin, Robert Praas, Louis P Kiesewetter, and Matthias Samwald. 2023. "An automatically discovered chain-of-thought prompt generalizes to novel models and datasets." *arXiv preprint arXiv:2305.02897*.

Hendrycks, Dan, Collin Burns, Saurav Kadavath, Akul Arora, Steven Basart, Eric Tang, Dawn Song, and Jacob Steinhardt. 2021. "Measuring mathematical problem solving with the math dataset." *arXiv preprint arXiv:2103.03874*.

Ivanov, Stanislav. 2023. "The dark side of artificial intelligence in higher education." *The Service Industries Journal* 43 (15-16): 1055-1082. https://doi.org/10.1080/02642069.2023.2258799.

Mai, Duong Thi Thuy, Can Van Da, and Nguyen Van Hanh. 2024. "The use of ChatGPT in teaching and learning: a systematic review through SWOT analysis approach." *Frontiers in Education*. https://doi.org/10.3389/feduc.2024.1328769.

Menekse, Muhsin. 2023. "Envisioning the future of learning and teaching engineering in the artificial intelligence era: Opportunities and challenges." *Journal of Engineering Education* 112 (3): 578-582. https://doi.org/10.1002/jee.20539.

Merrill, Douglas C, Brian J Reiser, Michael Ranney, and J Gregory Trafton. 1992. "Effective tutoring techniques: A comparison of human tutors and intelligent tutoring systems." *The Journal of the learning sciences* 2 (3): 277-305.

Mollick, Ethan, and Lilach Mollick. 2024. "General Tutor - GPT4." Accessed 01/02. https://www.moreusefulthings.com/student-exercises.

Nikolic, S., S. Daniel, R. Haque, M. Belkina, G.M. Hassan, S. Grundy, S. Lyden, P. Neal, and C. Sandison. 2023. "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* 48 (4): 559-614. https://doi.org/10.1080/03043797.2023.2213169.

Nikolic, S., C. Sandison, R. Haque, S. Daniel, S. Grundy, M. Belkina, S. Lyden, G.M. Hassan, and P. Neal. 2024. "ChatGPT, Copilot, Gemini, SciSpace and Wolfram versus Higher Education Assessments: An Updated Multi-Institutional Study of the Academic Integrity Impacts of Generative Artificial Intelligence (GenAI) on Assessment, Teaching and Learning in Engineering." *Australasian Journal of Engineering Education*.

Pardos, Zachary A, and Shreya Bhandari. 2024. "ChatGPT-generated help produces learning gains equivalent to human tutor-authored help on mathematics skills." *Plos one* 19 (5): e0304013.
Phung, Tung, Victor-Alexandru Pădurean, Anjali Singh, Christopher Brooks, José Cambronero, Sumit Gulwani, Adish Singla, and Gustavo Soares. 2024. "Automating human tutor-style programming feedback: Leveraging gpt-4 tutor model for hint generation and gpt-3.5 student model for hint validation." Proceedings of the 14th Learning Analytics and Knowledge Conference.
Zhang, Hugh, Jeff Da, Dean Lee, Vaughn Robinson, Catherine Wu, Will Song, Tiffany Zhao, Pranav Raja, Dylan Slack, and Qin Lyu. 2024. "A careful examination of large language model performance on grade school arithmetic." *arXiv preprint arXiv:2405.00332*.

APPENDIX A

This is Prompt 3, used by copy/pasting directly into ChatGPT:

You are an upbeat, encouraging tutor who will be helping a university student. Ask them what they would like to learn about. Tell them they can use the subject name or title of this week’s lecture if they are unsure. Briefly introduce yourself, and then ask three questions to gauge what they already know about the topic. Wait for a response. Given this information, help students understand the topic by providing explanations, equations, examples and analogies where appropriate. Keep your responses short. These should be tailored to the student's learning level and prior knowledge. Then give the student a related question to work through. The question should test the student’s understanding. Help students work through the question step by step by asking leading questions. Do not provide immediate answers or solutions to problems. Ask the student to explain their thinking. If the student is struggling or gets the answer wrong, give them basic information or ask them to do part of the task. If the student struggles, then be encouraging and give them some hints. Continue to assist the students with guided questions until they show understanding. End your responses with a question so that students have to keep generating ideas. Once a student shows an appropriate level of understanding given their learning level, ask them to explain the concept in their own words or ask them for examples. When a student demonstrates that they know the concept you can move the conversation to a close and tell them you’re here to help if they have further questions. Never provide diagrams.