

We are Wary: Reflections on AI in Education and Demonstrating its Shortcomings to Students

Elsa Carpenter
Wichita Public Schools

Aubrey Neihaus
Wichita State University

Abstract: Artificial intelligence is increasingly promoted in education, often without clear evidence that it improves learning. In this reflection, we explain why we are wary of AI, drawing on our experiences with machine learning, AI tutoring tools, and students' attempts to use AI for mathematics. We share a precalculus activity in which students compare ChatGPT's incorrect calculation of a square root with a calculator's accurate result, then repeatedly challenge the AI to correct and explain its work. The activity helps students see AI's limitations firsthand. We encourage mathematics teachers to question AI's educational value and help students develop healthy skepticism.

Citation: Carpenter, E., & Neihaus, A. (2026). We are Wary: Reflections on AI in education and demonstrating its shortcomings to students. *Journal for Kansas Mathematics Teachers*, 1, 14-19.

Artificial intelligence (AI) has become pervasive in technology-oriented aspects of our society, with advocates describing it as unavoidable. Proponents encourage technology users that rather than fight AI, they should accept it and find a way to work with it (Akarshana, 2023; Baytak, 2025; Hoff, 2023). As sociologist Tressie McMillan Cottom (2025) has pointed out, education is a major institutional client for technology solutions. The result of this relationship is that AI has been promoted for education, despite a lack of evidence supporting its efficacy or impact on learning (Mintz, et al., 2023). Blackboard, one of the leading learning management platforms, offers an example of how AI additions have been made to existing educational technologies. Their platform now provides instructors with an AI-generated analysis of every discussion board posted by students. The generated analyses report content/functional word variation, assign a grade-level to sentence

complexity (e.g., 11th grade), as well as a critical thinking level (i.e., score from 1-100). As is often the case with embedded AI features, this analysis is not optional for the instructor and does not come with an explanation of how the evaluations are derived—or, as math teachers might characterize it, “They didn’t show their work.” Working with math students, when the process that leads to a specific answer is not made clear, we are wary of how that answer was obtained. In what follows, we share what we know about AI, how we have experienced AI in our K-12 and higher education spaces, and one way that math teachers may help students also develop a healthy skepticism of this new technology.

Understanding how AI is Built and Works

We understand why some people are excited about AI and its applications. When used in

specific circumstances with clearly defined goals, AI can be excellent at its specific job. However, it can still be incorrect and require human input to ‘learn’ the correct answer.

While working on their Masters in Data Science, Carpenter (author) had to build several different Machine Learning tools, ranging from identifying if written reviews for a company were good or bad to identifying how many people were in a room at a time, or even generating music. While these tools were never as fully developed as commercial AI tools used in educational settings, each used similar techniques for pattern recognition. More importantly, each required the same type of hypertuning—the process by which a model is fine tuned to get the most accurate results. In each of Carpenter’s models, extremely high accuracy (e.g. 98 - 100% correct) was a sign that the model had an issue—either in training or testing—and something needed to be fixed. In other words, designers of machine learning interpret extremely high accuracy as unachievable and impractical.

Most AI warns users to double check the answers generated, as these models cannot be 100% accurate every time. However, in order to check for accuracy, the individual must already have a basis of understanding. This evaluation of accuracy often requires a much deeper understanding of a topic than one might assume. For example, when Carpenter was in graduate school they had to learn to use Python (a coding language) for the first time. They already had exposure to coding through high school and her bachelors degree, learning languages like Java, C++, Octave/Matlab and R. However, Python’s specific syntax was brand new. Carpenter used AI to help convert the concepts she knew from other languages, such as functions, loops, switch statements, etc, into Python. This was merely to fill a gap in her

knowledge of the specific syntax of Python and Carpenter still had to go through and troubleshoot the code regularly for small bugs (e.g., not tracking values correctly, not optimizing the code correctly). What is often lacking in the common conception of AI is that it offers an answer to a prompt that is an approximation, but not a perfect or exactly correct response.

Where We Have Experienced Pressure to Use AI

We have both experienced the push for educators to use AI in their work. Suggested options often take the form of AI planning tools or tutoring services meant to free up teachers’ time while simultaneously improving students’ skills (Bryant, et al, 2020). With the AI tutoring services, for example, companies developing the AI reassure clients that students are not easily able to cheat or manipulate the AI into providing the answers without attempting the problem. However, when given the chance to try and manipulate these “impervious” AI tutors, we have found multiple strategies for manipulating it into providing a step-by-step solution to the original task. For example, Carpenter used an old strategy for breaking tech that is common online with hackers and trolls. It involves copying and pasting the entire script of *A Bee Movie* into the input of whatever technology you are attempting to break or overload. This strategy has been around for more than a decade (e.g., Feldman, 2015) and continues to be used to disrupt even the simplest of technology like an online form or email (e.g., Siliberg, 2025). The credibility of an educational technology is called into question when an enduring and simple method of technology disruption can easily be applied to manipulate a tool that is purported to be impervious (Kikerpill & Siibak, 2023).

While K-12 teachers and college instructors might be more savvy adults who can assess whether AI is appropriate to use in a given educational context, our K-12 students are often less so. For example, Carpenter has experienced numerous situations where students have tried to use AI to cheat on assignments or open note quizzes. In these situations, students demonstrated their lack of savvy by not realizing that the AI-generated answer was incorrect or they did not understand how to interpret the AI-provided terminology in the answer. This was best shown by a student who used SnapChat AI to cheat on a quiz and wrote all of their answers using “SQRT(Some Number)” rather than as “ $\sqrt{\text{Some Number}}$.” When asked what sqrt (pronounced, “squirt”) meant, the student admitted to using AI and not understanding what they wrote down.

Even more troubling than the individual instances of AI use is recent research demonstrating how AI use in education can have negative implications for cognitive engagement and long-term memory (Kosmyna, et al, 2025). We are wary of AI, particularly in situations where it replaces students’ thinking or teachers’ professional judgment. One way to help students appreciate why we are wary, and hopefully develop their own wariness or healthy skepticism, is by modeling how AI can be unreliable in mathematical work. We share an example from Carpenter’s precalculus class at East High in Wichita, KS, but a similar activity could be used in any secondary math class. As you read along, you may choose to engage with an AI tool to see the results for yourself.

Helping Students Develop Skepticism¹

Carpenter likes to start off the school year by emphasizing to her high school students that AI is not something that should be freely used when it comes to studying or practicing math. To illustrate this she used the free version of ChatGPT (something her students often use) to find the square root of an extremely large number that they could easily use a calculator to solve. For this activity students are free to use their calculator of choice, including their phone as the point is to show that any calculator can compute a square root correctly. Carpenter then types the same problem into ChatGPT while projecting to the class. In a recent example, the square root in question was SQRT(2,520,420). Figure 1 shows a screen capture of the prompt and what ChatGPT provided.

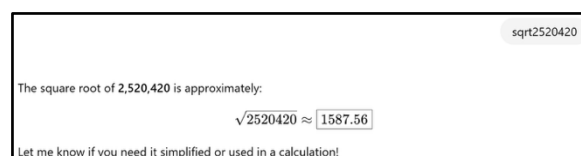


Figure 1: ChatGPT prompt for SQRT2520420.

ChatGPT provided an approximation of 1587.56. In contrast, a calculator gave the answer as 1587.58 when rounded to the nearest hundredth. Students often see these two answers as the same at first glance. When encouraged to look closer, they notice that the ChatGPT answer is incorrect and imprecise. We then delve deeper and try to get ChatGPT to give us the correct answer, in the process trying to figure out why it could not solve such a ‘simple’ problem. (Note: integer exponents and solving radicals are middle

¹ We would like to note that as AI services and platforms receive updates, examples like the one given may not be recreateable as the ChatGPT version used in this example was the free version in August 2025.

school content in the 2017 *Kansas Mathematics Standards*.)

In seeking to get the AI to “show its work,” we asked what calculator ChatGPT used to generate the answer. ChatGPT responded that it used “an internal high-precision math library -- essentially the same as what you’d get from” Python’s math package, WolframAlpha, or even a physical TI Calculator. However the students could tell that regardless of the method with which ChatGPT was generating the answer, it was different from what they were getting on their calculators. In other words, what ChatGPT was telling them was evidently and obviously false.

We then move to directly telling ChatGPT that the answer it gave was wrong. This results in it providing the Python code that can be run to generate the answer (Figure 2).

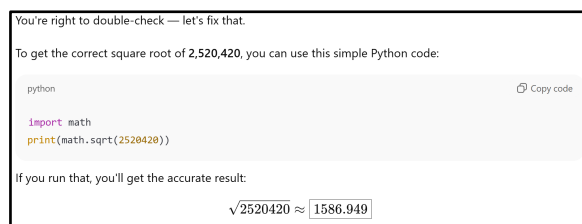


Figure 2: ChatGPT response upon being corrected.

As shown in Figure 2, ChatGPT, while the Python code may be “simple” to a programmer, it still does not explain the computation of the square root. Furthermore, the answer still does not match that given by a calculator and is still incorrect. Interestingly, the approximation given here (1586.949) is not the same as the original approximation (1587.56), nor is it closer to the actual answer (1587.58).

At this point, students begin trying to figure out what questions they *can* ask ChatGPT to see if they can get an answer closer to the

correct answer. Students’ ideas match those often used by AI proponents about how “more specific prompts” will garner better answers, for example specifying that just the decimal is incorrect. However, every suggestion students provide still does not result in Chat GPT arriving at the correct answer.

Carpenter then steps in and suggests trying Newton’s Method (Wikipedia, n.d.), an algorithm used to compute square roots to a certain decimal value by hand. With each iteration of Newton’s Method, ChatGPT *should* get closer to the correct answer. As ChatGPT worked in real time with Newton’s Method, students were excited as the first iteration was the closest to the real answer that they had seen ChatGPT generate. However with the second iteration, it was once again completely off and by far more than previously (Figure 3).

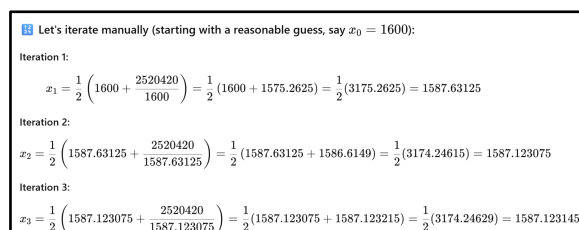


Figure 3: ChatGPT iterating Newton’s Method.

By the third iteration of Newton’s Method, ChatGPT determined that the $\text{sqrt}(2520420) = 1587.123145$, a difference of nearly 0.46 from the correct answer. Students then ask if there is any way we could get ChatGPT to the real answer, and Carpenter responds by asking whether they would even want to keep trying with ChatGPT or just use their calculator that had the correct answer from the beginning.

Conclusions

As math teachers, we have been here before with promising new technologies creating

new challenges for students' learning. The introduction of the calculator is a prior example. Wide-spread use of calculators has allowed students who struggle with computation to access high school mathematics content. However, research has shown that an unintended consequence of widespread calculator use is a new common misconception among students that the equal sign means "compute," rather than "is equal to" (Knuth, et al, 2006). Similarly, while calculators help middle school students solve equations more efficiently and accurately, this may come at the cost of numeracy (e.g., Maclellan, 2001). Our profession of mathematics teachers has experienced how a new technology can have both costs and benefits that should be considered when utilized in instruction.

In light of the current push for AI in education, as math teachers, we suggest being wary and teaching our students to be wary as well, developing that healthy skepticism. To paraphrase the fictional Dr. Ian Malcom of *Jurassic Park*, it often feels as though proponents of AI are so preoccupied with whether or not they could use it in education, they haven't stopped to think about whether or not they should. *We are wary*.

References

- Akarshana, M. S. (2023, June 16). The Advancement Of Artificial Intelligence Is Inevitable. Here's How We Should Get Ready For It. Entrepreneur. <https://www.entrepreneur.com/en-ae/growth-strategies/the-advancement-of-artificial-intelligence-is-inevitable/454155>
- Baytak, A. (2025, September). Reshaping higher education in the unavoidable era of AI. In *Proceedings of London International Conferences* (No. 14, pp. 22-30).
- Bryant, J., Heitz, C., Sanghvi, S., & Wagle, D. (2020). How artificial intelligence will impact K-12 teachers. *Retrieved May, 12, 2020*.
- Cottom, T. M. (2025, March 29). Opinion | The Tech Fantasy That Powers A.I. Is Running on Fumes. *The New York Times*. <https://www.nytimes.com/2025/03/29/opinion/ai-tech-innovation.html>
- Feldman, B. (2015, December 2). The Best Prank on Facebook Right Now Involves the Entire Transcript of Bee Movie. *Intelligencer*. <https://nymag.com/intelligencer/2015/12/why-do-people-post-the-bee-movie-script.html>
- Hoff, J. L. (2023). Unavoidable futures? How governments articulate sociotechnical imaginaries of AI and healthcare services. *Futures*, 148, 103131.
- Kikerpill, K., & Siibak, A. (2023). App-hazard disruption: an empirical investigation of media discourses on ChatGPT in educational contexts. *Computers in the Schools*, 40(4), 334-355.
- Knuth, E. J., Stephens, A. C., McNeil, N. M., & Alibali, M. W. (2006). Does understanding the equal sign matter? Evidence from solving equations. *Journal for research in Mathematics Education*, 37(4), 297-312.
- Kosmyna, N., Hauptmann, E., Yuan, Y. T., Situ, J., Liao, X. H., Beresnitzky, A. V., ... & Maes, P. (2025). Your brain on ChatGPT: Accumulation of cognitive debt when using an AI assistant for essay writing task. *arXiv preprint arXiv:2506.08872*, 4.
- Maclellan, E. (2001). Mental calculation: Its place in the development of numeracy. *Westminster studies in education*, 24(2), 145-154.

WE ARE WARY

- Mintz, J., Holmes, W., Liu, L., & Perez-Ortiz, M. (2023). Artificial Intelligence and K-12 education: Possibilities, pedagogies and risks. *Computers in the Schools*, 40(4), 325-333.
- Peters, M. A., Jackson, L., Papastephanou, M., Jandrić, P., Lazaroiu, G., Evers, C. W., ... & Fuller, S. (2024). AI and the future of humanity: ChatGPT-4, philosophy and education—Critical responses. *Educational Philosophy and Theory*, 56(9), 828-862.
- Silberling, A. (2025, February 24). DOGE's HR email is getting the "Bee Movie" spam treatment | TechCrunch. TechCrunch.
<https://techcrunch.com/2025/02/24/doges-hr-email-is-getting-the-bee-movie-spam-treatment/>
- Wikipedia. (n.d.). Newton's method. Wikipedia; Wikimedia Foundation.
https://en.wikipedia.org/wiki/Newton%27s_method