

Prompt Refinement in Calculus: Using Generative AI to Create Isomorphic Related Rates Problems

Barry J. Griffiths¹, Heather A. Edwards¹, Zhongzhou Chen¹, Katuscia C.B. Teixeira²

¹University of Central Florida, ²Oklahoma State University

In this article, we explore the possibility of integrating generative artificial intelligence (AI) technology into calculus instruction, specifically focusing on the automated creation of isomorphic related rates problems in a way that can be used in asynchronous testing for large classes. Our approach leverages large language models (LLMs) to construct diverse, pedagogically relevant problems, tailored to the demands of institutions where a common test, or a small number of test versions, is no longer feasible. By entering and reiterating the desired parameters, generative AI is used to create an arbitrary number of both generic and isomorphic questions, complete with solutions. With the more general goal being to use AI technology to aid teachers in generating problems that cater to individuals with different learning trajectories, this preliminary study demonstrates the feasibility of using LLMs to create a flexible testing scheme that offers a scalable solution to the challenge of providing large numbers of students with varied and engaging problems.

Keywords: Calculus; generative AI; testing; related rates; prompt refinement

Introduction

The rapid advance of generative AI has ushered in a transformative era for education, leading to a diverse range of articles considering the different applications, impacts, and challenges associated with integrating the new technology into educational settings (Baidoo-Anu and Ansah, 2023; Francis et al., 2025; Giannakos et al., 2024; Yusuf et al., 2024). This transformation is particularly significant in higher education, where large-class courses, such as calculus, present unique challenges in delivering engaging, scalable, and pedagogically sound instruction. As institutions grapple with balancing innovation and integrity, generative AI is a tool with the potential to revolutionize how complex subjects like calculus are taught and assessed, while also paying attention to ethical and practical implications. In this paper, we present a preliminary study that uses the topic of related rates to consider the feasibility of using LLMs to create a dynamic testing environment suitable for institutions where sections of calculus often involve hundreds of students. Calculus, which has an essential role in STEM education, demands a deep understanding of concepts such as related rates, but is often taught in auditorium settings, especially in the United States. As a result, traditional approaches to assessment result in static problem sets, which can limit variety, personalization, and scalability. This challenge is particularly acute when the students are required to take an exam at the same time. However, the development of generative AI offers a dynamic alternative. By leveraging LLMs, instructors can create an arbitrary number of diverse, contextually relevant problems tailored to specific learning objectives, complete with step-by-step solutions, which can be incorporated into practice exercises. This capability not only addresses the logistical demands of large classes but also aligns with pedagogical goals of fostering engagement and comprehension.

Research by Torres-Peña et al. (2024) highlighted a classroom experience where AI was used to update calculus teaching, demonstrating its potential to enhance student interaction with

mathematical concepts through innovative problem design. Similarly, Alvarez (2024) evaluated AI-powered tutors like MathGPT and Flexi 2.0, underscoring their role in improving calculus learning outcomes by adapting to individual student needs.

One of the most promising aspects of generative AI is its potential to personalize learning. Ruiz-Rojas et al. (2023) explored how AI tools can be integrated into instructional design matrices to empower education, tailoring resources to individual student performance. In the context of calculus, this adaptability could mean generating problems of varying difficulty based on a student's mastery of foundational concepts, as suggested by Baidoo-Anu and Ansah (2023). Such personalization not only enhances engagement but also supports inclusive education, a growing priority in higher education. Ogunleye et al. (2024) argued that AI-driven assessment practices can accommodate diverse learning needs, offering a scalable solution for large-class courses where individualized attention is often impractical.

Despite these opportunities, the integration of generative AI into calculus instruction is not without challenges. Ethical concerns, such as those raised by Nguyen (2025) and Yusuf et al. (2024), highlight the tension between innovation and academic integrity. The ease with which AI can generate solutions risks undermining student learning if not carefully managed, prompting calls for frameworks to ensure responsible use. Johnston et al. (2024) and Francis et al. (2025) further explored student and instructor perspectives, noting that while AI tools enhance efficiency, they also necessitate a re-evaluation of assessment validity and fairness. Kaldaras et al. (2024) emphasized the importance of developing valid assessments in the AI era, a critical consideration for ensuring that AI-generated calculus problems maintain mathematical rigor and pedagogical alignment.

Educators play a pivotal role in this evolving landscape. Simelane and Kittur (2025) examined engineering instructors' perspectives on using generative AI, suggesting that faculty must adapt their practices to harness its benefits effectively. This includes designing assessments that leverage AI's capabilities while fostering critical thinking – a skill that remains essential in calculus education. Giannakos et al. (2024) argued that the promise of AI lies in augmenting, rather than replacing, human instruction, which echoes the sentiment by Griffiths (2015). For large-class calculus courses, this might involve using AI to generate initial problem sets, which instructors refine to align with course objectives, ensuring balance between automation and human oversight.

As a result, we aim to utilise AI technology in a way that enhances student engagement and their comprehension of important calculus concepts. This preliminary study looks at the creation of assessment questions involving related rates. By refining an initial prompt, we try to efficiently generate a large bank of isomorphic problems that dissuades the rote memorisation of solutions. These questions will serve as a foundation for building a dynamic environment for learning and assessment that offers a scalable solution to the issue of large-class testing.

Methodology

The generation of related rates questions involved the following steps using the GPT-4o model:

Step 1: Identify the template problem type, which in this instance was initially generic related rates problems, followed by more specific problems involving a change of volume.

Step 2: Design a prompt that includes structural and contextual variations and determine the mathematical constraints. Structural variations refer to construct-relevant variations of the core structure of the problem that must stay within a more precise, user-defined range to ensure the

correctness and/or the appropriate difficulty of the problem. Contextual variations refer to variations in the surface features of a problem. These could include the students' reading level, language proficiency, and cultural background.

Step 3: Execute the prompt and iteratively improve it based on the outcome.

In this instance we designed the following two prompts, with the first generating a selection of related rates problems, and the second generating isomorphic questions focused on volume.

Task 1. We want to generate ten problem contexts for related rates problems.

Before creating the problems, here are some structural variations to consider: (1) For problems involving volume, give the volume formula within the problem body. (2) For problems involving similar triangles, provide the geometric relationship by giving the full dimensions of the shape (e.g., tank is 9m tall with radius 6m). (3) Students will not have access to a calculator, so any given information should be an integer not equal to 1. (4) At the end of the problem, instruct students to give their answer in exact form with the appropriate unit of measurement.

Based on the problem scenario and the structural variations described, generate 10 problems along with their solutions.

Task 2. We want to generate ten isomorphic problems that calculate the rate of change in volume.

Before creating the problems, here are some structural variations to consider: (1) A vessel - whose name and shape should vary (for example, spherical balloon and conical tank) - should be either filled or emptied with a substance. (2) The substance entering or leaving the vessel should vary (for example, air, water, oil, etc.). (3) Any calculation requiring the use of pi should ask for the answer to be written in terms of pi. (4) Any vessel in the shape of a cone should be pointed downwards. (5) The units of measurement should be used within the problem.

Based on the problem scenario and the structural variations described, generate 10 problems along with their solutions.

Results

Task 1: The GPT-4o model responded satisfactorily to the prompts, though multiple iterations were needed to ensure the quality of the output for certain questions.

One of the ten problems generated by the LLM is given below:

Problem 5: Filling a hemispherical bowl

A hemispherical bowl of radius 5 inches is being filled with liquid at a rate of 3 in³/min. Find the rate at which the height of the liquid is rising when the depth of the liquid is 2 inches. Use the formula for the volume of a spherical cap

$$V = \frac{1}{3}\pi h^2(3r - h)$$

Solution:

$$\frac{dV}{dt} = \pi h(3r - h) \frac{dh}{dt}$$

Substituting in the values, we get

$$\frac{dh}{dt} = \frac{3}{26\pi} \text{ in/min}$$

Here there was an error, and the prompt below was needed to correct the solution.

In problem 5, the derivative is calculated incorrectly. This problem needs to be reworked.

This time the error was removed and the correct solution given.

Task 2: The GPT-4o model generated ten isomorphic related rates problems that took into account the structural variations listed. Again, multiple iterations were needed to ensure the written and mathematical quality of the output for several questions.

One of the ten problems generated by the LLM is given below:

Problem 1: Conical barrel (water)

A conical barrel (pointed side down) with height 5 ft and radius 2 ft is being filled with water. The water depth is rising at 0.03 ft/min. The water conforms to the conical shape of the container. Find the rate of change of the volume of water when the water is 4 ft deep. Use

$$V = \frac{1}{3}\pi r^2 h \text{ with } r = \frac{2}{5}h.$$

Solution:

$$\begin{aligned} r = \frac{2}{5}h \rightarrow V &= \frac{1}{3}\pi \left(\frac{2h}{5}\right)^2 \rightarrow h = \frac{4\pi h^3}{75} \\ \rightarrow \frac{dV}{dt} &= \frac{4\pi}{75} \cdot 3h^2 \cdot \frac{dh}{dt} = \frac{4\pi}{25}h^2 \cdot \frac{dh}{dt} \\ &= \frac{4\pi}{25} \cdot 16 \cdot 0.03 = \frac{64\pi}{625} \text{ ft}^3/\text{min} \end{aligned}$$

Here there were several errors, which required the additional prompt below.

There are multiple errors in problem 1: Firstly, barrels are not typically in the shape of a cone, so use another word in place of barrel. Secondly, remove the sentence saying that water conforms to the conical shape of the container. Thirdly, there is an error in the equation for volume. Finally, $\frac{4\pi}{25} \cdot 16 \cdot 0.03$ is not equal to $\frac{64\pi}{625}$.

Again, the errors were corrected, and both a valid question and solution given.

Discussion

At the present time generative AI has not yet developed to the point where a single prompt will allow us to accurately create an arbitrary number of isomorphic related rates questions. Modifying the described procedure to include the prompt-chaining approach advocated by Chen (2025),

where the task is broken down into multiple subtasks and executed through a chain of prompts, will likely improve the efficiency of problem creation, but it will still be necessary to carefully review the generated outputs to ensure the desired structural and contextual constraints are satisfied, and that the mathematics is accurate. However, given the rapid development of LLM-based methods, it is only a matter of time before we can generate problems with minimal oversight.

A strong case can be made that student assessment in all subjects stands to benefit significantly from generative AI, especially when class sizes are large. Schmidt and Magzoub (2024) discussed the opportunities and pitfalls of using ChatGPT to write exam questions in the health professions, noting the balance between the increased efficiency of the process and the need to check for errors, and the same issues have been demonstrated by this study. However, we believe that in time, the ability to generate varied problems – such as related rates in dynamic systems – offers a scalable solution to the challenge of assessing hundreds of students simultaneously.

The literature also underscores the need to address pedagogical concerns. In calculus, this might involve validating AI-generated problems against established learning outcomes, as suggested by Mittal et al. (2024). Barrett and Pack (2023) noted that student and teacher perspectives on AI tools reveal both enthusiasm and scepticism, highlighting the importance of stakeholder input.

Conclusion

This study shows that by carefully refining an initial prompt, an arbitrary number of isomorphic questions can quickly be created covering the topic of related rates. However, at this time the prevalence of errors in the output means that caution should be exercised when using the technology. It can be reasonably assumed that the number of errors will diminish as the technology evolves, and that generative AI holds immense potential to transform large-class calculus assessment by offering scalable, personalized, and engaging solutions to longstanding challenges. Its ability to generate diverse problems aligned with real-world applications can deepen student understanding of related rates, while adaptive features can be used to cater to individual learning trajectories. Realizing this potential requires educators to be mindful of ethical, pedagogical, and practical considerations. Educators should work with researchers and policymakers to develop guidelines that ensure AI enhances, rather than replaces, human instruction. Maximising the power of generative AI in education hinges on a collaborative approach that prioritizes equity, rigor, and student success. Our study builds on these insights, examining how LLMs can create a dynamic assessment environment in calculus, offering a model for future innovation in higher education.

This paper discusses a preliminary step towards the creation of isomorphic test questions involving related rates that could be used in a calculus course. In subsequent studies, the techniques used will be refined and expanded to a wider range of calculus topics, with the overall goal being to create entire exams that would be well-suited to large-class and asynchronous testing situations.

References

- Alvarez, J. (2024). Evaluating the impact of AI-powered tutors MathGPT and Flexi 2.0 in enhancing calculus learning. *Jurnal Ilmiah Ilmu Terapan Universitas Jambi*, 8(2), 495-508. <https://doi.org/10.22437/jiituj.v8i2.34809>
- Baidoo-Anu, D., & Ansah, L. O. (2023). Education in the era of generative artificial intelligence (AI): Understanding the potential benefits of ChatGPT in promoting teaching and learning. *Journal of AI*, 7(1), 52-62. <https://doi.org/10.61969/jai.1337500>

- Barrett, A., & Pack, A. (2023). Not quite eye to AI: Student and teacher perspectives on the use of generative artificial intelligence in the writing process. *International Journal of Educational Technology in Higher Education*, 20(1), Article 59. <https://doi.org/10.1186/s41239-023-00427-0>
- Chen, Z. (2025). Reliable generation of isomorphic physics problems using ChatGPT with prompt-chaining and tool use. *arXiv preprint arXiv:2508.14755*.
- Francis, N., Jones, S., & Smith, D. (2025). Generative AI in higher education: Balancing innovation and integrity. *British Journal of Biomedical Science*, 81, Article 14048. <https://doi.org/10.3389/bjbs.2024.14048>
- Giannakos, M., Azevedo, R., Brusilovsky, P., Cukurova, M., Dimitriadis, Y., Hernandez-Leo, D., Järvelä, S., Mavrikis, M., & Rienties, B. (2025). The promise and challenges of generative AI in education. *Behaviour & Information Technology*, 44(11), 2518–2544. <https://doi.org/10.1080/0144929X.2024.2394886>
- Griffiths, B. (2015). Perspectives of exchange students on the role of classroom technology: A Law of Diminishing Returns? *International Journal for Infonomics*, 8(1), 974-978. <https://doi.org/10.20533/iji.1742.4712.2015.0114>
- Johnston, H., Wells, R., Shanks, E., Boey, T., & Parsons, B. (2024). Student perspectives on the use of generative artificial intelligence technologies in higher education. *International Journal for Educational Integrity*, 20(2), 1-21. <https://doi.org/10.1007/s40979-024-00149-4>
- Kaldaras, L., Akaeze, H., & Reckase, M. (2024). Developing valid assessments in the era of generative artificial intelligence. *Frontiers in Education*, 9, Article 1399377. <https://doi.org/10.3389/feduc.2024.1399377>
- Mittal, U., Sai, S., & Chamola, V. (2024). A comprehensive review on generative AI for education. *IEEE Access*, 12, Article 142733. <https://doi.org/10.1109/ACCESS.2024.3468368>
- Nguyen-Viet, K. (2025). The use of generative AI tools in higher education: Ethical and pedagogical principles. *Journal of Academic Ethics*, 23, 1435-1455. <https://doi.org/10.1007/s10805-025-09607-1>
- Ogunleye B., Zakariyyah, K., Ajao, O., Olayinka, O., & Sharma, H. (2024). Higher education assessment practice in the era of generative AI tools. *Journal of Applied Learning & Teaching*, 7(1), 46-56. <https://doi.org/10.37074/jalt.2024.7.1.28>
- Ruiz-Rojas, L., Acosta-Vargas, P., De-Moreta-Llovet, J., & Gonzalez-Rodriguez, M. (2023). Empowering education with generative artificial intelligence tools: Approach with an instructional design matrix. *Sustainability*, 15(15), 11524. <https://doi.org/10.3390/su151511524>
- Schmidt, H., & Magzoub, M. (2024). Using ChatGPT to write examination questions: Opportunities and pitfalls. *Health Professions Education*, 10(1), Article 1. <https://doi.org/10.55890/2452-3011.1067>
- Simelane, P., & Kittur, J. (2025). Use of generative artificial intelligence in teaching and learning: Engineering instructors' perspectives. *Computer Applications in Engineering Education*, 33(1), e22813. <https://doi.org/10.1002/cae.22813>
- Torres-Peña, R. C., Peña-González, D., Chacuto-López, E., Ariza-Echeverri, E. A., & Vergara, D. (2024). Updating Calculus Teaching with AI: A Classroom Experience *Education Sciences*, 14(9), Article 1019. <https://doi.org/10.3390/educsci14091019>
- Yusuf, A., Pervin, N., & Román-González, M. (2024). Generative AI and the future of higher education: A threat to academic integrity or reformation? Evidence from multicultural perspectives. *International Journal of Educational Technology in Higher Education*, 21(1), Article 21. <https://doi.org/10.1186/s41239-024-00453-6>