

# Large Language Models (LLMs) in STEM Education

Gavin Huang, Prof. Andrea Belz

Daniel J. Epstein Department of Industrial and Systems Engineering and Bridge Institute, University of Southern California, Los Angeles, CA, USA



Bridge UnderGrad Science (BUGS) Summer Research Program

## Abstract

Generative Artificial Intelligence (GAI) offers the capability to transform middle-school mathematics (MSM) education, but this facility is still in its infancy. In this project, we explore the use of GAI, informed by large language models (LLMs) to solve MSM problems and the more difficult task of generating them. Here we review the literature on automated MSM problem-solving and generation; and then we propose draft guidelines for MSM educators to implement immediately with the popular ChatGPT system. This work should inform MSM education and the rapidly evolving field of GAI that uses LLMs.

## Using Large Language Models to Solve Problems

- Chain-of-Thought (CoT)<sup>[1]</sup>
  - Going step by step increases accuracy
  - Limited to LLMs with large size
- Few-shot Learning (FSL)
  - Use of similar problems improves accuracy
  - Trains LLM on example problems
- Tool-augmented LLMs<sup>[2]</sup>
  - Outside APIs improve LLMs
  - Giving ChatGPT a “calculator”
- Self-verification<sup>[3]</sup>
  - Uses LLM to check its own answer
  - Higher cost

## References

Wei, Jason, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Brian Ichter, Fei Xia, Ed Chi, Quoc Le, and Denny Zhou. 2022. “Chain-of-Thought Prompting Elicits Reasoning in Large Language Models.” *arXiv [cs.CL]*. arXiv. <http://arxiv.org/abs/2201.11903>.

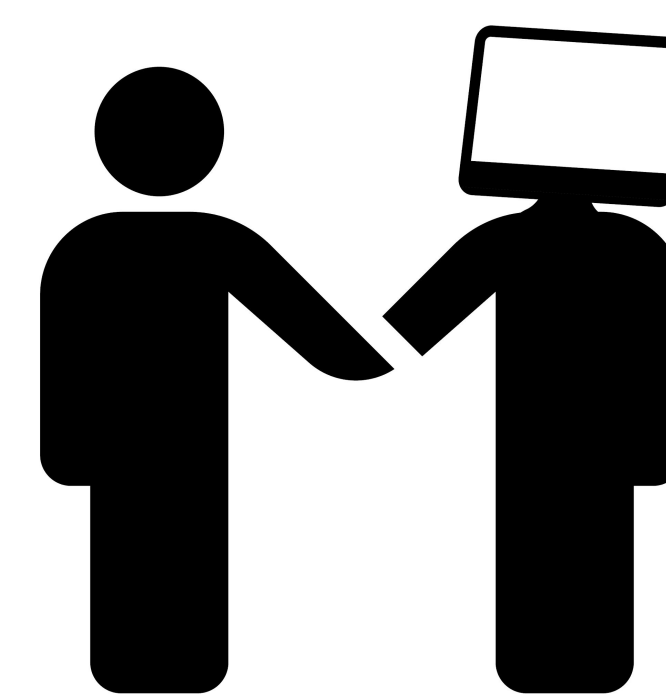
Zhang, Beichen, Kun Zhou, Xilin Wei, Wayne Xin Zhao, Jing Sha, Shijin Wang, and Ji-Rong Wen. 2023. “Evaluating and Improving Tool-Augmented Computation-Intensive Math Reasoning.” *arXiv [cs.CL]*. arXiv. <http://arxiv.org/abs/2306.02408>.

Weng, Yixuan, Minjun Zhu, Fei Xia, Bin Li, Shizhu He, Kang Liu, and Jun Zhao. 2022. “Large Language Models Are Better Reasoners with Self-Verification.” <http://arxiv.org/abs/2212.09561>.

## Vision

- Use LLMs to improve math education
- Improve cognition and math skills
- Create math education that attracts a diverse workforce

How can we help teachers partner with LLMs?



## Using Large Language Models to Create Math Problems

- Advantages
- Fast
  - Adaptable

- Disadvantages
- Disconnected from the curriculum
  - May be plagiarized
  - Solutions it gives may be incorrect
  - Generates only text (no diagrams or graphs)

**G** You are ChatGPT, an AI that generates high quality math problems. I am a middle school math teacher and need questions on finding the sum of finite and infinite geometric series. Can you generate 3 questions for me? The answers should be integers and not have a high magnitude.

**G** Of course! Here are three math problems on finding the sum of finite and infinite geometric series:

Question 1:  
Find the sum of the first 5 terms of the geometric series: 2, 6, 18, 54, ...

Question 2:  
Determine the sum of the infinite geometric series: 3, 6, 12, 24, ...

Question 3:  
A ball rebounds to a fraction of its previous height each time it bounces. If the first bounce reaches a height of 5 meters and each subsequent bounce reaches 60% of the previous height, find the total distance the ball travels (up and down) after 6 bounces. (Round the answer to the nearest meter)

Remember to use the formula for the sum of a finite geometric series, and for the sum of an infinite geometric series, use the formula  $S = a / (1 - r)$ , where 'a' is the first term and 'r' is the common ratio. Happy teaching! Let me know if you need any further assistance.

- Do
- Specify what you want
  - Give context
  - Verify the solutions
  - Customize results

- Don't
- Use problems without verification
  - Make prompts too long
  - Ask for visual output

## Relevance

- Math is required in life sciences to:
  - Analyze data
  - Create models of biological systems
- Project contributes to cognition research: Understanding teaching is to understand how a human learns
- Research Question: How can we conduct prompt engineering to generate meaningful problems?

## Our Work

- Surveyed literature on solving and generating math problems
- Composed guidelines for teachers on using LLMs
- Experimented with one LLM (ChatGPT) in generating problems

## Next Steps Toward LLMs in the Classroom

- Guide teachers to use and understand LLMs
- LLMs are the sidekick, not the superhero

## CONTACT US

[bridge.usc.edu/bugs](http://bridge.usc.edu/bugs), [abelz@usc.edu](mailto:abelz@usc.edu)