Roles and Research Trends of Artificial Intelligence in Mathematics Education: A Bibliometric Mapping Analysis and Systematic Review

Hwang, G., & Tu, Y. (2021). Roles and Research Trends of Artificial Intelligence in Mathematics Education: A Bibliometric mapping analysis and Systematic review. *Mathematics*, *9*(6), 584. https://doi.org/10.3390/math9060584

Introduction

"This research aims to conduct a bibliometric mapping analysis and systematic review to explore the role and research trends of AI in mathematics education by searching for the relevant articles published in the quality journals".

My **main takeaway** is how wide-ranging AI can be in education. We often think of AI as tools like ChatGPT, but, as evidenced by this systematic review, which precedes the birth of generative AI's, the landscape of applications of AI in education is far broader and more diverse that extends far beyond these mainstream tools.

Summary

The authors answer the following (nine) research questions on AI in mathematics education (AIME)

What are the most ...			
Common journals	Cited papers	Productive authors	Used keywords
of AIME research?			

What are the roles of AI in mathematics education?

What are the ...				
the application domains	sample groups selected	the research methods adopted	adopted AI algorithms	research issues investigated
in AIME research?				

The authors review 43 papers ranging from 1996 to 2020. Given the nature of a systematic review, the results are quantitative results based on frequency. For example, the research methods adopted by the AIME studies are Quantitative (79.07%), followed by mixed methods (18.60%) and qualitative methods (2.33 %)

For each of the research questions the authors presented informative and clear graphs summarizing their findings. Some notable results are:

- "The **most frequently adopted application** for AIME studies was discrete mathematics/algebra, followed by general/foundations. On the other hand, geometry and topology, applied mathematics, mathematics literacy and across-disciplines (e.g.,STEM)

were seldom included in those AIME studies. This implies that AIME applications remain in the beginning stage; that is, researchers mainly focused on using AI technologies to solve fundamental problems in mathematics courses"

- "The **most frequent role played by AI in mathematics education** was "intelligent tutoring systems (ITS)", followed by "profiling and prediction" and "adaptive systems and personalization." This is consistent with the finding regarding the research issue, that is, evaluating students' learning performance is the main focus of AIME studies. The main purpose of developing ITS is to evaluate students' learning problems and to provide instant support to them, which aims to improve their learning performance."
- "**Most AIME studies investigated** students' learning achievements (cognition dimension) and learning motivation and attitude (affect dimension). This is because the objective of mathematics education is to foster students' cognition competences. Moreover, since mathematics courses are generally considered by students as being challenging, investigating students' learning motivation or attitude is hence an important research focus."

Critique

In my opinion, the article is a landmark on AI in mathematical education. From their literature review they noticed that "AI is becoming increasingly influential in mathematics education" but there was no article about it. Thus, they decided to conduct a systematic review of the existing literature on AIME (from 1996 till 2020). On one hand, they created a valuable piece of information for the math education community, the 283 citations testify it. On the other hand, the article was written in 2021, just before the launch of ChatGPT, hence it is a landmark between the past and the present of Artificial Intelligence.

I enjoyed reading the article. Their objectives are precise, concise and pertinent. The article answers each of them thoroughly and presents the findings in informative and clear graphs.

The one thing I missed was a table with each article that they review and their tags. For example, say that *article 1* has discrete mathematics as an application field, the sample group was higher education students and the role of AI that the article research is Intelligent tutoring systems.

Reflection

My planned mentored teaching project is about the use of AI tools like Mathway or ChatGPT by undergraduate students when solving homework and the positive or negative consequences of the use. With this reading, I learned about 'intelligent tutoring systems (ITS)'. a field dedicated to develop AI technologies that ideally could 'simulate teachers' intelligence, providing personalized guidance, feedback or supports to individual students during the learning process. Before 2022, the current technology allowed the diagnosis of deficiencies in the learning process and providing feedback to them and curating learning materials for individuals. After 2022, this task can be highly improved by allowing the input of natural language directly from individuals. The student can receive immediate feedback and support and ideally establish an interaction that resembles a human tutor.

Moreover, reading this article I have learned how wide-ranging AI is in education. I would be interested to explore the different types of research and software presented in the article. For example, adaptive systems and personalization.

ChatGPT: A revolutionary tool for teaching and learning mathematics

Wardat, Y., Tashtoush, M. A., AlAli, R., & Jarrah, A. M. (2023). ChatGPT: A revolutionary tool for teaching and learning mathematics. *Eurasia Journal of Mathematics Science and Technology Education*, *19*(7), em2286. https://doi.org/10.29333/ejmste/13272

Introduction

This study aims to examine the perspectives and experiences on the use of ChatGPT in teaching mathematics. The results of this investigation propose a number of avenues for research that ought to be explored in order to achieve educational goals through the cooperation of human tutors and machines like ChatGPT.

My **main takeaway** is that ChatGPT, and alike models, can be used to provide accurate, fast and helpful answers to user's queries, however they are not capable of reasoning and understanding in the same way as humans, which can sometimes lead to errors or misconceptions. In particular, relatively simple conceptual mathematical problems proved to be a challenge for ChatGPT. This is especially hurtful when a novice user (student) cannot evaluate the accuracy of the answer and it lead to wrong answers.

Summary

The authors were filling the gap of knowledge about the impact of ChatGPT teaching mathematics.

Stage 1: 30 interviewees were chosen for this study based on their public blogs that discussed their use of ChatGPT in mathematics.

> The majority of participants have a positive attitude towards ChatGPT in education. They see its potential to enhance the learning experience and improve students' mathematical abilities. They appreciate the effectiveness of ChatGPT in conveying a thorough comprehension of various and challenging mathematics topics in an understandable way. Also, they were amazed by ChatGPT's ability to solve complex mathematical problems.
>
> However, there are concerns about the inaccuracy or incompleteness of ChatGPT's answers. One of the participants said that "while ChatGPT may be able to discuss math at a surface level, it lacks a genuine understanding of mathematics. ChatGPT is unable to correct misunderstandings in mathematics and may even introduce its own inaccuracies. In addition, there may be instances, where it makes unexplainable mathematical mistakes".

Stage 2: Three experienced educators were recruited to use ChatGPT for a whole week, during which they tested different teaching scenarios. Scenario 1 & 2: solving mathematics equations, scenario 3 & 4: solving limit of functions, and scenario 5 & 6: ChatGPT is ready to teach geometry.

> From scenarios 1 to 4, ChatGPT 3.5 was capable of performing reasonable mathematical steps to solve the problems. Nevertheless, it suffered hallucinations. For example, ChatGPT claimed that $x^3 - 1$ has two complex roots contradicting the fundamental theorem of algebra. In other instances, it made a couple of mistakes on numerical calculations. For example, it claimed that 2*(-5) + 5 = -3.

For scenarios 5 and 6 ChatGPT calculated correctly the length of a leg of a right triangle given the hypotenuse and the other leg using Pythagoras' theorem. Also, it was successful generating an explanation on how to use Pythagoras' theorem in terms that a middle school math student would understand.

However, Chat GPT demonstrates a lack of geometrical understanding when asked about misconceptions about Pythagoras' theorem. For instance, it claimed that one misconception "is that right triangles must have a 90 degree angle in order for the theorem to hold true". Also, when ChatGPT was fed an impossible triangle "Find the size C of a triangle with sides A, B, and C, where A and B are both seven inches long and A and C's angle and A and B's angle are both 45 degrees", the model does not realize the problem is ill-posed and creates an answer for C.

The passage highlights the limitations of ChatGPT's. While it can produce large amounts of text using geometric vocabulary, it lacks a real understanding of basic geometry, resulting in misleading answers

Critique

The strengths of the article are the educational scenarios they experimented with. The scenarios were cleverly chosen. They used basic mathematical problems that a student could try to solve using ChatGPT. Within six scenarios, the article shows the strengths and deficiencies of ChatGPT.

The weakness of the article is the interview phase. Even though they collect valuable opinions about the use of ChatGPT, it is unclear what the methodology of the interview is. It does not specify who the interviewees are. It does not mention if the interview was in person, by email, the questions that were asked, etc …

Overall, it is a good article that poses provoking experiments and ideas for future research on how human tutors and machines like ChatGPT could cooperate to achieve educational goals in mathematics.

Reflection

I really enjoyed reading the results of the educational scenarios created. I tried them myself. On the one hand, for scenarios 1 to 4 ChatGPT 4.0 gives fast, clear and very detailed answers on how to solve limits and equations. I can envision how a student can get a valuable learning support from ChatGPT if no tutor or mathematical "expert" is around. Moreover, it does not make calculational errors in contrast with the article.

On the other hand, ChatGPT 4.0 cannot handle conceptual problems related to the Pythagorean theorem. For example, it still does not recognize the impossible triangle i.e. it does not capture the geometrical meaning of the triangle. Also, when it is asked for misconceptions about the theorem it says that if a triangle has sides 3 ,4 and 5, then *is not* necessarily a right triangle. However, the Pythagorean theorem is a double implication, actually the statement is true.

In the future, I would like to create this type of scenario where students try themselves to write the prompts and reflect about ChatGPT's answers. I would like to promote my students' critical thinking and enhance their mathematical understanding.

Using ChatGPT as a proof assistant in a mathematics pathways course

Park, H., & Manley, E. D. (2024). Using ChatGPT as a proof assistant in a mathematics pathways course. *The Mathematical Education*, *63*(2), 139–163. https://doi.org/10.7468/MATHEDU.2024.63.2.139

Introduction

"The purpose of this study is to examine the capabilities of ChatGPT as a tool for supporting students in generating mathematical arguments that can be considered proofs. To examine this, we engaged students enrolled in a mathematics pathways course in evaluating and revising their original arguments using ChatGPT feedback"

My **main take away** is learning from a real-life example of the use of ChatGPT as a proof tutor. (More on the reflection)

Summary

In this study, the authors "conjecture that ChatGPT can function as a proof assistant supporting students (novice provers) in their proving activities. To examine the capabilities of ChatGPT as a tool to assist students in generating arguments that can be considered proofs, we (the authors) purposefully created a set of ChatGPT-embedded proving activities for learners"

During the in-class ChatGPT activities, students were asked to pull up ChatGPT 3.5 and type in their argument in the prompt box, providing ChatGPT with what they tried to prove and asking ChatGPT to determine whether their argument looks like a proof. Based on ChatGPT feedback, students were asked to improve their arguments to be acceptable as proofs. During these activities, the instructor engaged students in a discussion about the limitations of ChatGPT and the possibility that it may sometimes hallucinate or generate erroneous information. Students evaluated ChatGPT responses and whether they agreed or disagreed with the feedback. Students also shared their proof-writing experiences with ChatGPT assistance with their classmates during a whole-class discussion

The experiment

29 students were given homework similar to the in-class activity. "The students constructed an argument describing a method to find an area of a triangle in terms of its side lengths and explain why that method works. Then, the students were asked to review and improve their argument based on ChatGPT feedback. They were also asked to answer a question about their experience using ChatGPT in constructing proofs.

The results

Based on the understanding of what proof is, 21 out of 29 students determined their arguments could be accepted as proof.

When students asked for feedback from ChatGPT, it provided suggestions to improve:

1. **Clarity.** For example, ChatGPT suggested defining all the variables on their arguments to help a reader follow their reasoning more easily.
2. **Justification**. For example, ChatGPT encouraged students to give reasons for the statement.
3. **Generalization.** In some cases where students attempted to find a triangle area using specific numbers, ChatGPT tried to direct them to generalize their arguments for all triangles.
4. **Mistaken claims**. Most students made errors by making incorrect assumptions (e.g., assuming a given shape is one type of triangle such as isosceles, equilateral, or scalene triangle) and ChatGPT usually went along with the assumption students made and provided suggestions that fit within the same assumptions.

ChatGPT also **gave confusing or undesirable advice**. ChatGPT suggested rewriting the whole proof using Heron's formula or suggested a circular reasoning to prove why a standard area formula for a triangle works. Neither were the goal of this task, and the answers confused the students.

Twenty students (out of 29) improved their original arguments, while six made no substantial change to their arguments. Three students changed their arguments, but the result was no better than the original.

Most students attempted to revise their original arguments to make them more proof-like, considering ChatGPT feedback by improving clarity, justifying each step, proving generalizations, correcting errors, or revising their claims. ***They did so because ChatGPT's suggestions made sense to them and aligned with their perspectives on proof and proving****.*

Students' reactions to using ChatGPT as a proof assistant to improve their arguments were mostly positive, with 17 of 29 students saying only positive things about their proof writing experiences with ChatGPT. They found ChatGPT to be like a replacement for a peer or a coach that can give feedback and notice flaws they would not have found on their own.

Six other students have mixed opinions, saying some positive things but with caution about some aspects of ChatGPT, while six students had a primarily negative reaction. Among the negative aspects of their interaction with ChatGPT as a proof assistant, one common complaint was how difficulty is communicating with the language model; students had a hard time understanding ChatGPT's suggestions. The other negative comments focused mainly on limitations in ChatGPT's reliability and the students' inability to verify the content it generated.

Critique

The strength of the article is the experiment they created to prove their conjecture. Their semester-long field experiment offers valuable insights into the real-world impact of AI on mathematical education. In particular on a critical topic: the teaching and learning of mathematical proof writing.

In my opinion, the article is a masterpiece. They present a relevant literature review, and their research aims to fill the gap of the potential use of ChatGPT as a proof assistant in the context of mathematics classrooms.

The methodology of the experiment was well though. It allowed them to guide their students into a meaningful experience of learning while collecting valuable data to assess how effective is ChatGPT as a proof assistant and discover how students felt about in the process of improving arguments using ChatGPT's feedback.

The article provides a comprehensive analysis of its results. Their findings are rigorously supported by a combination of their students' reflections, ChatGPT outputs, and their own expertise as educators.

Overall, they managed to demonstrate the potential and limitations of ChatGPT as proof tutor.

Reflection

On the one hand, this reinforces my belief that as educators and students, to maximize the benefits we can obtain from using large language models we need to enhance our mathematical knowledge and understand the strengths and limitations of the software. For example, in the study, most students were able to improve their original arguments because ChatGPT's feedback made sense to them and aligned with their perspectives on proof and proving. Hence, having a fair understanding of a subject allows us to have effective support of ChatGPT.

On the other hand, this insightful article has provided a roadmap for my mentored teaching project. The authors' meticulous attention to detail, from research design to data analysis, will inform my own study.

I am particularly excited to find my research questions, the meaningful tasks my students could perform, how I will collect qualitative and quantitative data, how I will analyze it, etc ...