AI and Core Level Mathematics

Igor Egin, Jeff Morgan, Mansib Mursalin, Karthik Raghavendra Rao, Jatindera Walia

A Short History Lesson

- Research related to AI started in 1950.
- Huge breakthroughs took place in the early 2000s.
- Computer architecture improved tremendously.
- Large language models were introduced to the public in Nov 2022.
- People let their imaginations run wild.

The Initial Public Concern

Students are going to cheat!

My Thoughts

- Students with money have had access to assistance for a long time!

- This group has always represented a small percentage of our students.

- Modern LLMs and AI have brought assistance to the masses.

- I welcome the change!

Can AI Help Students in Core Math?

To answer this question, we need to investigate ChatGPT's mathematical capabilities.

Today's Discussion

- My mathematical experiences with AI and ChatGPT starting in Spring 2023 (followed by CASA IT investigations with AI).

- ChatGPT's potential to help students in UH Math courses.

My Experiences

- January 2023 – I created an OpenAI account.
- ChatGPT 4.0 took my linear algebra class in Spring 2023 and received a failing score.
- However, it gave some impressive answers to difficult questions, and demonstrated knowledge of key concepts.

Example Problem From Linear Algebra
(2 statements of the same problem)

- Find the number of ways that integers a, b and c can be chosen so that $|a| \leq 10$, $|b| \leq 10$, $|c| \leq 10$, $a > 0$ and $ab - c^2 > 0$.

- Find the number of 2 by 2 symmetric positive definite matrices with integer entries less than or equal to 10 in absolute value.

  Note: A 2 by 2 real symmetric matrix has the form $\begin{pmatrix} a & c \\ c & b \end{pmatrix}$. This matrix is positive definite if and only if $a > 0$ and $ab - c^2 > 0$.

Responses from ChatGPT 4.0

- In Spring 2023 it missed both questions.

- In February 2024 it answered the first question perfectly and provided correct Python code to count the number of occurrences.

- **It restated the second question using the language of the first question!** It also gave functioning Python code, but a logical error caused the answer to be incorrect.

Example Problem from Calculus

(February 2024)

Determine the value z between 0 and 1 so that the length of the graph of $f(x) = \tan\left(\sqrt{x^2 + 1}\right)$ on the interval $[0, z]$ is 1.

Response from ChatGPT 4.0

ChatGPT 4.0 gave the correct formula

$$\int_0^z \sqrt{1 + \left(f'(x)\right)^2}\, dx$$

for the length of the graph, formulated $f'(x)$ correctly, identified that the integral cannot be computed exactly, and provided Python code to give an excellent numerically approximation and find z between 0 and 1 so that

$$\int_0^z \sqrt{1 + \left(f'(x)\right)^2}\, dx = 1.$$

Further Investigation

- College Board released the written questions from the Spring 2023 AP Calculus Exam on May 8, 2023.

- I fed these questions to ChatGPT 4.0, and although it made arithmetic errors, it clearly passed!

- Conclusion – The new ChatGPT 4.0 has some mathematical skill. It just can't do arithmetic.

Fundamental Question

Can ChatGPT help students in UH Math courses?

It Certainly is Capable in Other Areas

- New Chat GPT scored 90th percentile on the Bar exam (old 10th).

- New ChatGPT scored 99th percentile in the USA Biology Olympiad (old 31st).

- New ChatGPT 4.0 demonstrated sufficient programming skills to be hired by Amazon (if it were human).

- New ChatGPT 4.0 scored 0 on the recent Mathematics Olympiad! But this is a *brutal math exam*.

Our Investigation

Many Math Courses at UH have large problem banks for weekly online quizzes (containing many thousands of questions). For example,

- MATH 1314 - College Algebra
- MATH 1324 - Finite Math with Applications
- MATH 1325 - Calculus for Business and Social Sciences
- MATH 1342 - Elementary Statistical Methods
- MATH 2312 - Precalculus
- MATH 2413 - Calculus I
- MATH 2414 - Calculus II
- MATH 2415 - Calculus III
- MATH 3321 – Engineering Math
- MATH 3339 - Statistics for the Sciences

Starting Point – Calculus I
(because ChatGPT 4.0 passed the Written AP Calculus Exam in Spring 23)

The initial results were not promising!

ChatGPT 3.5 0613* Turbo – 38% Correct

ChatGPT 4.0 1106 Turbo – 58% Correct

*our runs in 3.5 1106 had some issues

What Happened?

(recall ChatGPT 4.0 passed the Written AP Calculus Exam in Spring 23)

I fed many individual questions through my OpenAI account yesterday.

The solution process on most problems was correct, but the final answer was typically flawed due to simple errors.

My Conclusion

ChatGPT 4.0 can do mathematics, and when a new version or a competitor's version conquers arithmetic, it's going to be a 24/7 assistant for our students.

An improved tool will eventually be available to our students.

Am I Concerned About Students Cheating?

No

Students who are interested in succeeding will use similar tools to improve their understanding of course content.

Students who simply use these tools to cheat on weekly quizzes will not learn the material and will fail the proctored exams.

This Motivates Future Plans!

(perhaps) Train an **open-source** version of AI to create an online mathematics tutor for our students.

Starting Point: Llama Mixtral

https://ollama.com/library/mixtral:latest

We have loaded Llama Mixtral on one of our servers and will test its ability to perform on the question banks for our math courses.

Is Open-Source AI Up to the Task?

Definitely!!

Check this out!
AlphaGeometry AI: Google DeepMind
https://www.youtube.com/watch?v=WKF0QgxmGKs

Read
https://deepmind.google/discover/blog/alphageometry-an-olympiad-level-ai-system-for-geometry/

DeepMind Alpha Go Accomplishments

- **Recall:** New ChatGPT 4.0 scored 0 on the recent Mathematics Olympiad!

- Recently, DeepMind AlphaGo created a Mathematics AI that is 100 times smaller than ChatGPT 4.0. **It taught itself mathematics** and works **without human intervention. It recently scored just slightly below the top score on the most recent Mathematics Olympiad!**

- **It's open source!!**

Results from New ChatGPT 4.0 Turbo 0125

Math 1314	College Algebra	
Math 1324	Finite Math	
Math 1325	Business Calculus	
Math 1342	Elementary Stats	
Math 2305	Precalculus	
Math 2312	Calculus I	
Math 2413	Calculus II	
Math 2415	Calculus III	
Math 3321	Engineering Math	
Math 3339	Statistics for Sciences	