An Automated Graphing System for Mathematical Pedagogy

Arya Bulusu, Brandon Man, Ashish Jagmohan, Aditya Vempaty, Jennifer Mari-Wyka

MERLYN MIND™

Introduction

- Generative AI has significant potential to simplify the tools available to teachers in the classroom
- By creating automated tool-using systems, we allow teachers to more easily control tools such as graphing software
- We present an automated graphing system which converts utterances into mathematical expressions and graphs them with the Desmos interface
- Our design incorporates a mathematical solver into an LLM system, leading to more accurate, consistent responses
- The LLM allows for natural language understanding while the mathematical solver allows for mathematical accuracy

Dataset

Based on the Common Core standards, we identify a set of learning objectives that can be fulfilled through graphing. These categories form the basis of our two evaluation datasets. The utterance-focused dataset contains simple, single-step commands teachers would use in the classroom to demonstrate intermediate steps. The textbook-focused dataset contains multi-step, complicated problems requiring tool use to solve.

Example Row from Utterance-Focused Dataset	
Natural Language Utterance	Graph Input
Reflect y equals five x minus four across the y-axis	y = -5x - 4

Autoevaluator

Traditional evaluation metrics for text similarity cannot judge mathematical equivalence, so we need a method to precisely compare mathematical statements. We use the computer algebra system SymPy to compare LLM-produced mathematical expressions to our ground-truth expressions, with an LLM as a backup if SymPy cannot parse a given expression. We compare the results of the LLM-only and LLM+SymPy autoevaluators and find that the addition of Sympy significantly increases the accuracy of evaluations.

Results

We compare the performance of the LLM+Solver system and the LLM-only system, using GPT-4 as the LLM and Wolfram Alpha as the solver for evaluation.

Accuracy of LLM-only and LLM+Solver Systems		
	LLM-only	LLM+Solver
Utterance-Focused Dataset	63%	**85%**
Textbook-Focused Dataset	55%	**75%**

The addition of the solver results in a significant performance increase on both datasets. We see the greatest performance increase in categories with difficult, multi-step problems, such as Local Minima and Maxima and Tangents to Circles. The LLM-only system cannot solve these problems as it lacks mathematical reasoning. However, it is easy to write a Wolfram Alpha query to solve these problems, so the LLM+Solver consistently produces accurate answers.

Input question

```
Graph y equals x cubed minus 6 x squared plus 9 x plus 4 and find the relative extrema
```

LLM rephrases question into Wolfram Alpha query

```
Find the relative extrema of y = x^3 - 6x^2 + 9x + 4
```

Wolfram Alpha provides result and step-by-step solution

```
Step-by-step Solution: (truncated)
max{x^3 - 6 x^2 + 9 x + 4} = 8 at x = 1
Find and classify the local extrema of the following function using the first derivative test:
f(x) = x^3 - 6 x^2 + 9 x + 4
Find the critical points of f(x):
Compute the critical points of x^3 - 6 x^2 + 9 x + 4
To find all critical points, first compute f'(x):
d/(dx) (x^3 - 6 x^2 + 9 x + 4) = 3 x^2 - 12 x + 9
 = 3 (x - 3) (x - 1):
f'(x) = 3 (x - 3) (x - 1)
...
Answer: |
 | f(x) has a local maximum at x = 1
f(x) has a local minimum at x = 3
```

LLM writes detailed explanation based on Wolfram Alpha's output

```
Explanation: (truncated)
The problem is asking us to graph the function y = x^3 - 6x^2 + 9x + 4 and find the relative extrema. The relative extrema of a function are the local maximum and minimum points, where the function changes direction.

The first step in finding the relative extrema is to find the critical points of the function. Critical points are where the derivative of the function is either zero or undefined.

To find the critical points, we first need to find the derivative of the function. The derivative of x^3 - 6x^2 + 9x + 4 is 3x^2 - 12x + 9. This can be factored to 3(x - 3)(x - 1).

Setting this equal to zero gives us the critical points x = 1 and x = 3. The derivative exists everywhere, so these are the only critical points.

Next, we partition the domain of the function into intervals using the critical points as endpoints. This gives us the intervals (-∞, 1), (1, 3), and (3, ∞).

...
```

LLM writes Desmos expressions and chain of thought

```
Thought: I need to create an expression for the equation f(x) = x^3 - 6x^2 + 9x + 4. In order to graph the local maximum and minimum, I need to solve for the y-values at x = 1 and x = 3.
f(1) = (1)^3 - 6(1)^2 + 9(1) + 4 = 1 - 6 + 9 + 4 = 8
f(3) = (3)^3 - 6(3)^2 + 9(3) + 4 = 27 - 54 + 27 + 4 = 4
The local maximum occurs at (1, 8) and the local minimum occurs at (3, 4).
Desmos expressions: [
{"action": "setExpression", "expression": {"type":"expression","id":"function","latex":"f(x) = x^3 - 6x^2 + 9x + 4","hidden":false}},
{"action": "setExpression", "expression": {"type":"expression","id":"local_maximum","latex":"(1,8)","hidden":false}},
{"action": "setExpression", "expression": {"type":"expression","id":"local_minimum","latex":"(3,4)","hidden":false}}]
```

Graph is produced with Desmos