

Introducing ChatGPT as a Learning Tool to Students

Robert J. McKeown* mckeow99@yorku.ca
*York University

1. Large Language Models (LLMs) can improve learning and productivity

- LLMs like ChatGPT helped new workers move down the learning curve (Brynjolfsson et al. (2023), Noy and Zhang (2023)).
- LLMs can act like a personal tutor (Mollick and Mollick (2023)), and personal tutoring improved student outcomes by 0.7 to 2.0 standard deviations (VanLehn (2011), Bloom (1984)).

2. Introducing students to an LLM

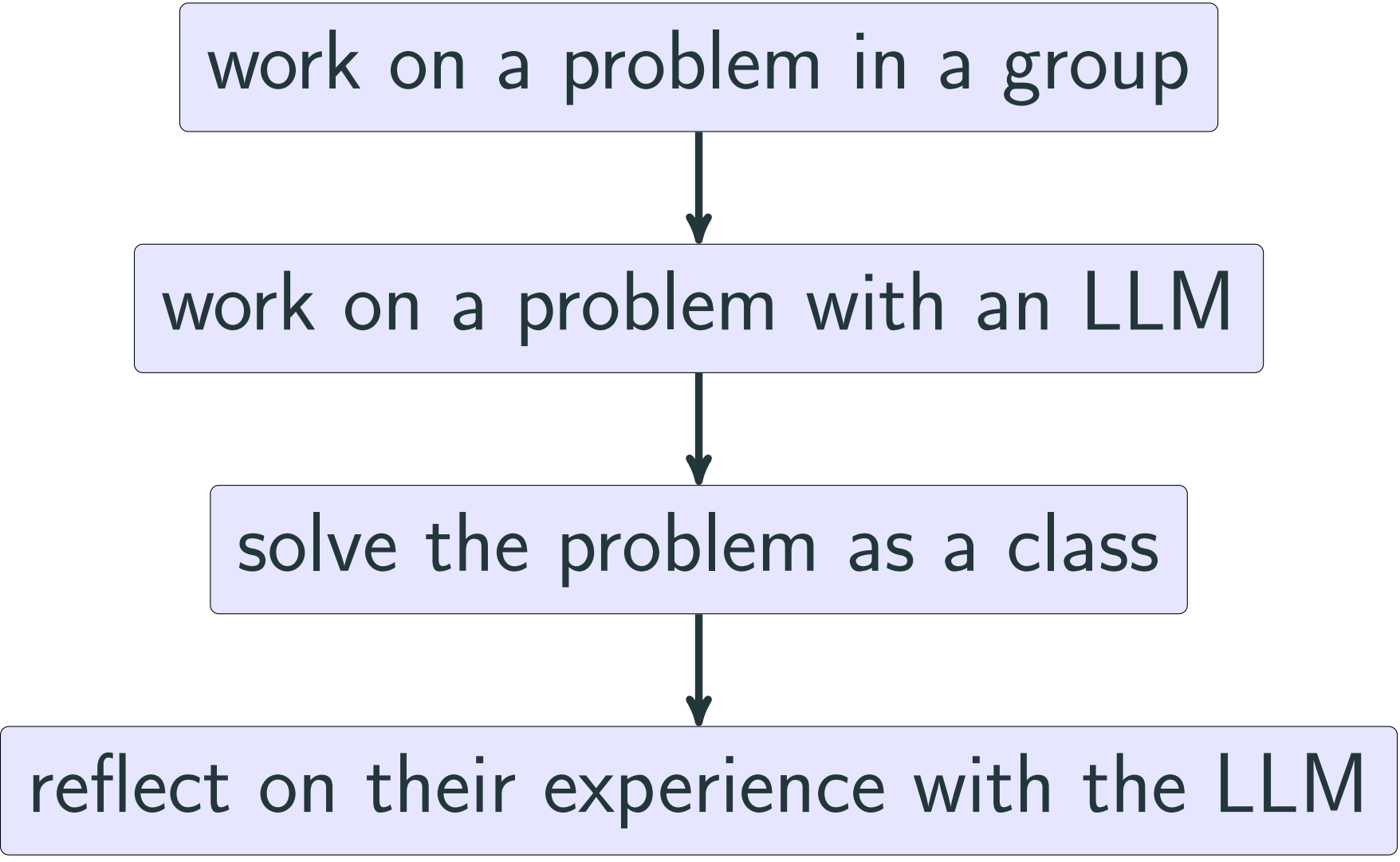
- Introduce students to using an LLM as a personal tutor under instructor supervision.
- Take advantage of a flipped classroom format for a class exercise.
- Catch and correct any hallucinations, mistakes, or misconceptions in LLM output.
- Keep the exercise short (15-20 minutes), so all course content can be covered.

3. Instructor Prepares the Exercise

- Chooses *one problem* for the LLM to tutor students with.
- Prepares a prompt for the LLM.
- Prepares instructions (i.e. how to setup an account) for students to access LLM.
 - Bing AI is free.
- Asks students to bring their laptop or smartphone to class.
- Deliver the problem, instructions, and prompt to students.

4. Execute the Exercise

Students:



- Next, students worked on a longer **problem set** in groups where using a LLM was permissible but not required.

5. Example of a Problem with Prompt

- **Problem:** Illustrate the chain rule to find dy/dx when $y = u^5$ and $u = 1 - x^3$
- **Prompt:** *Act as if you are a math tutor and I am your tutee who is a first-year undergraduate. You are a capable, energetic, and supportive tutor. Use the Socratic method, asking questions to help me understand the concepts and solve the problem myself. Do not jump to the final answer. Make your answer easy to understand for a student new to calculus. Reference a composite function and the inner and outer functions in your answer.*

Introduced first-year undergraduates to using ChatGPT as a learning tool in a flipped (reverse) classroom.

On average, students reported ChatGPT was just as helpful as their groupmates in understanding course concepts.

Students who completed in-class problem sets earned higher grades on a calculus term test

References

Bloom, Benjamin S (1984). "The 2 sigma problem: The search for methods of group instruction as effective as one-to-one tutoring." *Educational researcher*, 13, 4–16.

Brynjolfsson, Erik, Danielle Li, and Lindsey R Raymond (2023). "Generative ai at work." Technical report, National Bureau of Economic Research.

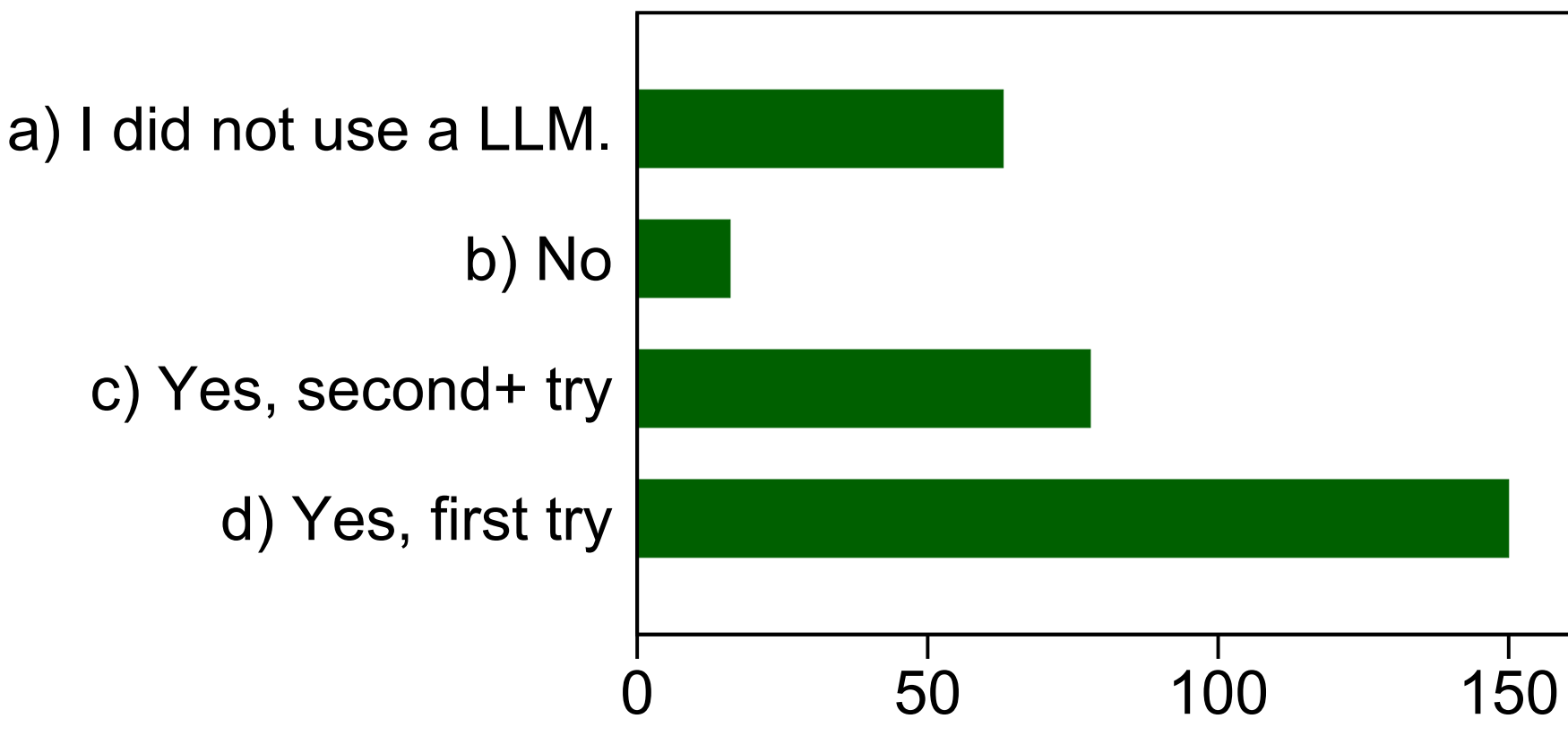
Mollick, Ethan R and Lilach Mollick (2023). "Using AI to implement effective teaching strategies in classrooms: Five strategies, including prompts." *March 17, 2023*.

Noy, Shakkeed and Whitney Zhang (2023). "Experimental evidence on the productivity effects of generative artificial intelligence." *Available at SSRN 4375283*.

VanLehn, Kurt (2011). "The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems." *Educational psychologist*, 46, 197–221.

6. Evaluating Accuracy: LLM’s Correct Answer Count for Assigned Problem

Figure 1: Did the LLM solve the problem correctly today?



Note: observations: 307

7. Comparative Helpfulness: Groupmates v LLM

Table 1: How helpful was your LLM (groupmates) in understanding today’s problem set

	Oct19	Oct26	Nov3	Nov10	Nov24	Dec1	Total
The LLM was:							
more helpful,	9	6	0	4	4	4	27
equally helpful,	15	17	18	14	14	17	95
less helpful than my groupmates	7	7	6	2	2	2	26

Students separately evaluated their groupmates and their LLM experience on their helpfulness in learning using a 1-to-5 Likert scale. This table summarizes whether the groupmates or LLM received a higher helpfulness score. Observations: 128.

8. Comparative Helpfulness: Adaptive Learning Software (ALS) v LLM

Figure 2: Which software should be used again in this class?

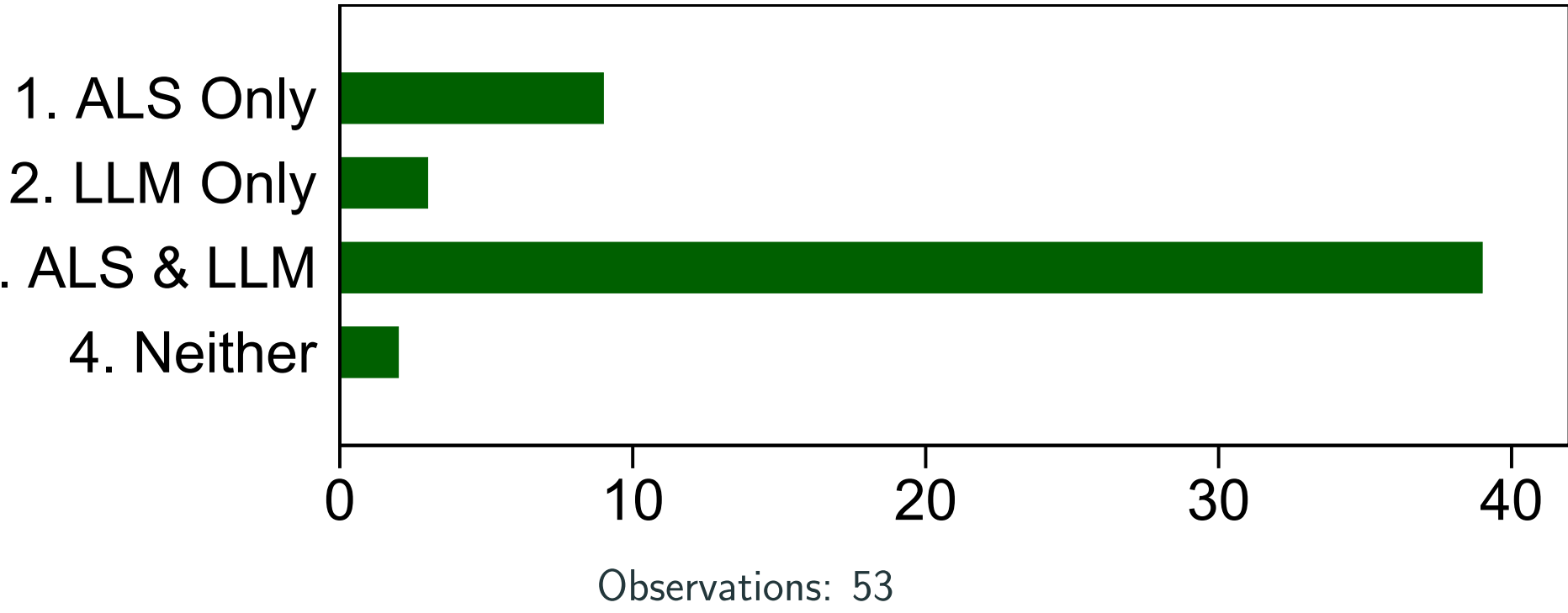
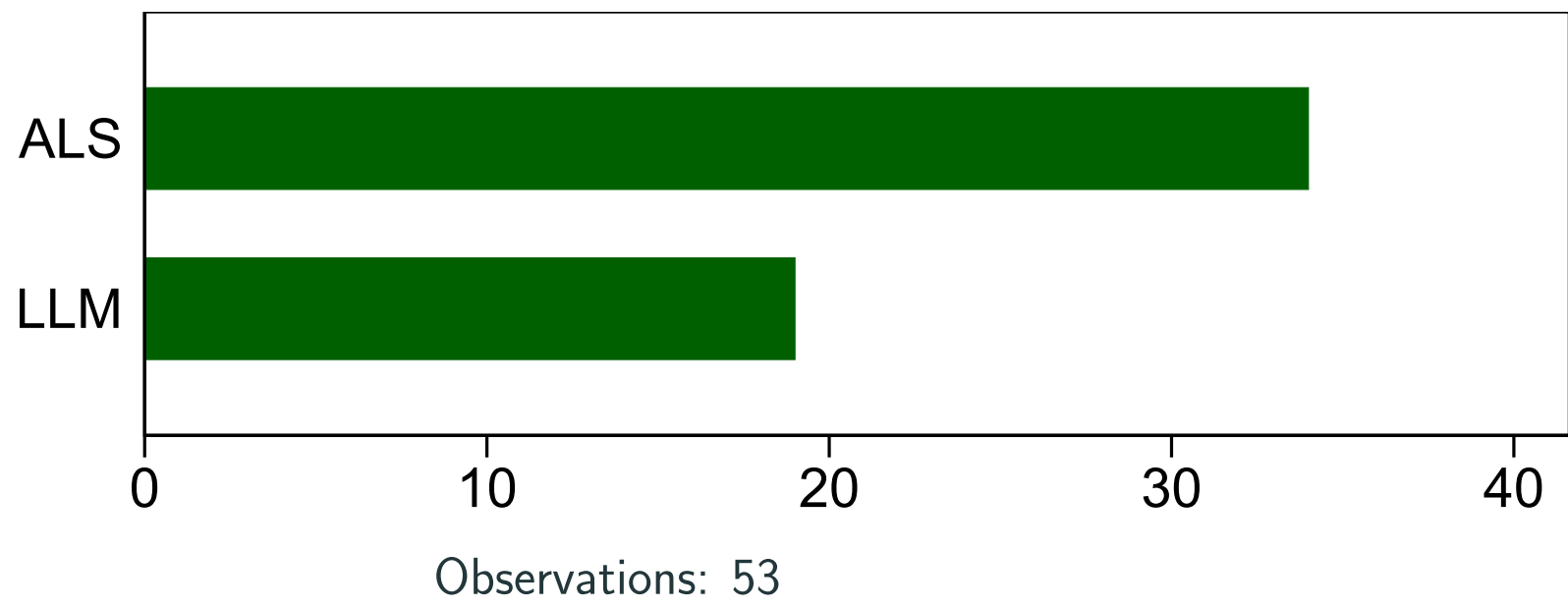
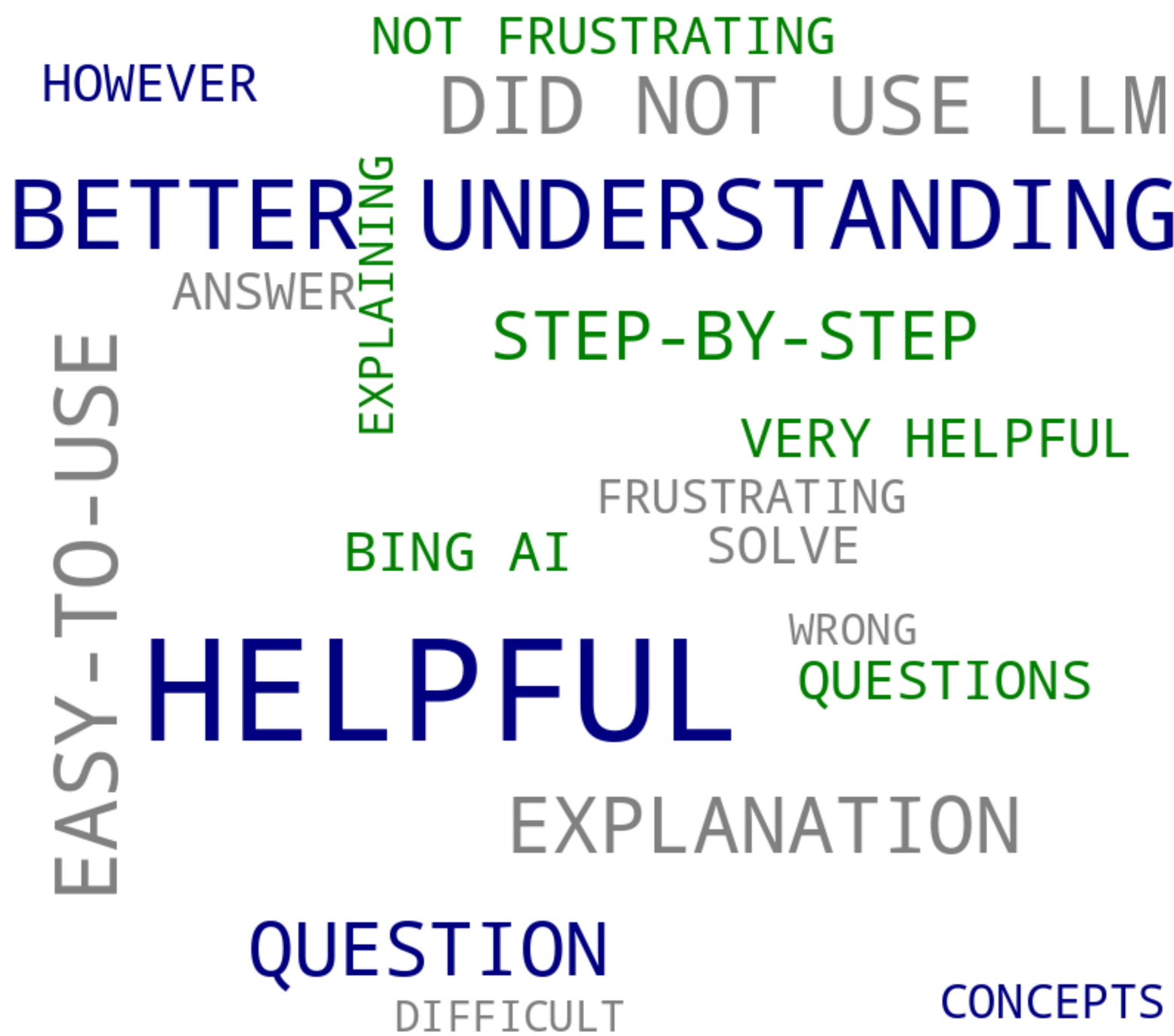


Figure 3: Which software contributed more to your learning



9. Qualitative Feedback

Figure 4: Describe your experience using an LLM



10. Regression Analysis

Table 2: Calculus term test grade regression results

Dep.Var.:	Calculus Term Test	Coef.	Std.Err.	t-value
Other Controls		–	–	–
HS calculus grade		0.004	0.006	0.67
Used LLM before		0.168	0.163	1.03
English spoken at home		0.405	0.178	2.28 **
Lecture video questions		0.096	0.104	0.92
Online practice questions		0.244	0.116	2.11 **
Online quizzes		0.130	0.095	1.37 *
Term test 1 (prcalculus)		0.603	0.117	5.17 ***
In-class problem set		0.493	0.266	1.85 *
Used LLM on problem set		0.402	0.289	1.39
Constant		-0.897	0.367	-2.44 **
Number of obs	69		R ²	0.757
F-test	11.02		Adj. R ²	0.688
F-test (In-class problem set + Used LLM on problem set = 0)				4.34 **

- Students who completed 100% of their in-class problem sets scored 0.493 standard deviations higher on a calculus term test.