

Large language models are effective for summarizing student feedback



Presenter:
Nicholas T. Young¹



Download poster

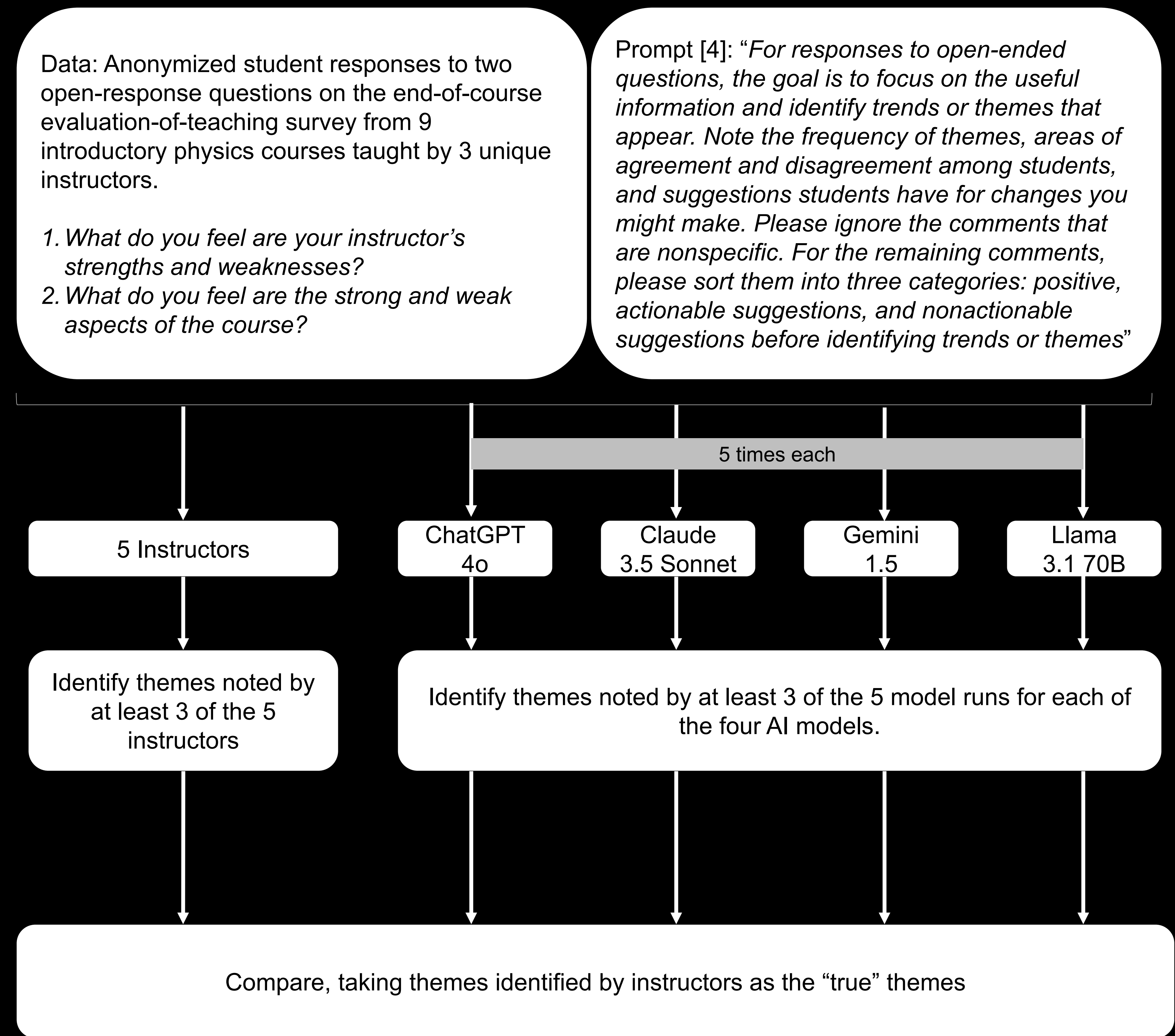
Christopher Overton¹, Ania Majewska², Hina Shaikh^{1,3}, Nandana Weliweriya¹

¹ Department of Physics and Astronomy, University of Georgia
² Department of Physiology and Pharmacology, College of Veterinary Medicine, University of Georgia
³ Institute for Astronomy and Astrophysics, Eberhard Karls University of Tübingen

Introduction

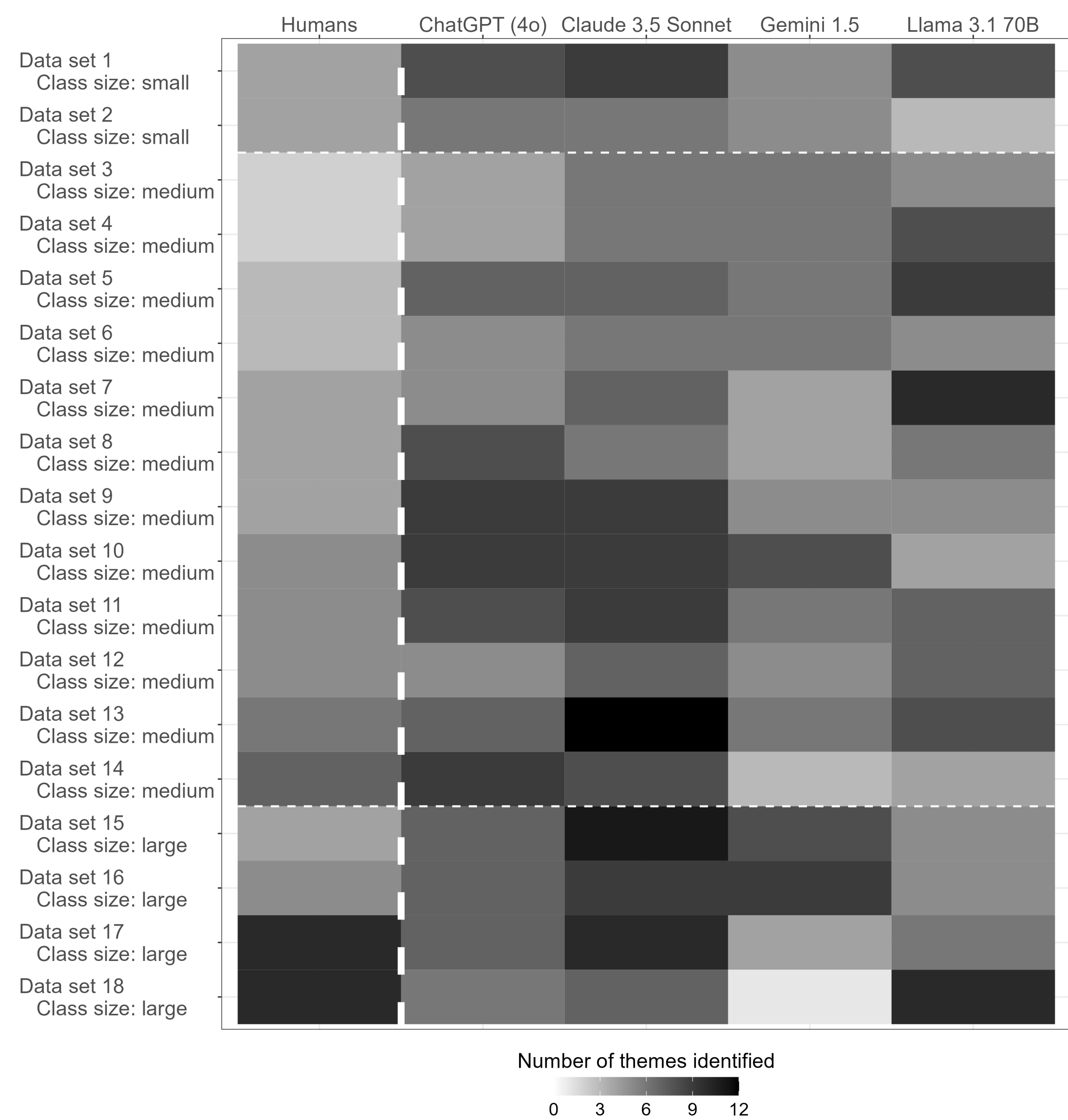
- Collecting and acting on student feedback is an important method for instructors to adapt their teaching to student needs [1].
- Analyzing feedback from students in large-enrollment introductory courses, such as introductory physics at large universities, can be time-consuming for instructors.
- Generative AI tools are effective at producing summaries of text [2, 3] and therefore offer a potential solution for instructors to quickly extract key points from student feedback.
- Here, we compare generative AI’s ability to extract key themes and trends from student feedback compared to university instructors.

Methods



Result: Generative AI tools, such as ChatGPT, can extract themes from student feedback about as effectively as instructors can.

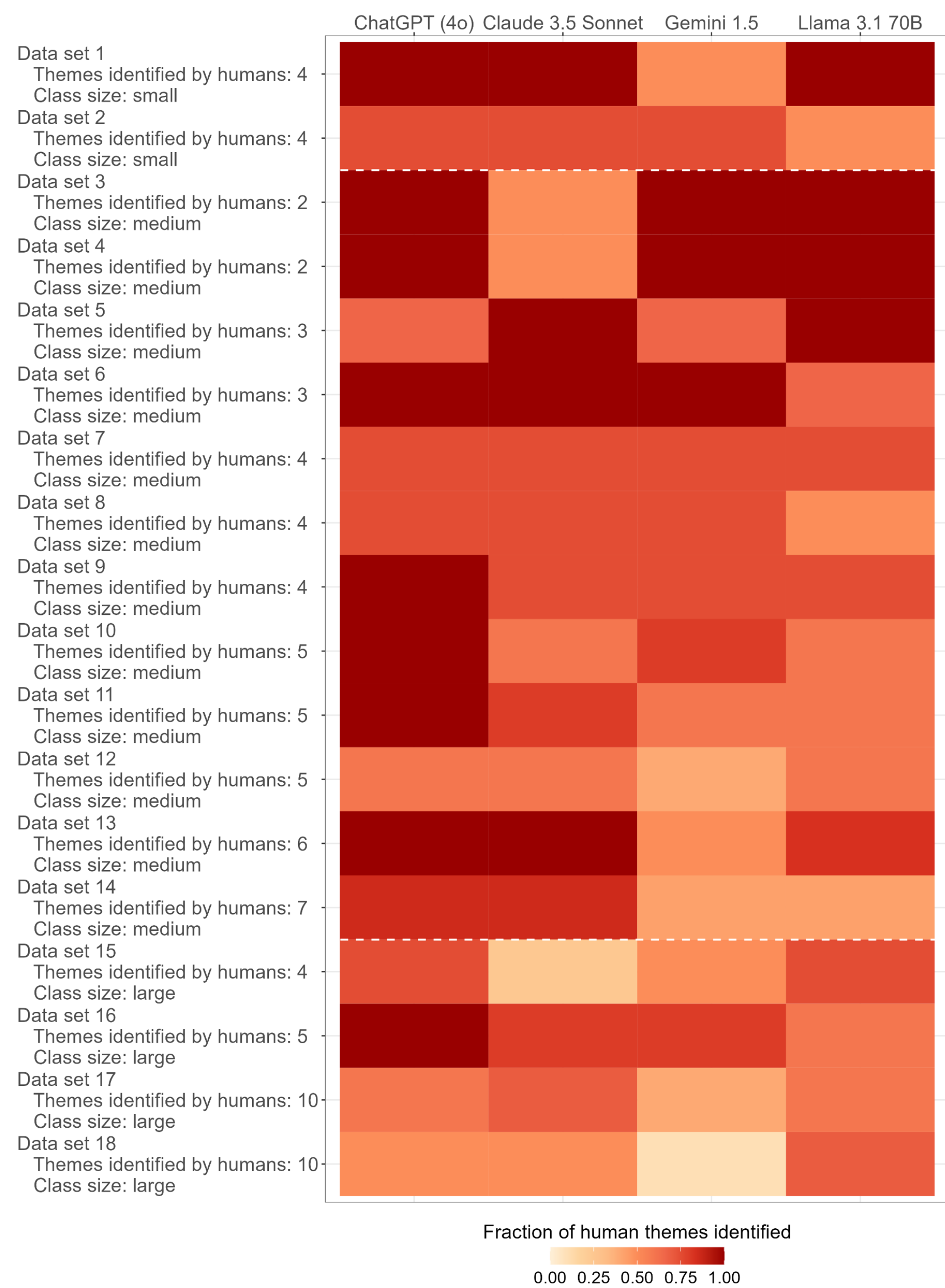
How do models compare in their ability to find themes in the student feedback?



ChatGPT and Claude tended to find more themes in the course feedback than instructors did.

Gemini and Llama tended to have mixed results.

Did the themes identified by AI tools align with what instructors identified?



For ChatGPT and Claude, yes. For Gemini and Llama, less so.

In many cases, ChatGPT identified all of the themes the instructors identified while this rarely happened for Gemini or Llama.

ChatGPT was the best or tied for best performing model on 15 of the 18 data sets.

Performance of all models was generally best for data sets with fewer human-identified themes.

References

[1] Spooren, P., Brockx, B., & Mortelmans, D. (2013). *Review of Educational Research*, 83(4), 598-642.

[2] Pu, X., Gao, M., & Wan, X. (2023). *arXiv preprint arXiv:2309.09558*.

[3] Parker, M. J., Anderson, C., Stone, C., & Oh, Y. (2024). *International journal of artificial intelligence in education*, 1-38.

[4] Based on <https://ctl.uga.edu/teaching-resources/feedback-and-evaluation-of-teaching/interpreting-responding-to-student-evaluations-of-teaching/>