### Large language models are effective for summarizing student feedback



Presenter: Nicholas T. Young<sup>1</sup>

Christopher Overton<sup>1</sup>, Ania Majewska<sup>2</sup>, Hina Shaikh<sup>1,3</sup>, Nandana Weliweriya<sup>1</sup>

<sup>1</sup> Department of Physics and Astronomy, University of Georgia <sup>2</sup> Department of Physiology and Pharmacology, College of Veterinary Medicine, University of Georgia

<sup>3</sup> Institute for Astronomy and Astrophysics, Eberhard Karls University of Tübingen

#### Introduction

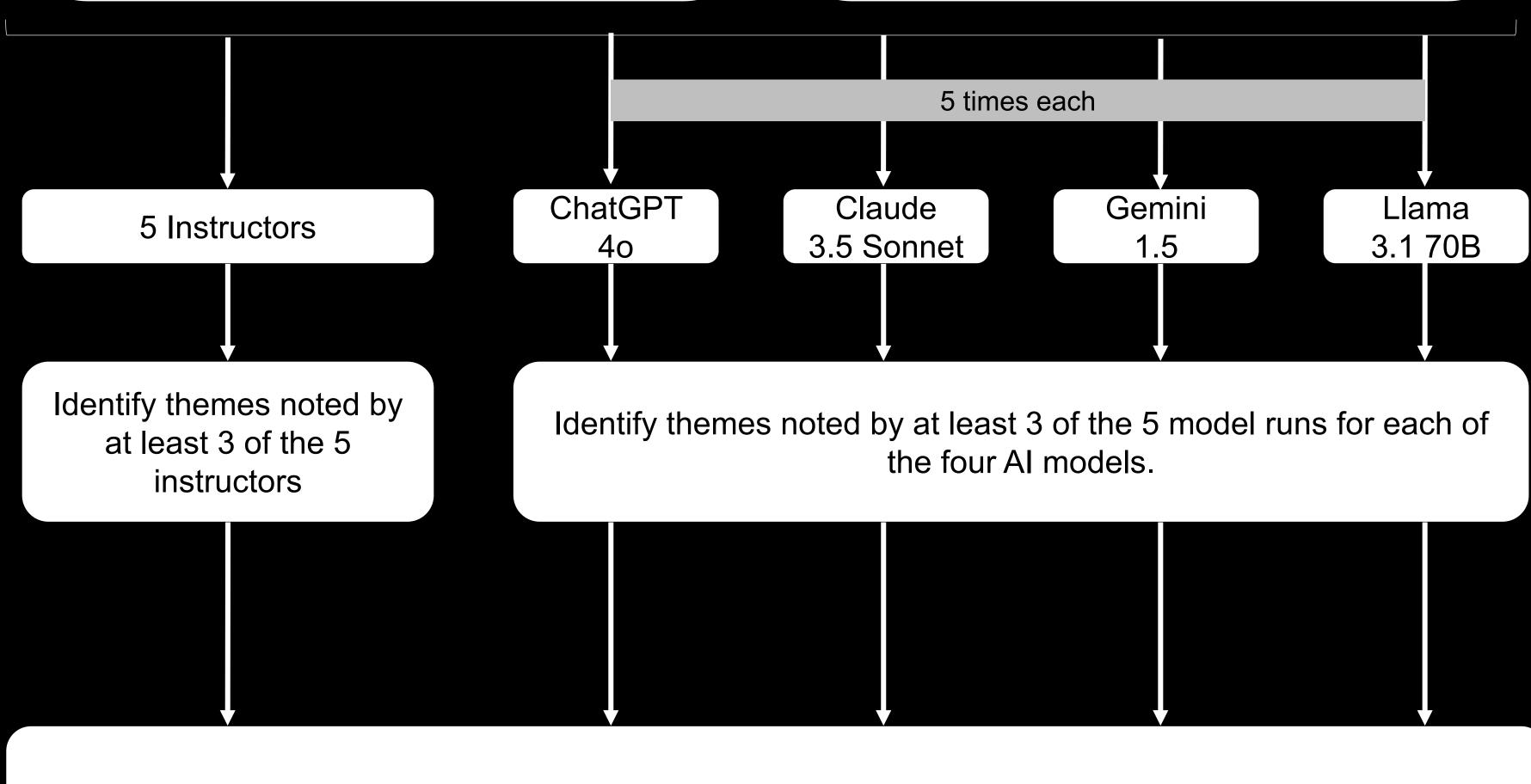
- Collecting and acting on student feedback is an important method for instructors to adapt their teaching to student needs [1].
- Analyzing feedback from students in large-enrollment introductory courses, such as introductory physics at large universities, can be time-consuming for instructors.
- Generative 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

Data: Anonymized student responses to two open-response questions on the end-of-course evaluation-of-teaching survey from 9 introductory physics courses taught by 3 unique instructors.

- 1. What do you feel are your instructor's strengths and weaknesses?
- 2. What do you feel are the strong and weak aspects of the course?

Prompt [4]: "For responses to open-ended questions, the goal is to focus on the useful information and identify trends or themes that appear. Note the frequency of themes, areas of agreement and disagreement among students, and suggestions students have for changes you might make. Please ignore the comments that are nonspecific. For the remaining comments, please sort them into three categories: positive, actionable suggestions, and nonactionable suggestions before identifying trends or themes"



Compare, taking themes identified by instructors as the "true" themes



**Georgia Physics and Astronomy Education Research** Franklin College of Arts and Sciences **UNIVERSITY OF GEORGIA** 





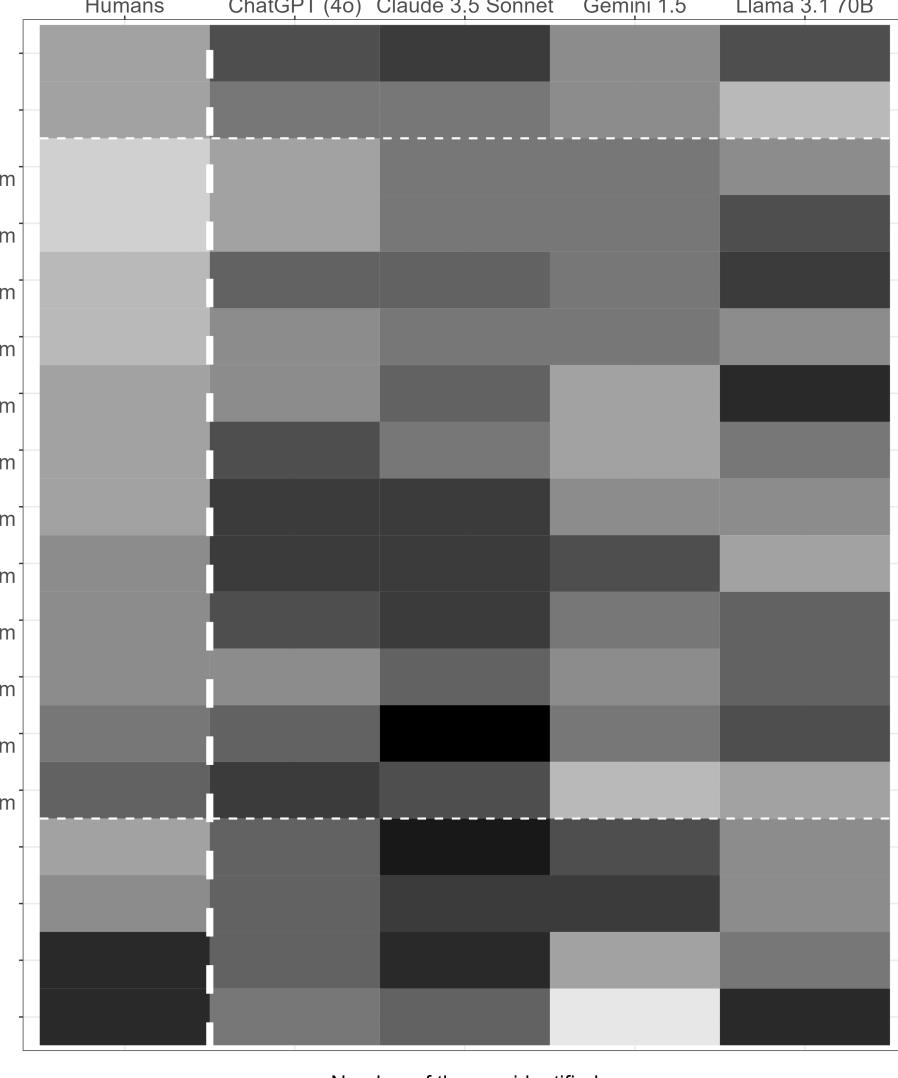


## Result: Generative AI tools, such as ChatGPT, can extract themes from student feedback about as effectively as instructors can.

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

Data set 1 Class size: small Data set 2 Class size: small Data set 3 Class size: medium Data set 4 Class size: medium Data set 5 Class size: medium Data set 6 Class size: medium Data set 7 Class size: medium Data set 8 Class size: medium Data set 9 Class size: medium Data set 10 Class size: medium Data set 11 Class size: medium Data set 12 Class size: medium Data set 13 Class size: medium Data set 14 Class size: medium Data set 15 Class size: large Data set 16 Class size: large Data set 17 Class size: large Data set 18

Class size: larg



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

Data set 1 Themes identified Class size: small Data set 2 Themes identified Class size: small Data set 3 Themes identified Class size: mediu Data set 4 Themes identified Class size: mediu Data set 5 Themes identified Class size: mediu Data set 6 Themes identified Class size: mediu Data set 7 Themes identified Class size: mediu Data set 8 Themes identified Class size: mediu Data set 8 Themes identified Class size: mediu Data set 9 Themes identified Class size: mediu Data set 10 Themes identified Class size: mediu Data set 11 Themes identified Class size: mediu Data set 11 Themes identified Class size: mediu Data set 12 Themes identified Class size: mediu Data set 13 Themes identified Class size: mediu Data set 13 Themes identified Class size: mediu Data set 14 Themes identified Class size: mediu Data set 15 Themes identified Class size: large Data set 16 Themes identified
Themes identified Class size: small Data set 2 Themes identified Class size: small Data set 3 Themes identified Class size: mediu Data set 4 Themes identified Class size: mediu Data set 5 Themes identified Class size: mediu Data set 6 Themes identified Class size: mediu Data set 7 Themes identified Class size: mediu Data set 8 Themes identified Class size: mediu Data set 8 Themes identified Class size: mediu Data set 9 Themes identified Class size: mediu Data set 10 Themes identified Class size: mediu Data set 10 Themes identified Class size: mediu Data set 11 Themes identified Class size: mediu Data set 12 Themes identified Class size: mediu Data set 13 Themes identified Class size: mediu Data set 13 Themes identified Class size: mediu Data set 14 Themes identified Class size: mediu Data set 15 Themes identified Class size: large Data set 16
Data set 2 Themes identified Class size: small Data set 3 Themes identified Class size: mediu Data set 4 Themes identified Class size: mediu Data set 5 Themes identified Class size: mediu Data set 6 Themes identified Class size: mediu Data set 7 Themes identified Class size: mediu Data set 8 Themes identified Class size: mediu Data set 8 Themes identified Class size: mediu Data set 9 Themes identified Class size: mediu Data set 10 Themes identified Class size: mediu Data set 11 Themes identified Class size: mediu Data set 11 Themes identified Class size: mediu Data set 11 Themes identified Class size: mediu Data set 12 Themes identified Class size: mediu Data set 13 Themes identified Class size: mediu Data set 13 Themes identified Class size: mediu Data set 14 Themes identified Class size: mediu Data set 15 Themes identified Class size: large Data set 16
Class size: small Data set 3 Themes identified Class size: mediu Data set 4 Themes identified Class size: mediu Data set 5 Themes identified Class size: mediu Data set 6 Themes identified Class size: mediu Data set 7 Themes identified Class size: mediu Data set 8 Themes identified Class size: mediu Data set 8 Themes identified Class size: mediu Data set 9 Themes identified Class size: mediu Data set 10 Themes identified Class size: mediu Data set 11 Themes identified Class size: mediu Data set 12 Themes identified Class size: mediu Data set 12 Themes identified Class size: mediu Data set 13 Themes identified Class size: mediu Data set 13 Themes identified Class size: mediu Data set 13 Themes identified Class size: mediu Data set 14 Themes identified Class size: mediu Data set 15 Themes identified Class size: large Data set 16
Themes identified Class size: medium Data set 4 Themes identified Class size: medium Data set 5 Themes identified Class size: medium Data set 6 Themes identified Class size: medium Data set 7 Themes identified Class size: medium Data set 8 Themes identified Class size: medium Data set 9 Themes identified Class size: medium Data set 9 Themes identified Class size: medium Data set 10 Themes identified Class size: medium Data set 11 Themes identified Class size: medium Data set 12 Themes identified Class size: medium Data set 12 Themes identified Class size: medium Data set 13 Themes identified Class size: medium Data set 13 Themes identified Class size: medium Data set 13 Themes identified Class size: medium Data set 14 Themes identified Class size: medium Data set 15 Themes identified Class size: large Data set 16
Data set 4 Themes identified Class size: mediu Data set 5 Themes identified Class size: mediu Data set 6 Themes identified Class size: mediu Data set 7 Themes identified Class size: mediu Data set 8 Themes identified Class size: mediu Data set 9 Themes identified Class size: mediu Data set 10 Themes identified Class size: mediu Data set 11 Themes identified Class size: mediu Data set 12 Themes identified Class size: mediu Data set 13 Themes identified Class size: mediu Data set 13 Themes identified Class size: mediu Data set 13 Themes identified Class size: mediu Data set 14 Themes identified Class size: mediu Data set 15 Themes identified Class size: large Data set 16
Class size: medium Data set 5 Themes identified Class size: medium Data set 6 Themes identified Class size: medium Data set 7 Themes identified Class size: medium Data set 8 Themes identified Class size: medium Data set 9 Themes identified Class size: medium Data set 10 Themes identified Class size: medium Data set 11 Themes identified Class size: medium Data set 12 Themes identified Class size: medium Data set 12 Themes identified Class size: medium Data set 13 Themes identified Class size: medium Data set 13 Themes identified Class size: medium Data set 14 Themes identified Class size: medium Data set 14 Themes identified Class size: medium Data set 15 Themes identified Class size: large Data set 16
Themes identified Class size: mediu Data set 6 Themes identified Class size: mediu Data set 7 Themes identified Class size: mediu Data set 8 Themes identified Class size: mediu Data set 9 Themes identified Class size: mediu Data set 10 Themes identified Class size: mediu Data set 11 Themes identified Class size: mediu Data set 12 Themes identified Class size: mediu Data set 13 Themes identified Class size: mediu Data set 13 Themes identified Class size: mediu Data set 14 Themes identified Class size: mediu Data set 14 Themes identified Class size: mediu Data set 15 Themes identified Class size: large Data set 16
Data set 6 Themes identified Class size: medium Data set 7 Themes identified Class size: medium Data set 8 Themes identified Class size: medium Data set 9 Themes identified Class size: medium Data set 10 Themes identified Class size: medium Data set 11 Themes identified Class size: medium Data set 12 Themes identified Class size: medium Data set 12 Themes identified Class size: medium Data set 13 Themes identified Class size: medium Data set 13 Themes identified Class size: medium Data set 14 Themes identified Class size: medium Data set 14 Themes identified Class size: medium Data set 15 Themes identified Class size: large Data set 16
Themes identified Class size: medium Data set 7 Themes identified Class size: medium Data set 8 Themes identified Class size: medium Data set 9 Themes identified Class size: medium Data set 10 Themes identified Class size: medium Data set 11 Themes identified Class size: medium Data set 12 Themes identified Class size: medium Data set 13 Themes identified Class size: medium Data set 13 Themes identified Class size: medium Data set 14 Themes identified Class size: medium Data set 14 Themes identified Class size: medium Data set 15 Themes identified Class size: large Data set 16
Data set 7 Themes identified Class size: medium Data set 8 Themes identified Class size: medium Data set 9 Themes identified Class size: medium Data set 10 Themes identified Class size: medium Data set 11 Themes identified Class size: medium Data set 12 Themes identified Class size: medium Data set 13 Themes identified Class size: medium Data set 13 Themes identified Class size: medium Data set 14 Themes identified Class size: medium Data set 14 Themes identified Class size: medium Data set 15 Themes identified Class size: large Data set 16
Class size: medium Data set 8 Themes identified Class size: medium Data set 9 Themes identified Class size: medium Data set 10 Themes identified Class size: medium Data set 11 Themes identified Class size: medium Data set 12 Themes identified Class size: medium Data set 13 Themes identified Class size: medium Data set 14 Themes identified Class size: medium Data set 14 Themes identified Class size: medium Data set 15 Themes identified Class size: large Data set 16
Themes identified Class size: medium Data set 9 Themes identified Class size: medium Data set 10 Themes identified Class size: medium Data set 11 Themes identified Class size: medium Data set 12 Themes identified Class size: medium Data set 13 Themes identified Class size: medium Data set 14 Themes identified Class size: medium Data set 14 Themes identified Class size: medium Data set 15 Themes identified Class size: large Data set 16
Data set 9 Themes identified Class size: medium Data set 10 Themes identified Class size: medium Data set 11 Themes identified Class size: medium Data set 12 Themes identified Class size: medium Data set 13 Themes identified Class size: medium Data set 14 Themes identified Class size: medium Data set 14 Themes identified Class size: medium Data set 15 Themes identified Class size: large Data set 16
Class size: medium Data set 10 Themes identified Class size: medium Data set 11 Themes identified Class size: medium Data set 12 Themes identified Class size: medium Data set 13 Themes identified Class size: medium Data set 14 Themes identified Class size: medium Data set 14 Themes identified Class size: medium Data set 15 Themes identified Class size: large Data set 16
Themes identified Class size: medium Data set 11 Themes identified Class size: medium Data set 12 Themes identified Class size: medium Data set 13 Themes identified Class size: medium Data set 14 Themes identified Class size: medium Data set 15 Themes identified Class size: large Data set 16
Data set 11 Themes identified Class size: medium Data set 12 Themes identified Class size: medium Data set 13 Themes identified Class size: medium Data set 14 Themes identified Class size: medium Data set 15 Themes identified Class size: large Data set 16
Themes identified Class size: medium Data set 12 Themes identified Class size: medium Data set 13 Themes identified Class size: medium Data set 14 Themes identified Class size: medium Data set 15 Themes identified Class size: large Data set 16
Data set 12 Themes identified Class size: medium Data set 13 Themes identified Class size: medium Data set 14 Themes identified Class size: medium Data set 15 Themes identified Class size: large Data set 16
Class size: medium Data set 13 Themes identified Class size: medium Data set 14 Themes identified Class size: medium Data set 15 Themes identified Class size: large Data set 16
Themes identified Class size: mediu Data set 14 Themes identified Class size: mediu Data set 15 Themes identified Class size: large Data set 16
Data set 14 Themes identified Class size: mediu Data set 15 Themes identified Class size: large Data set 16
Class size: mediu Data set 15 Themes identified Class size: large Data set 16
Themes identified Class size: large Data set 16
Data set 16
Thomas identified
Class size: large
Data set 17 Themes identified
Class size: large Data set 18
Themes identified Class size: large

# Research, 83(4), 598-642.

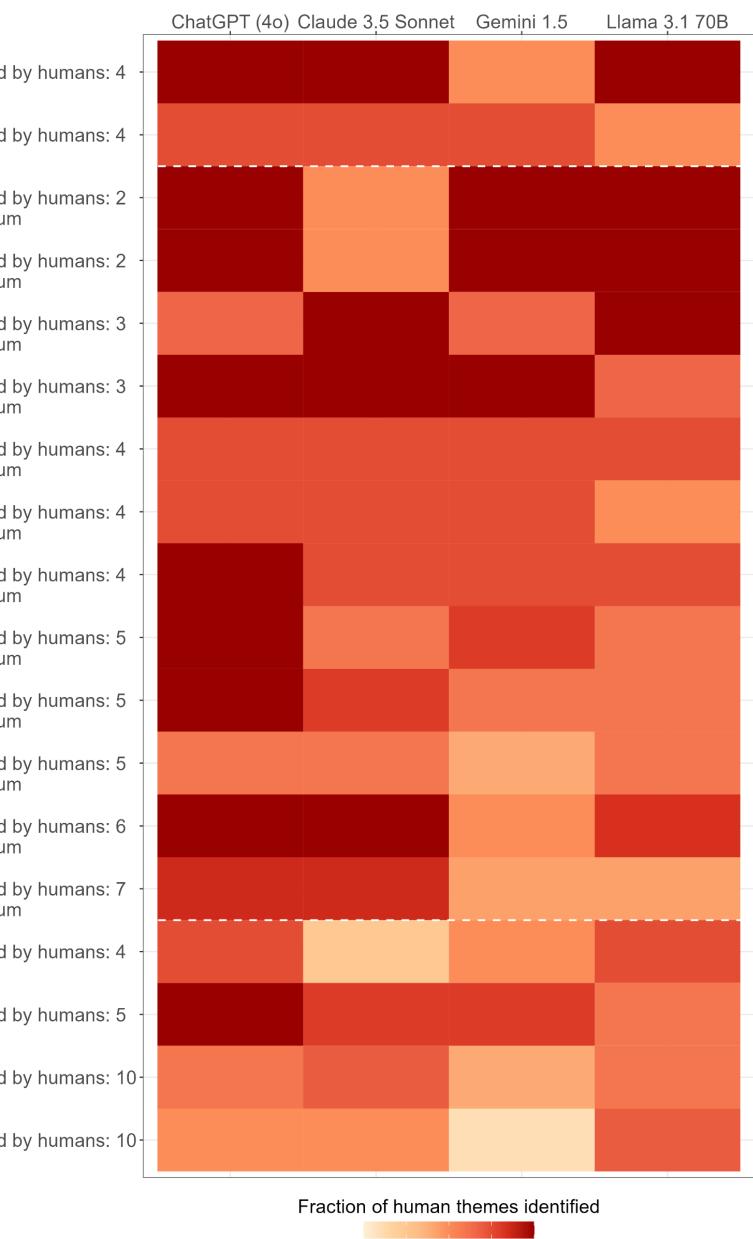
[2] Pu, X., Gao, M., & Wan, X. (2023). arXiv preprint arXiv:2309.09558. [3] Parker, M. J., Anderson, C., Stone, C., & Oh, Y. (2024). International journal of artificial intelligence in education, 1-38.

ChatGPT (4o) Claude 3.5 Sonnet Gemini 1.5 Llama 3.1 70B

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

Gemini and Llama tended to have mixed results.

Number of themes identified 0 3 6 9 12



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.

#### 0.00 0.25 0.50 0.75 1.00

#### References

[1] Spooren, P., Brockx, B., & Mortelmans, D. (2013). *Review of Educational*