IOWA STATE UNIVERSITY
Department of Psychology

Honors Poster Presentation
May 1, 2024

Nicholas Rogers Jason C.K. Chan Dahwi Ahn Zohara Assadipour

Streamlining Data Coding: GPT Automation of Human Responses in Behavioral Sciences

Background and Objective

Coding experimental data is an essential part of quantitative analysis in behavioral sciences, but this process can be resource-intensive. Researchers have to be trained and dedicate many hours to code verbal responses. By increasing the number of researchers, the time it takes and the burden on individuals is reduced. However, this is not always feasible, especially for smaller labs. OpenAI's GPT has the potential to transform data coding because it does not require a massive data set to train an ML algorithm. This study examines the effectiveness of GPT for the idea units coding relative to trained human coders.

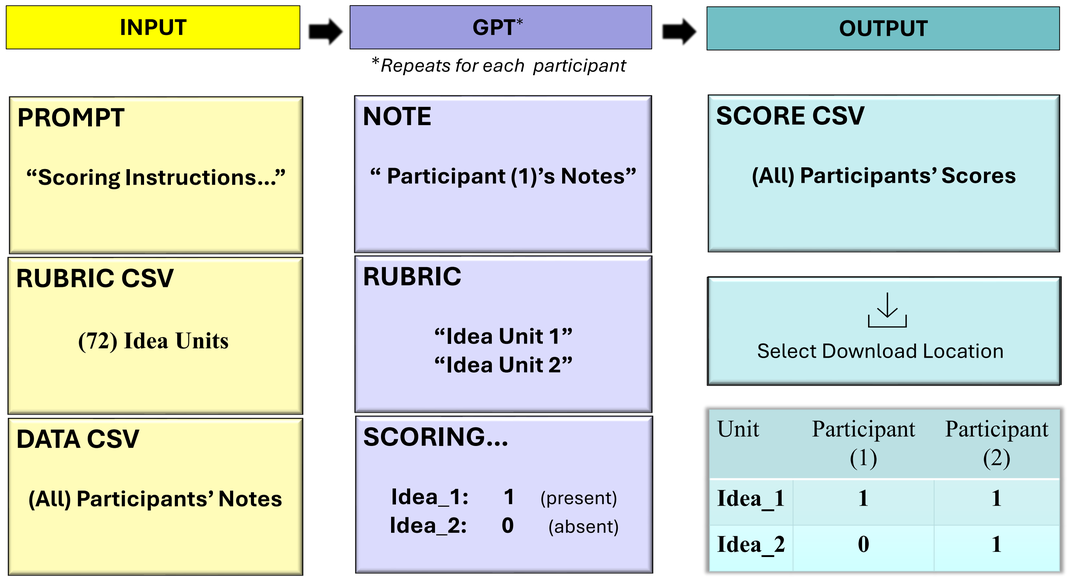
Method

Behavioral Experiment:
Undergraduate participants at ISU watched a lecture video about a STEM topic. They were instructed to take notes while watching the video by typing their notes into the computer. For the lecture video, we identified a total of ~72 idea units. Four undergraduate research assistants were trained to code the note-taking data for the presence/absence of each idea unit. We then compared the results of human coding to GPT's coding.

Prompt Engineering:
1. Researchers engineered initial prompt and asked GPT to scored a small set of data.
2. GPT produced coding. Researchers identified errors, asked GPT for explanations, and then modified the prompt/rubric or asked GPT for prompt modification suggestions.
3. GPT coded data with the modified prompt.
4. Repeat 1-3.

Program: A bespoke program is written for automated data scoring. User uploads two CSV files: A data file that contained all subjects' notes and a rubric file that contains all of the idea units to be identified. The program goes interfaces with OpenAI's GPT API and records GPT's data coding output into a CSV file. In addition, the program identifies missing outputs from GPT and requests GPT to re-code until no missing data remain.

Procedure



Results

% Agreement	Coder 2	Coder 3	Coder 4	GPT
Coder 1	91%	88%	84%	**85%**
Coder 2		89%	86%	**84%**
Coder 3			87%	**83%**
Coder 4				**79%**

Conclusions

	Avg % agreement with all coders
Coder 1	87%
Coder 2	88%
Coder 3	87%
Coder 4	84%
GPT	**83%**

Automated coding of complex textual data by GPT produced results that are comparable to those of trained human coders.
These results demonstrate the viability of using GPT to code human responses in behavioral studies without the need to train a specialized ML algorithm.

GitHub Repository

Documentation for the program will be provided via a public GitHub page. Included will be a copy of the code, a how to guide with example input and output data, and annotated code snippets broken down by section.

References

Hariri, W. (2023). Unlocking the potential of ChatGPT: A comprehensive exploration of its applications, advantages, limitations, and future directions in natural language processing. *arXiv preprint arXiv:2304.02017*.

Lee, G. G., Latif, E., Wu, X., Liu, N., & Zhai, X. (2024). Applying large language models and chain-of-thought for automatic scoring. *Computers and Education: Artificial Intelligence*, 100213.