The Viability of Generative AI in Data Analytics and Summarization

Parsa Fares, BSE, Electrical Engineering
Mentor: Dr. Chao Wang, Associate Teaching Professor
Arizona State University

Research Question:

With the advent of Large Language Models, this project seeks to address whether modern generative AI tools such as OpenAI's ChatGPT are viable in analyzing datasets, particularly student surveys that have write-in answers, and giving an accurate readable synopsis of the survey data.

Image Credits: OpenAi ChatGPT, Google Gemini, GitHub Copilot

Methods & Power of the Prompt:

Most of the research composed of experimenting with/improving on various prompts using the Assistants API and tracking which prompts improved the accuracy of GPT-3.5 and GPT-4.0. Overall, the more background info and examples it is given, the better it performs. Much of our research also composed of repeating tests and measuring GPT's accuracy vs. traditional human data analytics (i.e., finding the mean, counting free responses)

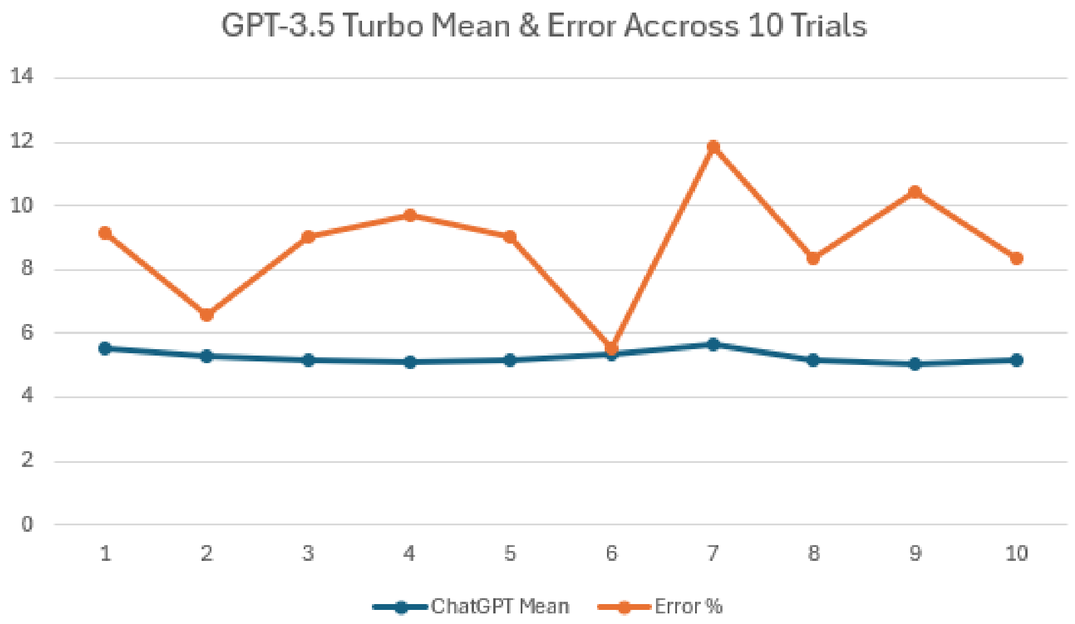
Motivation:

Through this research, new potential use-cases for generative AI in the field of Data Analytics will be explored, especially in the context of promoting teaching and learning. The instant analysis of student surveys will allow instructors to make course adjustments beneficial to students *faster than traditional data analysis tools and methods.*

Question 1: Is there any part of your experiences in the MATLAB in-class activities that sticks out in your mind?

- 6 students mentioned their enjoyment of learning about MATLAB.
- 6 students mentioned that they already knew MATLAB before the class.
- 4 students mentioned struggling to understand the MATLAB activities.

Sample output depicting ChatGPT's analysis of a free-response question



Findings & Obstacles

The major fallback of using ChatGPT to analyze large datasets was that it would crash when faced with large datasets. With smaller datasets, the different models of ChatGPT could identify the top 3 trends in student free responses but was not able to accurately count the trends. We also found that it was able to calculate the mean of numerical datasets with high accuracy. In both cases, GPT-4 had higher accuracy than GPT-3.5:

Datapoint	GPT-3.5 Turbo	GPT-4.0
Trend Identification Accuracy	62%	91%
Mean Calculation Accuracy	96.5%	100%
Price	$0.50 / 1M tokens	$30.00 / 1M tokens

Further Comments:

If LLM's can't do math, how do they find the mean so accurately? OpenAI's Code interpreter allows the model to write its own code using Python Libraries like NumPy/Pandas to complete mathematical tasks such as finding the mean of a dataset. While GPT-4.0 is more accurate, it is also *60X* more expensive!

```python
# load the CSV file
file_path = '/mnt/data/file-PTj6j4wsVPexUAf8X1oFKLfZ'
df = pd.read_csv(file_path)
# preview the first few rows in the dataframe
df.head()
# extract first character which is a number from each entry in the first column
first_column = df.columns[0]
df[first_column] = df[first_column].apply(lambda x: int(x.split(':')[0]))
# calculate the mean value
mean_value = df[first_column].mean()
mean_value
```

Sample code GPT-4.0 generated to find the mean

Acknowledgements:
The primary reference for this research was OpenAI's API and Prompt Engineering documentation:
https://platform.openai.com/docs/guides/prompt-engineering

FURI

Ira A. Fulton Schools of Engineering
Arizona State University