

An Analysis of My Interactions with ChatGPT: A Data Science Approach

Emil Gasimov

January 19, 2024

1 Introduction

Welcome to my not-average-academic-report on my interactions with ChatGPT. I present a fun and captivating exploration of the data derived from my personal conversations with the AI language model. Prepare for a ride through time stamps, message lengths, and sentiment analysis, all wrapped up with a sprinkle of data science magic.

2 Data Preparation

I consider myself quite the active ChatGPT user, with 20 MBs worth of conversation data to prove it. Initially, my computer had problems while rendering this file, but a conversion to json lines format made it behave again. The size shrank to a manageable 8 MBs when I removed out the message contents, making it easier to work with. The conversations.json file structure is quite detailed, so it is not mentioned in this report.

3 File Structure

The structure of each conversation entry is as follows:

```
{
  "title": "The title of the conversation.",
  "create_time": "The timestamp of when the conversation was created.",
  "update_time": "The timestamp of when the conversation was last updated.",
  "mapping": "A dictionary that maps message IDs to their corresponding messages.",
  "moderation_results": "An empty list for moderation results.",
  "current_node": "The ID of the current message in the conversation.",
  "plugin_ids": null,
  "conversation_id": "The ID of the conversation.",
  "conversation_template_id": null,
  "gizmo_id": null,
  "is_archived": "A boolean indicating whether the conversation is archived.",
  "safe_urls": "A list of safe URLs.",
  "id": "The ID of the conversation entry."
}
```

Each message within the mapping field has the following structure:

```
{
  "id": "The ID of the message.",
  "message": "The content of the message.",
  "author": "The role, name, and metadata of the author of the message.",
  "create_time": "The timestamp of when the message was created.",
  "update_time": "The timestamp of when the message was last updated.",
  "content": "The content type and parts of the message.",
  "status": "The status of the message.",
  "end_turn": "A boolean indicating whether the message ends the turn.",
  "weight": "The weight of the message.",
  "metadata": "Additional metadata of the message.",
}
```

```

"recipient": "The recipient of the message.",
"parent": "The ID of the parent message.",
"children": "A list of IDs of the child messages."
}

```

4 Analysis and Findings

4.1 Conversation Counts and Patterns

I've had a total of 1149 chats with ChatGPT. Peeking at the distribution of new chats over time (see Figure 1), we see peaks and valleys, coinciding with my academic calendar's breaks and flows. Thursdays seem to be a hot day for our conversations (Figure 2), likely due to assignment deadlines on weekends. My chat times (Figure 3) are quite predictable, with four spikes throughout the day that align with my daily routine, namely waking up, after classes, after dinner, and late night studies.

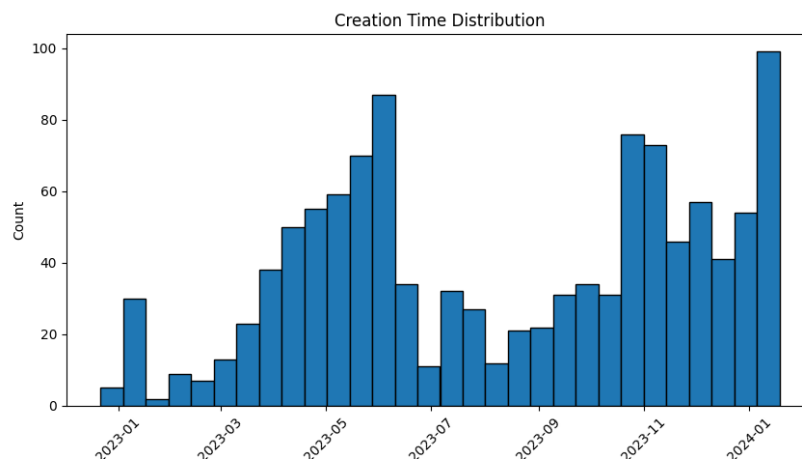


Figure 1: Distribution of New Chats Over Time

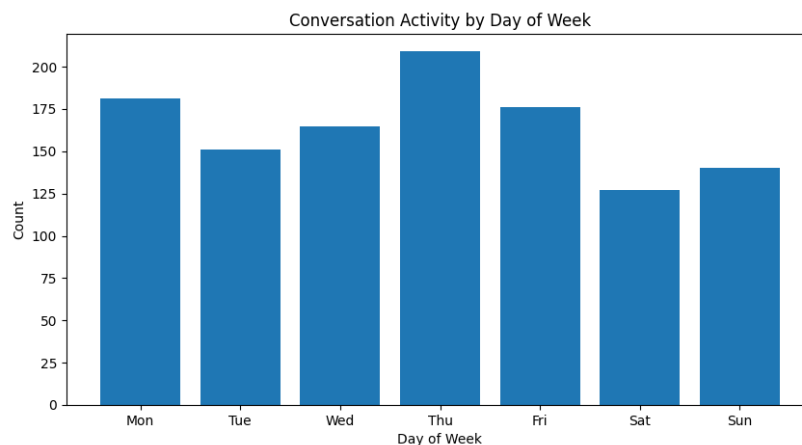


Figure 2: Activity by Day of the Week

4.2 Inference Speed and Message Distribution

The median wait time for a ChatGPT response is 90 seconds (Figure 4); thus, I've waited a total of 36 hours for all replies. Notably, the distribution of messages between us isn't quite 50/50 (Figure 5) due

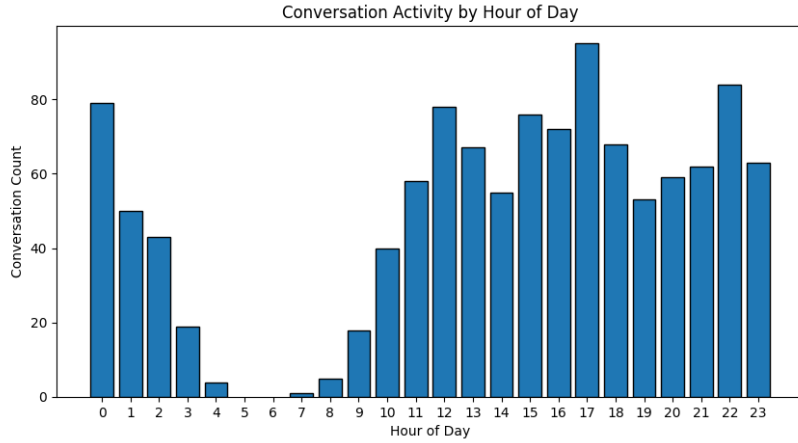


Figure 3: My Active Chat Times

to the 'regenerate' button, which created new messages from ChatGPT, while not increasing the counter of my messages, and occasional unanswered questions from server issues.

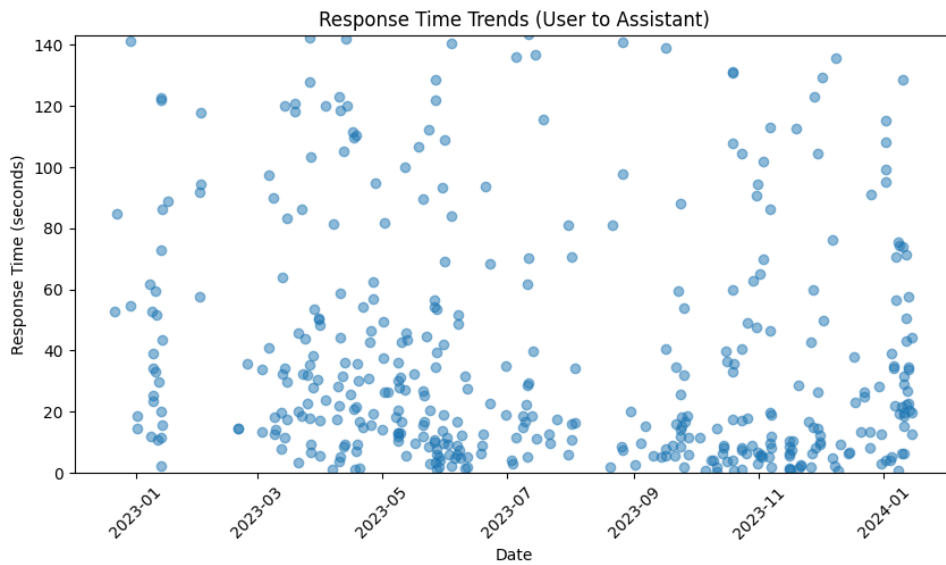


Figure 4: Median Response Time from ChatGPT

4.3 Message Lengths and Users Laziness

ChatGPT's messages are nearly twice as long as mine on average (Figure 6), and we've both increased our message lengths over time (Figure7). This contradicts to the common opinion that GPT got lazier, may be at the end it is the users that got lazier, and ChatGPT is just behaving accordingly.

4.4 Message Branching

Another useful but unpopular feature is branching of a chat. One of my chats has the record of 39 branches, meaning that there are 39 different instances of a chat. I tried to visualize this, although unsuccessfully (Figure 8).

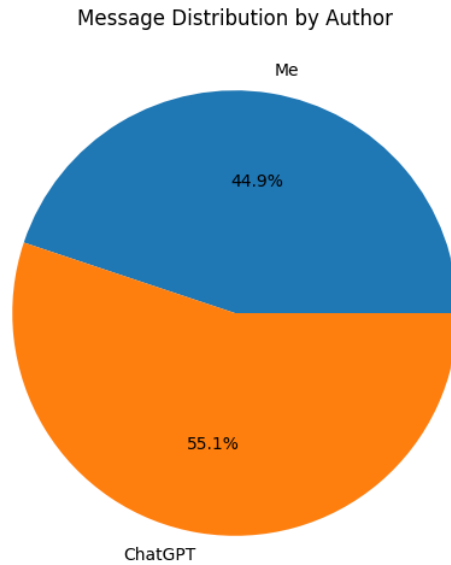


Figure 5: Distribution of Messages Between Me and ChatGPT

4.5 Content Analysis

Word clouds reveal our most common words (Figure 9), and TF-IDF analysis highlights our top keywords, showing my imperative approach versus ChatGPT’s responsive nature. ChatGPT is known for unnecessary apologies(Figure 10), but my reactions are mostly neutral instead of getting angry, as seen in the sentiment analysis of our messages (Figure 11).

4.6 Language Complexity

GRE word usage analysis shows ChatGPT’s fondness for complex vocabulary (Figure 12), however, I, in contrast to ChatGPT, use GRE words with more variety. ChatGPT, on the other hand, tends to repeat the same words more often (Figure 13).

5 Conclusion

This project has been an interesting dive into the depths of my digital dialogues with ChatGPT. I’ve uncovered many patterns of our AI-human interactions. It’s been a fun ride, and I hope this report has been as entertaining for you as it was enlightening for me.

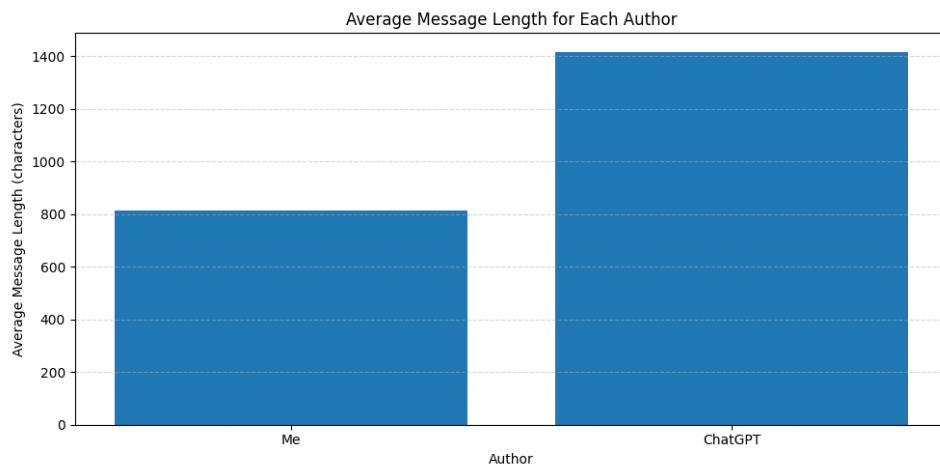


Figure 6: Average Message Length Comparison

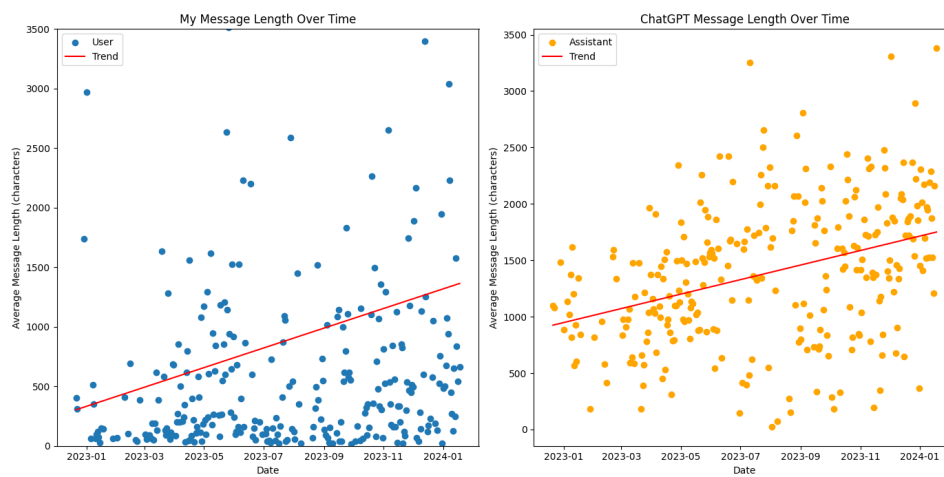


Figure 7: Average Message Length Over Time

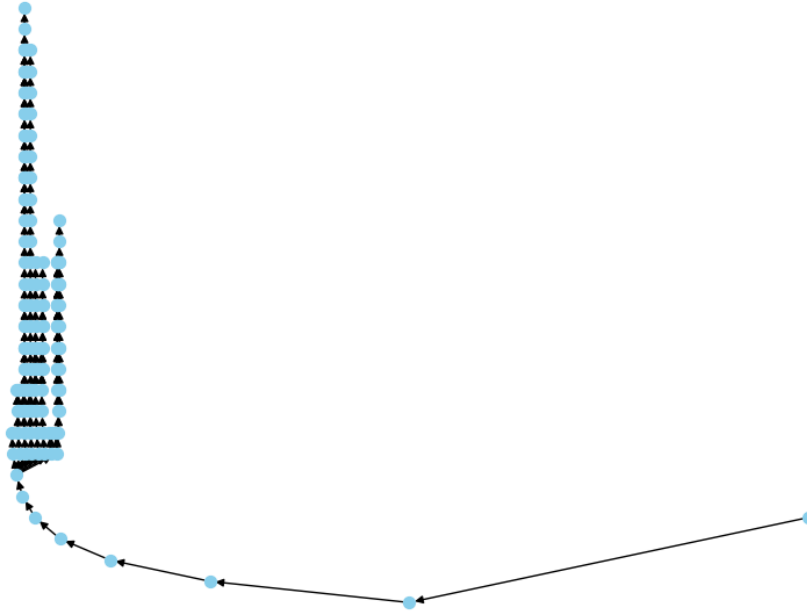


Figure 8: Visualization of a Heavily Branched Conversation



Figure 9: Word Clouds of Common Words Used by Both Authors

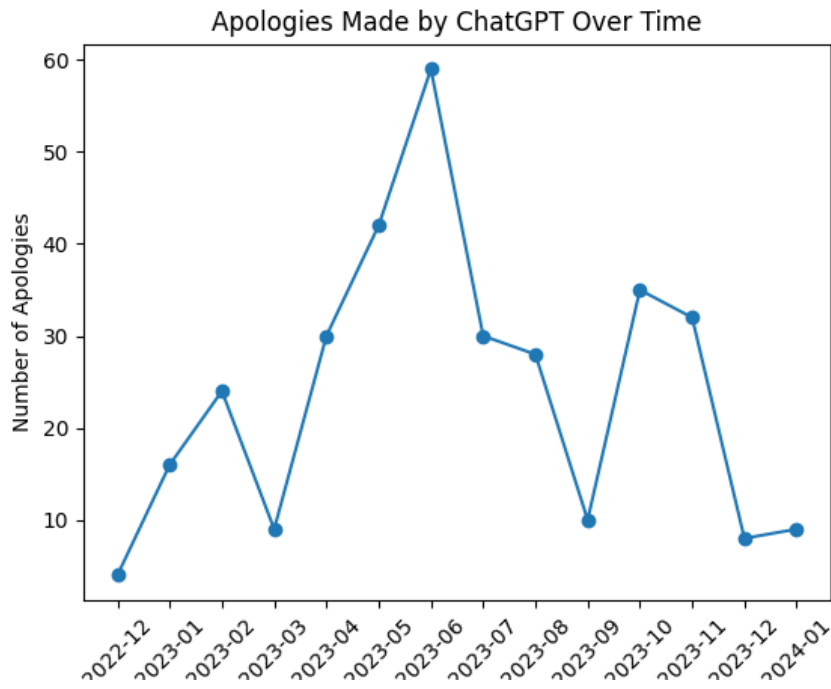


Figure 10: Frequency of ChatGPT's Apologies

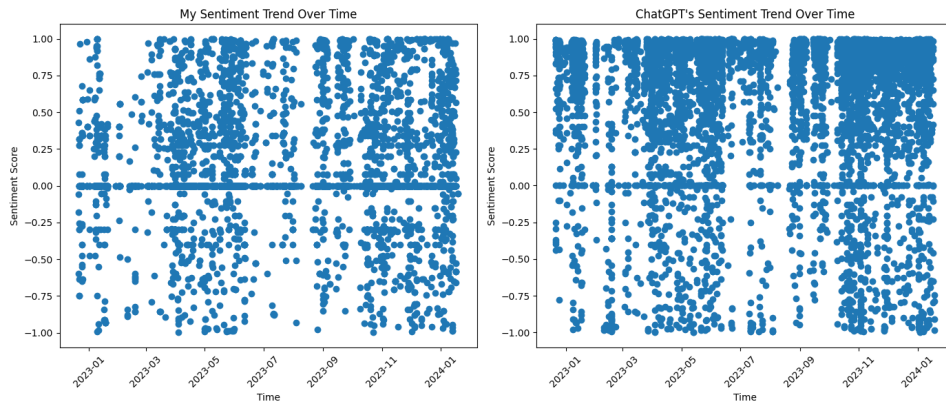


Figure 11: Sentiment Analysis of All Messages

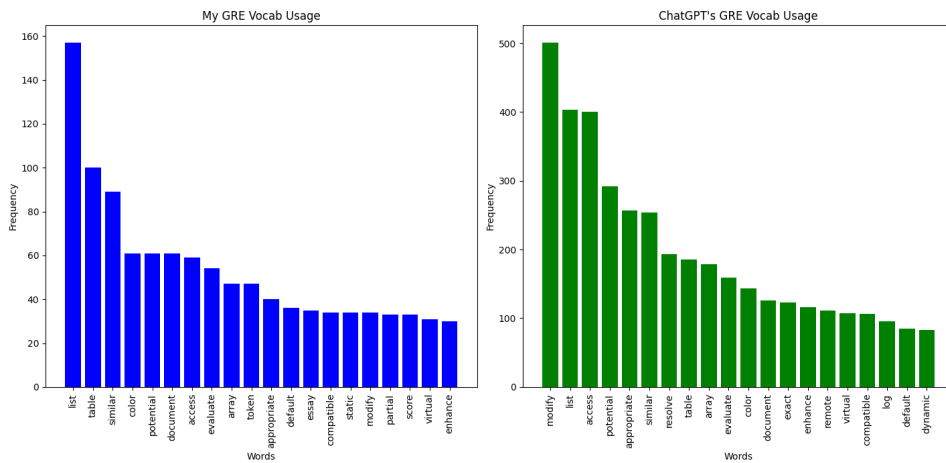


Figure 12: GRE Word Usage by Both Authors

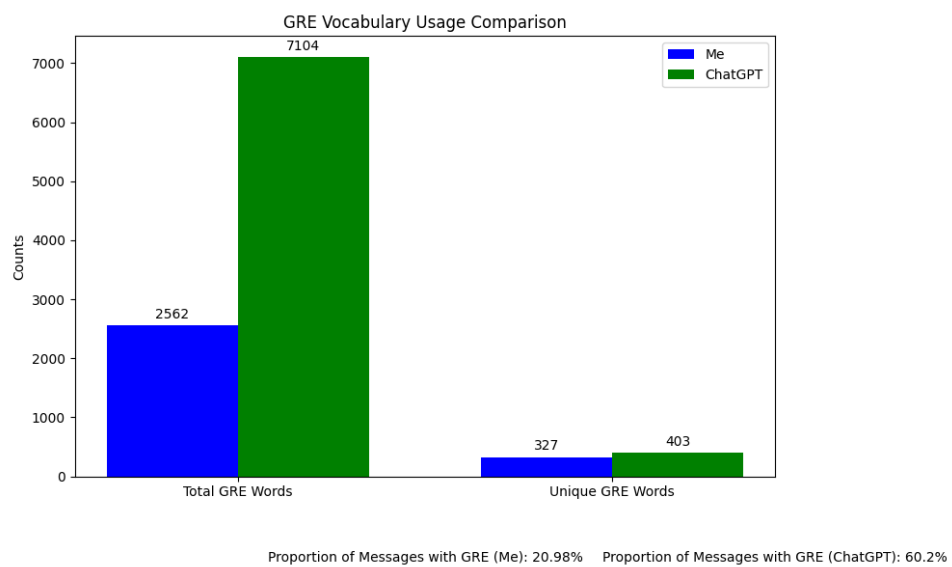


Figure 13: Unique GRE Words Used by Both Authors