Prompt Analysis : ChatGPT

Abhijeet Shah, Shahzeb Faheem, Tianyun Liu, Wonjong Kim, Yunpeng Wen

Research Question

I. Can we reliably predict whether a developer's issue will be resolved based on the initial conversation with ChatGPT?

II. If developers were to rerun their prompts with ChatGPT now and/or with different settings, would they obtain the same results?

Motivation

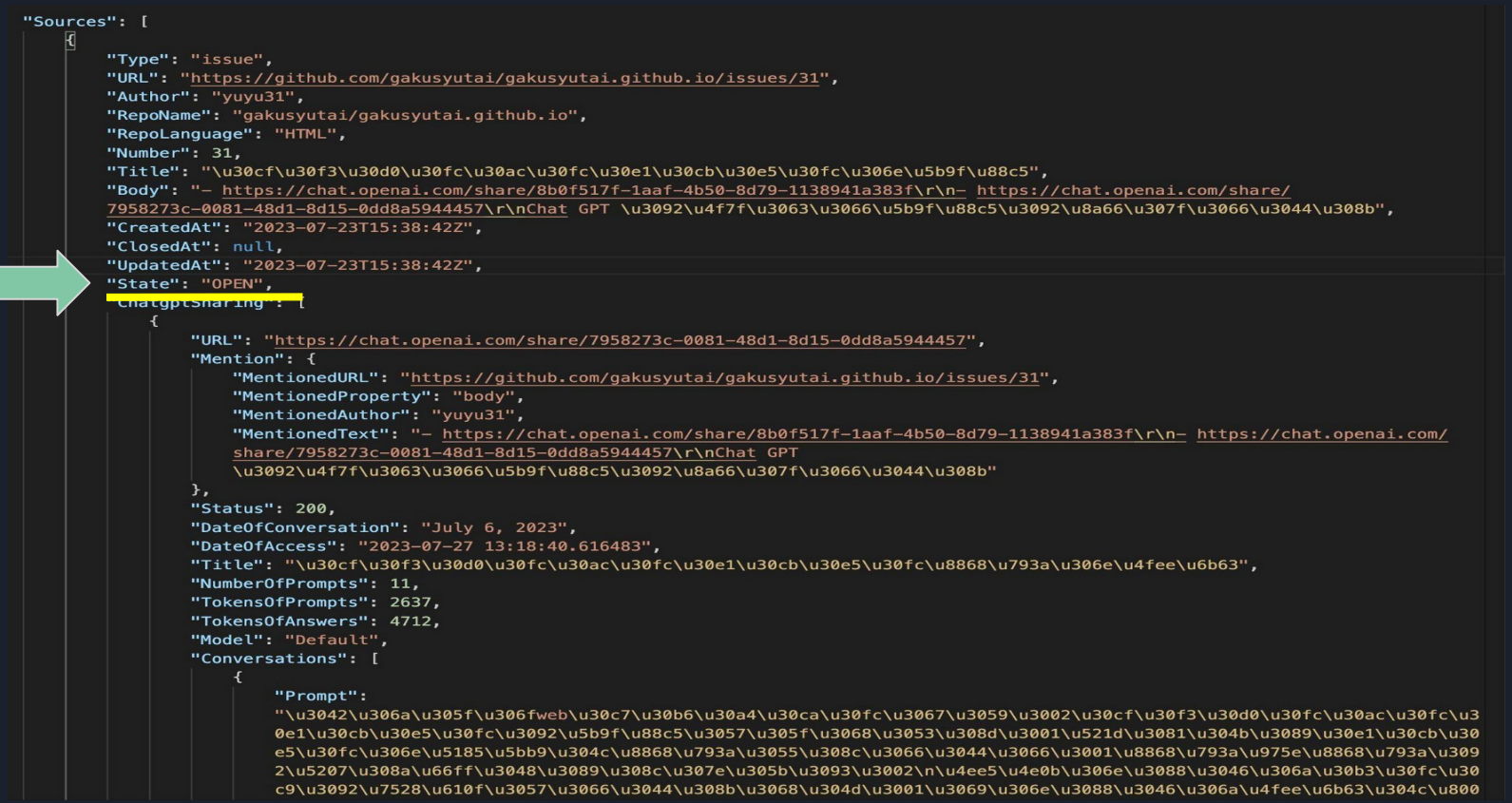
I. With the rise of ChatGPT, developers are shifting from StackOverflow
II. ChatGPT provides a real time discussion based experience as opposed to StackOverflow.
III. It is imperative to know how to manipulate ChatGPT in order to get successful answers.

What we have done so far

I. Extracted the data from the DevGPT compiled developer-centric issues (988 Conversation Threads).

II. Classified the data based on resolved and unresolved categories.

How have we classified the data?

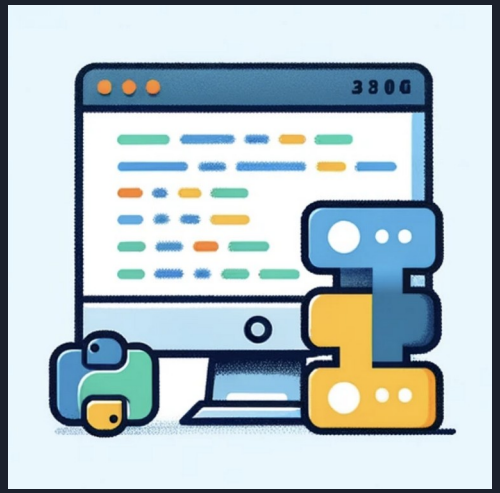


Design and Evaluation

DESIGN

I. Use of ML Classifier

- Create a Machine learning classifier with a baseline of 50% and train it on a dataset of 395 solved + 395 unsolved queries (80% of dataset) between the developer and ChatGPT.

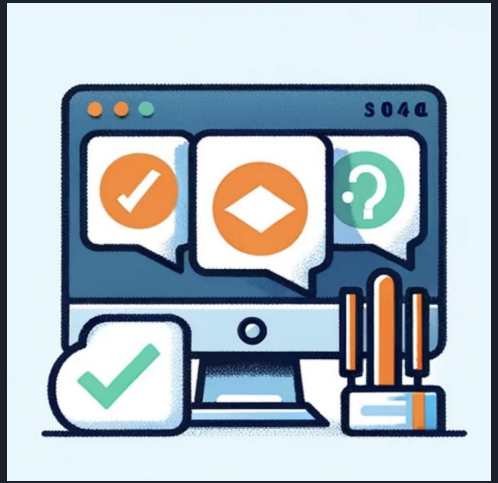


Proposed Design and Evaluation

DESIGN

II. Model Selection

- Random Forest classifier



Proposed Design and Evaluation

DESIGN

III. Feature Engineering

- Use TF-IDF vectorization for text features. It gives a score to each keyword on the number of times it appears in the text.

Proposed Design and Evaluation

DESIGN

IV. ChatGPT API

- Test the prompts from the issues dataset using ChatGPT API under various different settings like varying temperature, max tokens.

Proposed Design and Evaluation

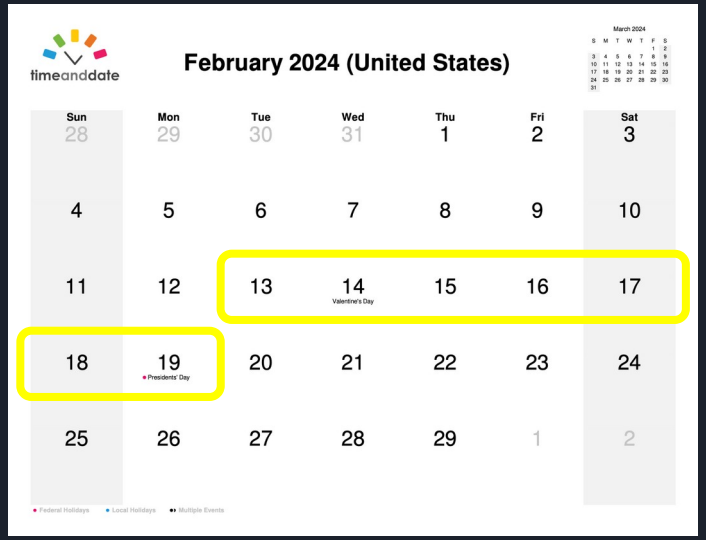
EVALUATION

I. **Metrics Evaluation**
   - Check for accuracy of prediction, precision, True Positive Rate.
   (First research question)

II. **Statistical Analysis**
   - Conducting Paired T-Test, i.e., Assess the significance of differences between responses by comparing the mean scores of paired samples calculated by cosine similarity.
   (Second research question)

Development Phase

TIMELINE

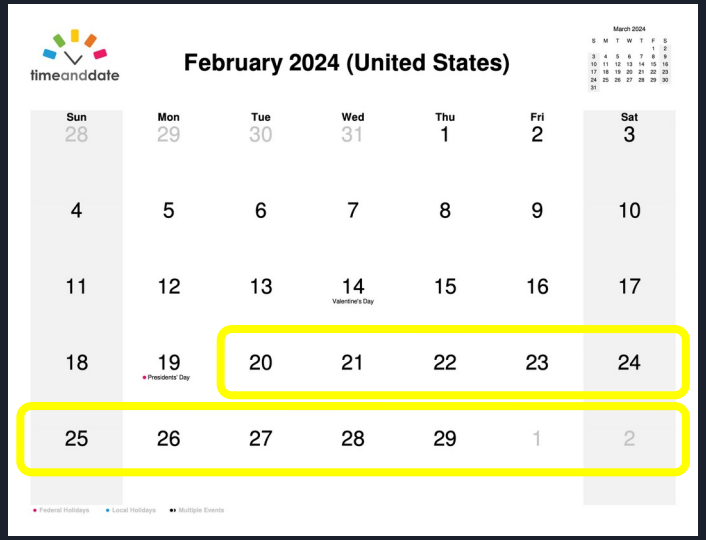


Develop ML Model

- Select the appropriate ML Model (Random Forest Classifier) and train it on a dataset of 395 solved + 395 unsolved issues conversation threads.

Evaluation Phase

TIMELINE



Testing the Trained ML Model

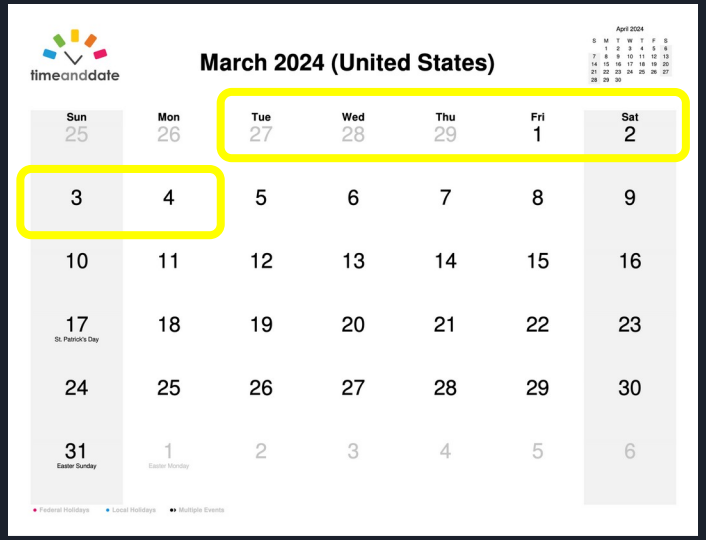
- Run the trained ML Model on the complete dataset and predict the accuracy.

ChatGPT API

- Test ChatGPT API on different settings and initiate statistical analysis (Paired t-Test) on the old vs new answers.

Testing Phase

TIMELINE

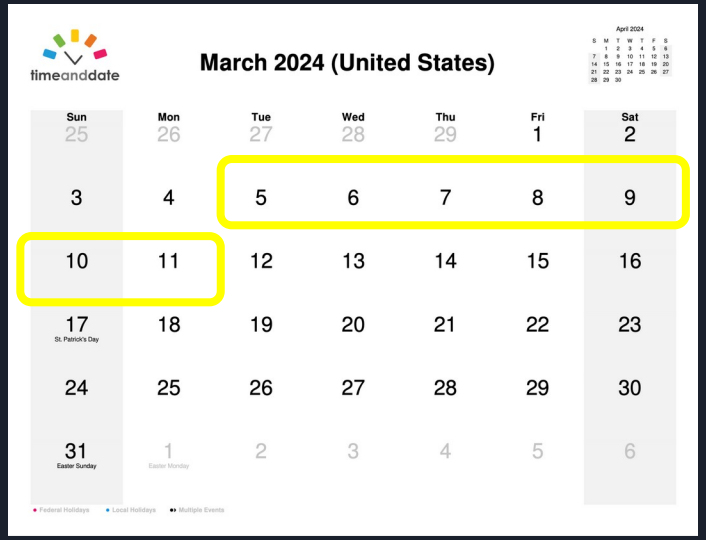


Report

- Compile findings into a comprehensive report and discuss and refine conclusions.

Conclusion

TIMELINE



Presentation Preparation

- Prepare slides for the Final Project and rehearse.

Responsibilities

Shahzeb - Lead Developer RQ2 | Presentation Lead

Wonjong - Support RQ2 | Report Lead.

Tianyun - Lead Statistical Analyst | Support Role : Report and Presentation.

Abhijeet - Data Miner | Support Developer RQ1 | Project Manager

Yunpeng - Lead Developer ML Model | Support Role : Report.

THANK YOU