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?

Problem Statement

I. To determine whether the initial conversations with ChatGPT are reliable enough to predict the resolution of a developer's issue.

II. To investigate whether the results are consistent when developers reattempt their prompts with ChatGPT with same or different conditions.

Proposed Design and Evaluation

DESIGN

I. Data Extraction

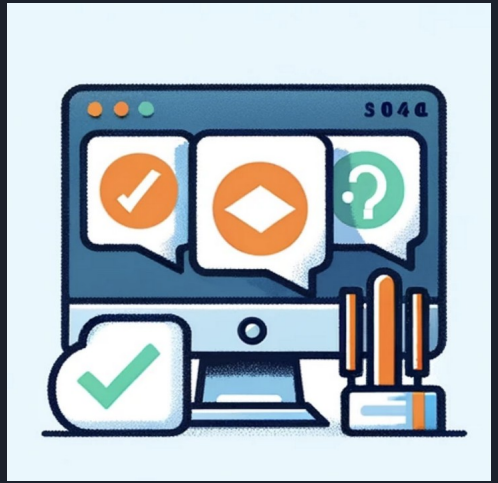
- Create a Python script to read and parse 36 JSON files from DevGPT, extracting data such as prompts, answers, and metadata. Organize this data into a CSV format.

Proposed Design and Evaluation

DESIGN

II. Data Cleaning

- Filter the dataset to focus on developer-related issues, removing duplicates and irrelevant data for consistency.



Proposed Design and Evaluation

DESIGN

III. Data Classification

- Randomly select 100 questions from 36 JSON files per person.
- Manually annotate solved and unsolved problems by team members.
- Conduct frequency analysis of keywords to identify common words in “solved” and “unsolved” categories.
- Use a Python script to separate the dataset into solved and unsolved categories

Proposed Design and Evaluation

DESIGN

IV. Data Analysis

- Use SQL queries or tools to load, clean, and analyze data, exploring patterns or trends

Proposed Design and Evaluation

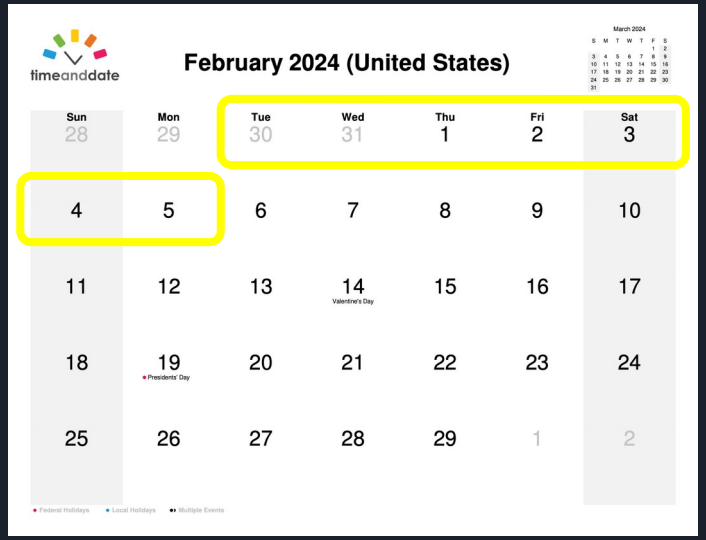
EVALUATION

I. **Statistical Analysis**
   - Conduct a detailed statistical analysis of the dataset to identify key patterns, "hot words," or "keywords," and the prompt format for consistent ChatGPT results, and analyze developer rerun outcomes.

II. **Practical Analysis**
   - Conduct extensive test runs using identified keywords and prompt formats, and apply basic statistical measures to assess the consistency or variability of these tests.

Plan

TIMELINE

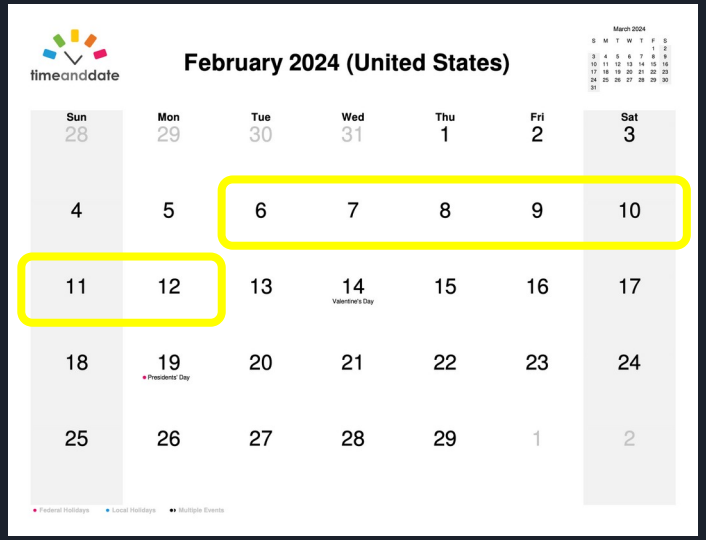


Data Preparation

- Develop and test python scripts for data extraction, organization, and cleaning.

Plan

TIMELINE

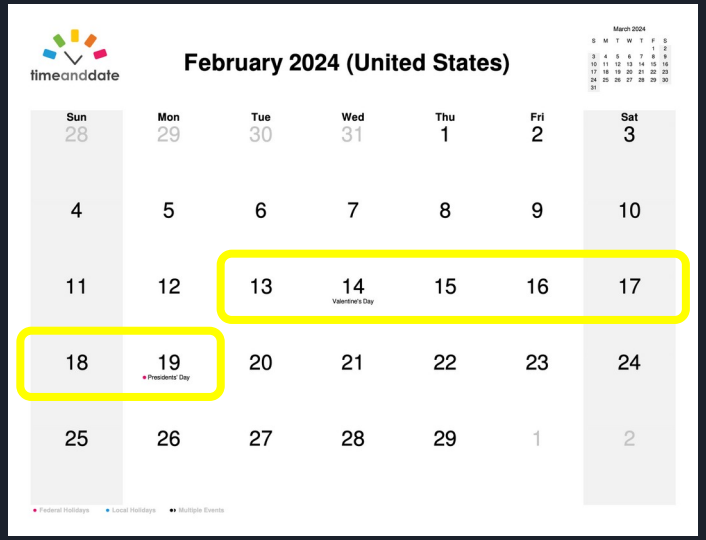


Initial Analysis

- Develop SQL queries for data analysis and explore patterns and trends in the dataset.

Plan

TIMELINE

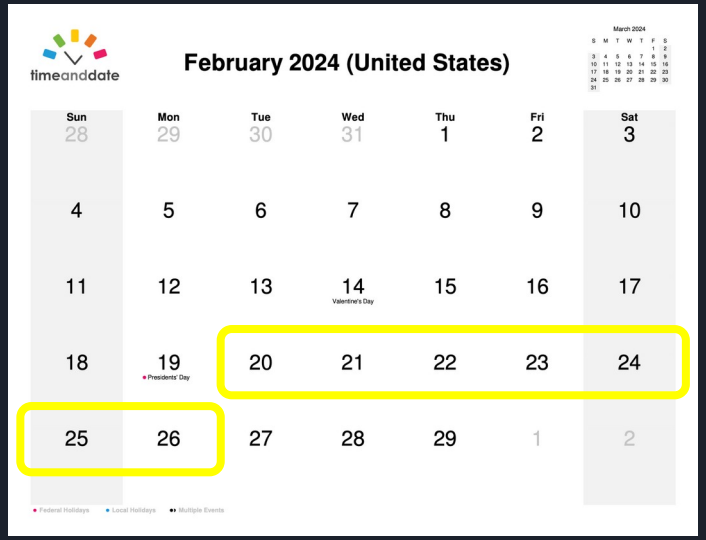


Statistical Analysis

- Conduct comprehensive analysis to identify keywords or prompt formats indicating consistent results.

Plan

TIMELINE

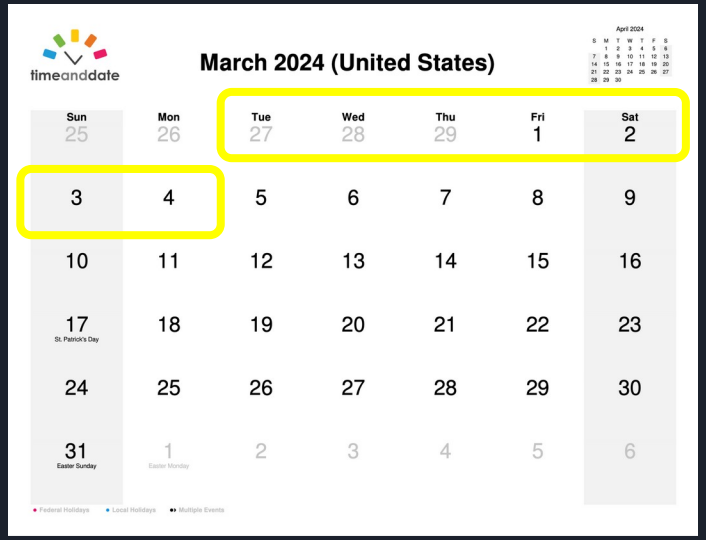


Practical Analysis and Testing

- Conduct extensive test runs using the identified keywords and prompt formats and assess the consistency and variability of the test runs using statistical measures.

Plan

TIMELINE

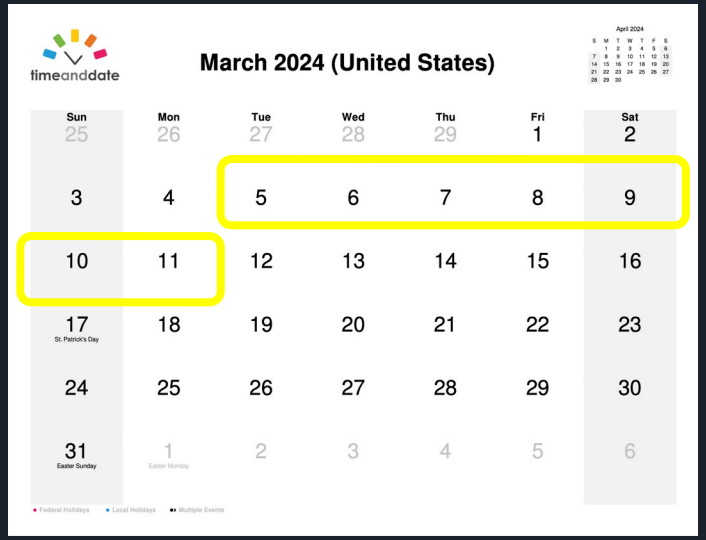


Report

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

Plan

TIMELINE



Presentation Preparation

- Prepare slides for the Final Project and rehearse.

Plan

Responsibilities

Shahzeb - Develop SQL queries and begin organizing and cleaning the dataset to identify the keywords.

Wonjong - Conduct test runs using identified keywords and prompt formats for practical analysis and testing.

Tianyun - Explore patterns and trends in the dataset and conduct the test runs using identified keywords and prompt formats.

Abhijeet - Develop and test python scripts for data extraction and conduct comprehensive analysis. Develop SQL queries for data analysis.

Yunpeng - Research and incorporate relevant external sources in to the project report and assist in gathering findings and insights from individual analysis

THANK YOU