SAP VT Global Hackathon

20-21 July 2023

**Digital Assistant – Generative AI Chat Assistant**

Problem Statement

Gathering data to complete surveys, questionnaires or even train models is extremely time consuming and the process of locating the right data source is tedious.

Use Cases

Generating dummy data for surveys can be useful in several real-world scenarios, particularly in industries where privacy, security, or data protection are major concerns. Some potential use cases include:

Software Development and Testing: During the development and testing phase of software applications that involve data collection and analysis, it is often necessary to simulate large-scale user interactions. Dummy data can be used to mimic real user responses, helping developers identify and fix any issues before deploying the application to a live environment.

Data Analytics and Visualization: Data analysts and visualization specialists often require sample datasets to demonstrate the capabilities of their tools or to create compelling visuals for presentations. Dummy data can help in generating charts, graphs, and other visuals without exposing actual sensitive information.

Training and Education: In educational settings, instructors may need to teach data analysis or survey techniques to students without accessing or sharing real-world data. Dummy datasets allow students to practice their skills in a safe and controlled environment.

Market Research: Market researchers may need to evaluate the performance of a survey platform or conduct practice surveys to refine their questionnaires before launching a real survey. Dummy data can provide a way to perform such trials without compromising the privacy of actual respondents.

Compliance Testing: In industries where regulatory compliance is crucial (e.g., healthcare, finance), organizations must ensure that their data handling practices adhere to specific standards. Dummy data can be used for testing compliance procedures without exposing real sensitive information.

Cybersecurity Training: Organizations can use dummy data to simulate cyber-attacks or security breaches, allowing their security teams to practice incident response and enhance their cybersecurity measures.

Public Demonstrations and Prototypes: When showcasing new products or applications at conferences or public demonstrations, using dummy data can protect the privacy of potential users while still demonstrating the features and capabilities of the offering.

Benchmarking and Performance Testing: In scenarios where systems need to handle large volumes of data, using dummy data for benchmarking and performance testing can help assess the system's responsiveness and scalability.

In all these cases, using dummy data ensures that sensitive or confidential information is not at risk and helps maintain compliance with data protection regulations. Industries such as healthcare, finance, government, market research, and software development are particularly concerned about data privacy and security, making them ideal candidates for the use of dummy data in surveys and other data-related activities.

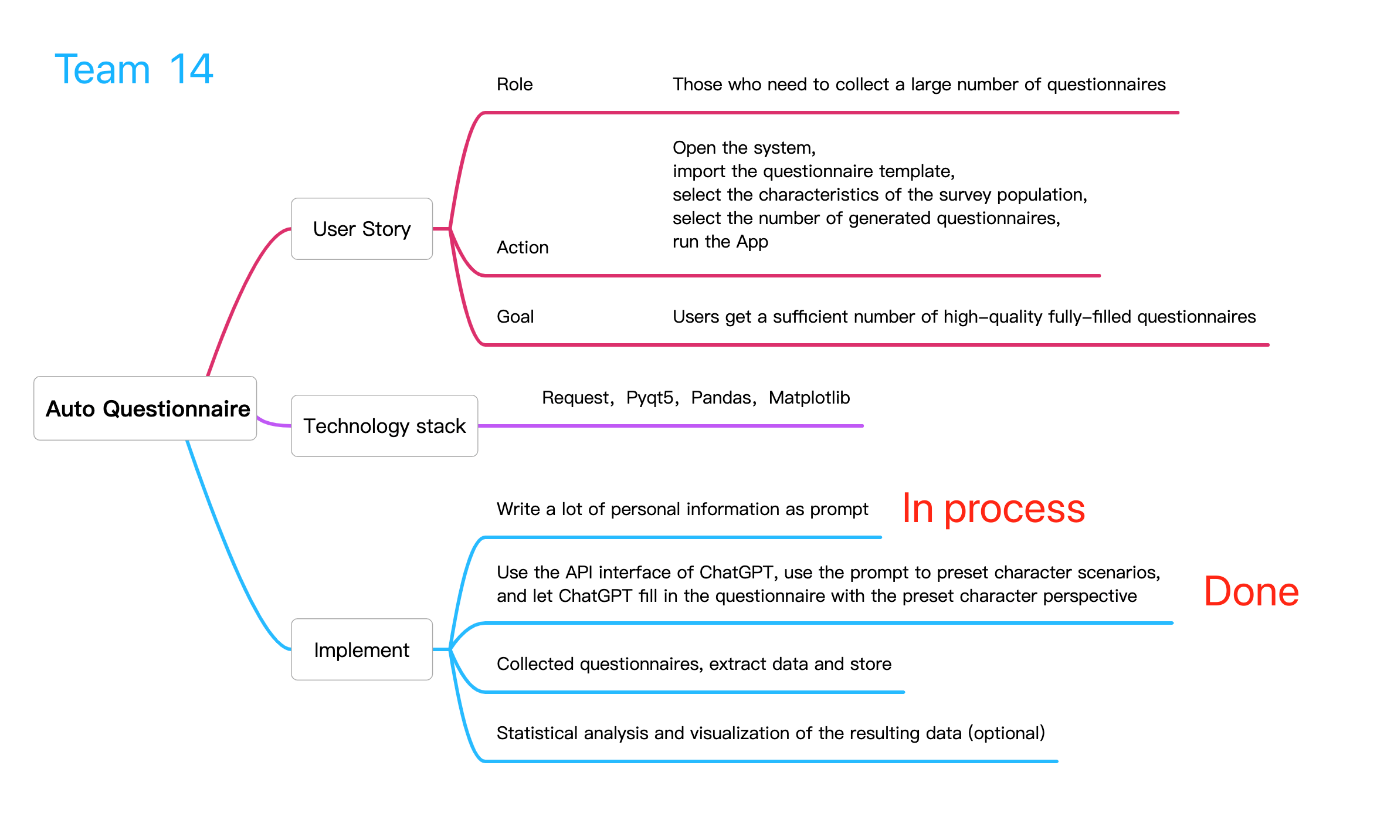
Solution Overview

ChatGPT for answers via an API

Filled questionnaire results can be statistically analysed and visualized to show generate insights

Personal information is stored and used later as prompts:

* Race: Chinese
* Background: Middle-Class
* Parents Occupation: Orchard farm owners
* Gender: Male
* Age:20 years old
* Height and Weight: 175cm and 70kg
* Educational Qualifications: Undergraduate, Computer science,
* Hobbies: Reading



Limitations

1. Quality of Responses

ChatGPT may still produce responses that are not always accurate or relevant. Some answers generated by ChatGPT could be misleading or incomplete, leading to unreliable data in the questionnaires. E.g The weight of a person is less than 30, or age is 5 years old but he has already completed graduation.

2. Bias in Responses:

ChatGPT's responses might be influenced by biases present in its training data. This could result in biased or unfair answers to certain questions, affecting the integrity of the collected information.

3. Inability to Clarify Ambiguities:

ChatGPT cannot ask clarifying questions to respondents when faced with ambiguous or unclear queries. This can lead to incorrect or irrelevant responses, impacting the reliability of the collected data.

4. Privacy Concerns:

When using ChatGPT's API to collect questionnaire information, there may be privacy and security concerns. Sensitive or personal data shared by respondents could potentially be mishandled or misused.

5. Limited Control Over Responses:

While ChatGPT's responses can be fine-tuned to some extent using prompts, there might still be limited control over the exact wording and style of the generated answers.

6. Ethical Considerations:

Automating data collection through ChatGPT's API raises ethical questions, particularly when dealing with sensitive topics or vulnerable populations. Ensuring informed consent and ethical research practices is essential.

Future Expansions (not yet designed or implemented)

* Enhanced AI Models: As AI technology advances, more sophisticated language models with improved contextual understanding and reduced biases will become available. Upgrading to these enhanced models can lead to better and more accurate questionnaire responses.
* Multilingual Support: Expanding the app's capabilities to support multiple languages can broaden its user base and make it accessible to people from diverse linguistic backgrounds.
* Real-Time Feedback: Incorporating real-time feedback mechanisms, such as sentiment analysis, can help assess the quality of responses and enable improvements to the data collection process.
* Active Learning Techniques: Implementing active learning techniques can help the AI system identify and prioritize ambiguous responses, enabling it to ask follow-up questions to gather more precise information.
* Privacy and Security Enhancements: Addressing privacy and security concerns with robust encryption, data anonymization, and compliance with data protection regulations can enhance user trust.
* Mobile Application: Developing a mobile app version of the platform can increase accessibility and convenience for users who want to collect data on the go.
* Benchmarking and Comparison: Providing benchmarking and comparison features to assess the app's performance against other data collection methods can help users understand its strengths and limitations.
* Social Impact: Exploring potential applications in fields like public health, social sciences, and market research can have a significant positive impact on understanding human behavior and societal trends.
* Integration with SuccessFactors to automate the filling of various forms and documents
  + User inputs the desired user ID (or i-number/email address) into the prompt function
  + ChatGPT will request to pull the relevant information (e.g name, department, dietary requirements, etc)
  + The target user will receive a request from ChatGPT via email to authorize the extraction of data
  + Form/Administrative item/questionnaire will be automatically populated

Conclusion