CNT 5410: Computer and Network Security

Final Project Report: ContextualAI - A Personal Assistant with Voice Recognition and Contextual Understanding

Snehit Vaddi
(Point of Contact)
vaddisnehit@ufl.edu

Sindhura Sriram
sindhura.sriram@ufl.edu

Saisanthosh Addla
saddla@ufl.edu

Sai Pujitha Reddy Chandragiri
chandragiri.s@ufl.edu

Bala Tejo Kiran Saggurthi
balatejosaggurth@ufl.edu

December 9, 2023

1 Introduction

In our endeavor to create a fluid and intuitive bridge between human communication and technology, we encounter a complex and intriguing challenge: developing a system that can accurately transcribe live audio and provide immediate, context-aware responses. This challenge is particularly demanding due to the necessity for real-time speech recognition and sophisticated natural language processing. Overcoming this challenge is vital, as it holds the potential to transform user experiences, notably in settings like live interviews where rapid and accurate responses are paramount.

To address this challenge, we propose the creation of ”ContextualAI - personal assistant with voice recognition,” an innovative system combining state-of-the-art speech-to-text technologies with advanced language models to enable dynamic transcription and response generation. This system is designed to harness the power of leading- edge technologies, such as AssemblyAI for transcription, Python's audio-to-text services for precision, OpenAI's ChatGPT for context-sensitive responses. Our objective is to blend these technologies effectively and economically, ensuring both precise transcription and coherent responses. We aim to develop ContextualAI on user-friendly platforms like Streamlit or Flask, guaranteeing straightforward user interaction and deployment. Furthermore, we intend to boost security and protect user privacy by integrating Streamlit Authenticator and applying data obscuring strategies. This approach will safeguard user data while preserving the system's effectiveness. With these technological advancements, ContextualAI is poised to fill the void between human interaction and machine comprehension, offering an unparalleled level of real-time interaction and accessibility.

2 Background & Related Work

The recent surge in artificial intelligence research, focusing on large language models (LLMs) such as ChatGPT, is highlighted in several key studies. Research [6] delves into the multilingual capabilities of ChatGPT, emphasizing the importance of evaluating LLMs in diverse linguistic environments, a crucial step towards achieving global inclusivity in AI technologies. In another study, [4] investigates the impact of ChatGPT and comparable models influence the field of public information sharing, highlighting both the advantages and risks linked with how information circulates in this essential area. This research underscores the challenges posed by the rapid dissemination of both accurate and misleading information, a phenomenon known as 'information overload.'

A broader perspective is offered in [10], which surveys the practical applications of LLMs, showcasing their transformative potential across various industries. This study illuminates the growing influence and versatility of ChatGPT and related models in practical settings. The study [3] takes an analytical approach to assess whether ChatGPT's language processing is comparable to human interaction, a core concern for the advancement of conversational AI. This inquiry is significant in understanding how these models perform in real-world interactions.

Further, [5] explores the integration of conversational AI in design processes, positioning ChatGPT as not just a tool but also as a designer and product. This innovative perspective opens new avenues for AI in creative domains, highlighting its potential as a collaborative and creative force. In a comparative analysis, [1] shows that open-source LLMs can outperform human workers in text-annotation tasks, indicating the efficiency and accuracy of these models in language processing, nearing human-level performance. The evolution of transcription strategies and their challenges are discussed in [2] and [7], emphasizing the role AI and LLMs like ChatGPT could play in transforming and enhancing transcription processes.

A comprehensive review of ChatGPT is presented in [8], covering its applications, challenges, biases, ethics, and limitations. This review is pivotal in understanding the scope and limitations of current AI technologies. The study [9] addresses the crucial aspect of privacy and data protection in AI chatbots. In an age where data breaches are frequent, this research emphasizes the need for effective strategies to safeguard user information in AI applications.

3 Approach: Dataset(s) & Technique(s)

The development of the "ContextualAI" application has been marked by careful selection and integration of diverse technologies, each contributing significantly to its overall functionality and user experience. With the recent implementation of the Streamlit Authenticator, the application has taken a substantial leap forward in terms of security and user management. The Streamlit Authenticator is a crucial component that provides robust user authentication by using Bcrypt hashing algorithm, enhancing security and offering valuable insights into user demographics and behaviors for a more personalized and secure user experience.

Streamlit, serving as the foundation for the user interface, was initially chosen for its simplicity in creating interactive web applications. Its ability to facilitate rapid development and seamless integration with Python scripts made it the optimal choice over other frameworks like Flask or Django.

The application's audio processing capability, powered by PyAudio, was selected for its straightforward API, enabling efficient handling of real-time audio streams. Although alternatives like SoundDevice and SpeechRecognition were considered for their different levels of control and user-friendliness, PyAudio's balance of functionality and ease of use made it the ideal choice. Future enhancements in this area are focused on improving audio processing accuracy, potentially incorporating advanced noise reduction techniques to ensure clarity and reduce transcription errors.

For real-time communication, the application employs websockets to establish a reliable, bidirectional communication channel with AssemblyAI for speech-to-text conversion. This technology was chosen over alternatives such as Socket.IO or server-sent events (SSE) for its superior real-time capabilities. Ongoing improvements aim to enhance the stability and error handling of these websocket connections, ensuring consistent and reliable performance.

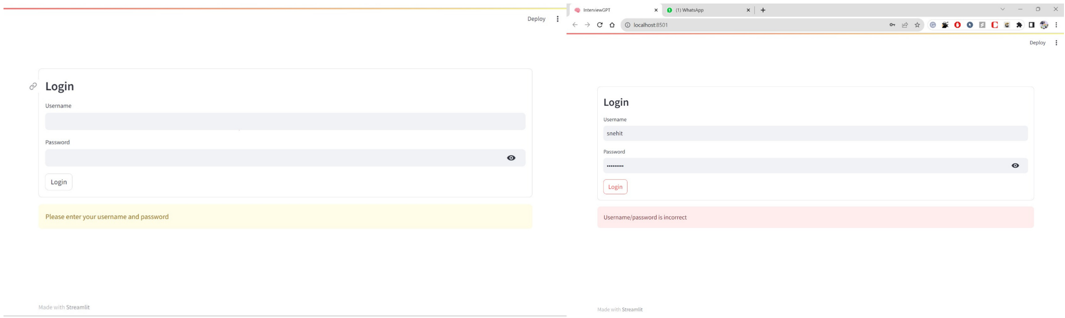
AssemblyAI was the selected service for speech-to-text conversion, prized for its accuracy and seamless integration with real-time audio streams. Competing services like Google Cloud Speech-to-Text and IBM Watson Speech to Text were also considered, but ultimately, AssemblyAI's performance in practical scenarios gave it the edge. In addition, the application integrates a form of data obscuring during speech-to-text conversion to protect user privacy. The conversation text undergoes a redaction process using spaCy's Named Entity Recognition (NER) to replace sensitive entities with "[REDACTED]", thereby enhancing privacy before storage.

As the application evolves, exploring more advanced features or switching to different services will be considered, depending on the needs for accuracy and cost-effectiveness.

The integration of OpenAI's GPT for generating responses was a strategic decision to harness its advanced natural language processing capabilities. While IBM Watson Assistant and Rasa offered different advantages, GPT's unmatched language understanding and generation abilities made it the most suitable option for our application. To maintain state-of-the-art response quality, the application will continually update with the latest GPT models and explore fine-tuning options for more tailored outputs.

4 Results

The development of "ContextualAI," a cutting-edge application, involved a series of meticulously chosen steps, each playing a crucial role in enhancing its functionality and user experience. At the forefront of its design was the integration of Streamlit, selected for its exceptional capacity to create interactive web applications effortlessly. This choice proved pivotal in crafting a user-friendly interface, significantly improving user interaction and engagement. Below is the image of the User Interface Design of the Login Page."



Login

Username

Password

Please enter your username and password

Login

Username: snehit

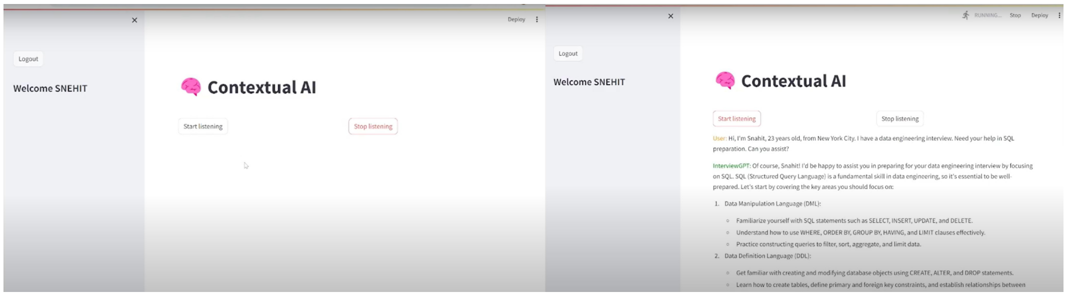
Password: ••••••••••

Username/password is incorrect

Figure 1: Web UI/UX of Login Page

An essential feature of the design was the incorporation of a robust authentication mechanism, crucial for secure and personalized user access. This addition not only reinforced the application's integrity but also strengthened user trust. The above image displays the User Interface Design of the Login Page, showcasing the implementation of the ContextualAI authentication process."

For real-time audio capture and processing, PyAudio was integrated for its proficiency, enabling the application to handle audio input efficiently, which is crucial for its speech transcription capabilities.



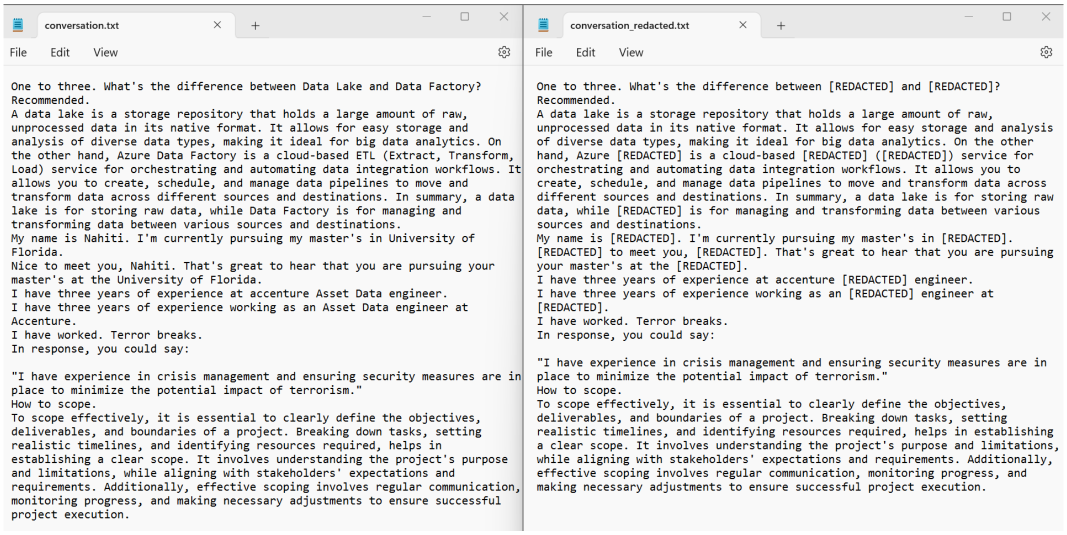
User: Hi, I'm Snahit, 23 years old, from New York City. I have a data engineering interview. Need your help in SQL preparation. Can you assist?

InterviewGPT: Of course, Snahit! I'd be happy to assist you in preparing for your data engineering interview by focusing on SQL. SQL (Structured Query Language) is a fundamental skill in data engineering, so it's essential to be well-prepared. Let's start by covering the key areas you should focus on:

1. Data Manipulation Language (DML):
   - Familiarize yourself with SQL statements such as SELECT, INSERT, UPDATE, and DELETE.
   - Understand how to use WHERE, ORDER BY, GROUP BY, HAVING, and LIMIT clauses effectively.
   - Practice constructing queries to filter, sort, aggregate, and limit data.
2. Data Definition Language (DDL):
   - Get familiar with creating and modifying database objects using CREATE, ALTER, and DROP statements.
   - Learn how to create tables, define primary and foreign key constraints, and establish relationships between

Figure 2: ContextualAI Application

In the above image, we see the main interface of the application features a personalized welcome message, ”Welcome SNEHIT”. It includes two primary interactive buttons: ”Start listening” and ”Stop listening”, integrated with PyAudio for real-time audio capture during interview simulations. The screen displays an active conversation, showcasing the application’s effective use of speech-to-text and AI response functionalities. Additionally, the interface reflects our developed methodology to handle asynchronous real-time requests and responses between the application and various APIs



```
One to three. What's the difference between Data Lake and Data Factory?
Recommended.
A data lake is a storage repository that holds a large amount of raw,
unprocessed data in its native format. It allows for easy storage and
analysis of diverse data types, making it ideal for big data analytics. On
the other hand, Azure Data Factory is a cloud-based ETL (Extract, Transform,
Load) service for orchestrating and automating data integration workflows. It
allows you to create, schedule, and manage data pipelines to move and
transform data across different sources and destinations. In summary, a data
lake is for storing raw data, while Data Factory is for managing and
transforming data between various sources and destinations.
My name is Nahiti. I'm currently pursuing my master's in University of
Florida.
Nice to meet you, Nahiti. That's great to hear that you are pursuing your
master's at the University of Florida.
I have three years of experience at accenture Asset Data engineer.
I have three years of experience working as an Asset Data engineer at
Accenture.
I have worked. Terror breaks.
In response, you could say:

"I have experience in crisis management and ensuring security measures are in
place to minimize the potential impact of terrorism."
How to scope.
To scope effectively, it is essential to clearly define the objectives,
deliverables, and boundaries of a project. Breaking down tasks, setting
realistic timelines, and identifying resources required, helps in
establishing a clear scope. It involves understanding the project's purpose
and limitations, while aligning with stakeholders' expectations and
requirements. Additionally, effective scoping involves regular communication,
monitoring progress, and making necessary adjustments to ensure successful
project execution.
```

```
One to three. What's the difference between [REDACTED] and [REDACTED]?
Recommended.
A data lake is a storage repository that holds a large amount of raw,
unprocessed data in its native format. It allows for easy storage and analysis
of diverse data types, making it ideal for big data analytics. On the other
hand, Azure [REDACTED] is a cloud-based [REDACTED] ([REDACTED]) service for
orchestrating and automating data integration workflows. It allows you to
create, schedule, and manage data pipelines to move and transform data across
different sources and destinations. In summary, a data lake is for storing raw
data, while [REDACTED] is for managing and transforming data between various
sources and destinations.
My name is [REDACTED]. I'm currently pursuing my master's in [REDACTED].
[REDACTED] to meet you, [REDACTED]. That's great to hear that you are pursuing
your master's at the [REDACTED].
I have three years of experience at accenture [REDACTED] engineer.
I have three years of experience working as an [REDACTED] engineer at
[REDACTED].
I have worked. Terror breaks.
In response, you could say:

"I have experience in crisis management and ensuring security measures are in
place to minimize the potential impact of terrorism."
How to scope.
To scope effectively, it is essential to clearly define the objectives,
deliverables, and boundaries of a project. Breaking down tasks, setting
realistic timelines, and identifying resources required, helps in establishing
a clear scope. It involves understanding the project's purpose and limitations,
while aligning with stakeholders' expectations and requirements. Additionally,
effective scoping involves regular communication, monitoring progress, and
making necessary adjustments to ensure successful project execution.
```

Figure 3: Redacted Conversation

Privacy is critical in data management, especially when handling sensitive information. Data obscuring, a key privacy mechanism in your project, addresses this need. It involves redacting or converting specific data elements into generic terms like ”[REDACTED]”, ensuring sensitive details like names and organizations are not exposed. This technique allows for the meaningful analysis of data while safeguarding individual and organizational privacy.This foundational approach paves the way for the integration of more sophisticated privacy mechanisms in future development stages, enhancing overall data security.

We evaluated speech-to-text accuracy using cosine similarity. This involved preprocessing the output to remove unnecessary labels and transforming the texts into numerical vectors through TF-IDF Vectorization. By computing the cosine similarity between the original and processed texts, we quantified their closeness, with values nearing 1 indicating high similarity. This approach demonstrated our system’s effectiveness in accurately transcribing speech, achieving approximately 92% similarity, even with accent variations.

The accuracy of ContextualAI's text redaction system was evaluated by comparing original texts with their redacted versions, focusing on the effectiveness of masking personal details such as names, dates, and locations. This method revealed a success rate of 72.22% in correctly redacting sensitive elements, providing a clear measure of the system's capability to maintain privacy in processing user data.

5 Conclusions

The "ContextualAI" is a web application designed to facilitate secure and interactive interview simulations. The Streamlit Authenticator has been implemented to enhance security by providing robust user authentication and gathering insights into user demographics and behaviors, contributing to a personalized user experience. Streamlit serves as the foundational framework for the user interface due to its simplicity and seamless integration with Python scripts.

PyAudio is employed for real-time audio processing, leveraging its straightforward API for efficient handling of audio streams. Websockets establish bidirectional communication with AssemblyAI, enabling speech-to-text conversion. AssemblyAI is chosen for its accuracy in this conversion process. The application incorporates spaCy's Named Entity Recognition (NER) to redact sensitive entities from conversation text, enhancing privacy.

OpenAI's GPT is strategically integrated for generating responses, utilizing its advanced natural language processing capabilities. The application's design is forward-looking, committing to regular updates with the latest GPT models and exploring fine-tuning options for more tailored responses. So, "ContextualAI" employs a combination of technologies for user authentication, real-time audio processing, speech-to-text conversion, privacy enhancement, and advanced natural language understanding to create an effective interview simulation platform.

References

[1] Meysam Alizadeh et al. Open-source large language models outperform crowd workers and approach chatgpt in text-annotation tasks. *arXiv:2307.02179 [cs.CL]*, 2023.

[2] V. Azevedo, M. Carvalho, F. Fernandes-Costa, S. Mesquita, J. Soares, F. Teixeira, and Â. Maia. Interview transcription: Conceptual issues, practical guidelines and challenges. *Revista de Enfermagem Referência*, September 2017.

[3] Zhenguang G. Cai et al. Does chatgpt resemble humans in language use? *arXiv:2303.08014 [cs.CL]*, 2023.

[4] Luigi De Angelis et al. Chatgpt and the rise of large language models: the new ai-driven infodemic threat in public health. *Frontiers in Public Health*, 11, 2023.

[5] A. Baki Kocaballi. Conversational ai-powered design: Chatgpt as designer, user, and product. *arXiv:2302.07406 [cs.HC]*, 2023.

[6] Viet Dac Lai et al. Chatgpt beyond english: Towards a comprehensive evaluation of large language models in multilingual learning. *arXiv:2304.05613 [cs.CL]*, 2023.

[7] S. Point and Y. Baruch. (re)thinking transcription strategies: Current challenges and future research directions. *Scandinavian Journal of Management*, 39:101272, 2023.

[8] P. P. Ray. Chatgpt: A comprehensive review on background applications key challenges bias ethics limitations and future scope. *Internet of Things and Cyber-Physical Systems*, 3:121–154, 2023.

[9] Glorin Sebastian. Privacy and data protection in chatgpt and other ai chatbots: Strategies for securing user information. *International Journal of Security and Privacy in Pervasive Computing*, 15(1), 2023.

[10] Jingfeng Yang et al. Harnessing the power of llms in practice: A survey on chatgpt and beyond. *arXiv:2304.13712 [cs.CL]*, 2023.