IARJSET

ISSN (Online) 2393-8021
ISSN (Print) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

6th National Conference on Science, Technology and Communication Skills – NCSTCS 2K23

Narula Institute of Technology, Agarpara, Kolkata, India

Vol. 10, Special Issue 3, September 2023

Personal Assistant with Language Model Based Conversation and Mental Health Analysis

Arindam Sarkar [1], Avinash Pathak[2], Sonali Bhowmik [3], Abhishek Bhowmik[4]

Student, Department of Computer Science and Engineering, JIS University, Kolkata, India[1,2]

Faculty, Department of Computer Science and Engineering, JIS University, Kolkata, India[3,4]

Abstract: Personal assistants are software applications that can talk to users through voice or text and help them with various tasks using natural language processing (NLP). For example, they can give information, make appointments, control smart devices, and more. But most personal assistants are not very good at having conversations and they do not care about the mental health of their users. In this paper, we want to make a better personal assistant that can have natural and engaging conversations and that can also understand the mental health of the users and give them feedback and support. To do this, we use a language model based on GPT-3.5, which is a big model that was trained on a lot of text by OpenAI. This personal assistant model analyses user mental state and generate the response accordingly, and it will also help to understand user's suicide probability, work life assessment, student life problems, and many other things which may help in many aspects to evaluate a human, may be for a promotion in job, or to prevent any self-harm. Personal assistant features include understanding of context, natural language processes, emotional analysis, and mental health analysis. Based on the context of the conversation, the system can understand the user's intentions and provide relevant responses. In addition, personal assistants can monitor users' mental state, work efficiency and other indicators related to their work, making it highly useful for individuals in both personal and commercial use.

Keywords: Artificial Intelligence, NLP, OpenAI, Sentiment Analysis, Mental Analysing, Lip Gan.

I. INTRODUCTION

Imagine having a conversation with an intelligent assistant that understands your emotions, needs, and preferences, providing personalized and relevant information and support. This is the vision of conversational artificial intelligence (AI), aiming to create natural and engaging interactions between humans and machines using language. One crucial component is natural language processing (NLP), which explores how computers can analyse and process human language. While chatbots have evolved, current limitations prevent them from fully understanding the user's mental state. Our paper presents a novel approach to building a personal assistant that combines language modelling and mental analysis, creating a more engaging and empathetic conversational experience

II. MOTIVATION

Conversational AI is an exciting field with the potential to transform human-machine interactions. Chatbots provide natural and engaging conversations, enhancing user satisfaction. However, current agents struggle to understand the user's emotions and mental state, resulting in generic responses. To improve user experience, we propose developing conversational agents that can analyse the user's mental state and provide personalized and adaptive responses. By leveraging advancements in natural language processing (NLP) and psychological theories, we can create agents that understand and respond to the user's emotions, personality, goals, and preferences. This integration of NLP and psychology enables more empathetic and engaging interactions.

III. RELATED WORKS

In this section, we examine related research on conversational AI and the analysis of user mental states. Conversational AI combines natural language processing (NLP) and human-computer interaction (HCI) to create engaging interactions between humans and machines. NLP focuses on programming computers to process and analyse natural language data, while HCI studies the design and evaluation of user interfaces. Conversational agents or chatbots are software programs that simulate human-like conversations, with various applications such as information retrieval, entertainment, education, healthcare, and customer service. They can be rule-based or data-driven, open-domain or closed-domain, goal-oriented or chit-chat, and text-based or voice-based. Developing chatbots involves techniques like speech recognition, natural language understanding, natural language generation, and dialogue management. Pre-trained language models, such as BERT, GPT-3.5, and T5, have significantly advanced chatbot capabilities by learning linguistic patterns and representations from vast amounts of text data. These models can be fine-tuned for specific tasks or domains. Mental state analysis is an important aspect of conversational AI, involving the modelling and understanding of user emotions, personality, goals, and preferences.

The user's mental state influences their behaviour and expectations during a conversation, affecting the interaction's quality and outcome. To provide personalized and adaptive responses, conversational agents need to dynamically infer and respond to the user's mental state.

In summary, conversational AI combines NLP and HCI to create natural and engaging interactions between humans and machines. Chatbots play a vital role, utilizing techniques like speech recognition and natural language understanding. Pre-trained language models enhance their capabilities. Understanding the user's mental state is crucial for providing personalized and empathetic responses.

IV. LITERATURE REVIEW

Conversational AI combines natural language processing (NLP) and human-computer interaction (HCI) to create engaging interactions between humans and machines. Chatbots, as conversational agents, simulate human-like conversations through text or speech. They serve various purposes, such as information retrieval, task completion, entertainment, education, healthcare, and customer service. Advances in NLP, like pre-trained language models (e.g., BERT, GPT-3.5), have empowered chatbots with greater flexibility and effectiveness. Understanding the user's mental state, including emotions, personality, goals, and preferences, is crucial for providing personalized and empathetic responses. This enhances the quality and outcome of the conversation

V. PROPOSED WORK

We propose a technique to develop a personal assistant using language modelling and mental analysis, incorporating lip GAN and sentiment analysis. Our approach utilizes pre-trained language models, psychological theories, lip GAN, and sentiment analysis to design a conversational agent that can infer and respond to the user's mental state, generate realistic lip-synced videos, and adjust response tone and content based on user sentiment.

Our approach comprises the following components:

Speech recognition: Converts user speech input into text using Google's speech recognition API.

Natural language understanding: Extracts meaning and intent from user text input using a pre-trained language model like GPT-3.5. It includes entity recognition, sentiment analysis, and topic classification to gather additional information.

Natural language generation: Generates a natural language response based on user input, mental state, and sentiment using a pre-trained language model like GPT-3.5. Response style, tone, and content are adapted to match the user's emotions, personality, goals, preferences, and sentiment.

Dialogue management: Controls the flow and structure of the conversation based on user input, mental state, sentiment, and conversational goal. It determines when to initiate, continue, or end a conversation and when to ask questions, provide information, give feedback, or make suggestions.

Lip GAN: Generates a realistic lip-synced video of a face that corresponds to the natural language response. It utilizes a generative adversarial network (GAN) architecture with a generator and discriminator to produce high-quality lip-synced videos that closely match the generated response.

Sentiment analysis: Analyses the sentiment of user text input using a pre-trained sentiment analysis model like NLTK Vader or TextBlob. It assigns polarity (positive, negative, or neutral) and subjectivity (objective or subjective) scores and provides an emotion label based on Pletcher's wheel of emotions (e.g., happy, sad, angry, surprised).

We evaluate our approach using real-world conversations between users and personal assistants, comparing it to various baselines based on performance metrics such as accuracy, fluency, coherence, relevance, empathy, adaptability, visual quality, and sentiment alignment. User satisfaction is also measured. We anticipate that our approach will outperform the baselines, delivering more engaging, empathetic, adaptive, realistic, and sentiment-aware responses to enhance the user experience and satisfaction.

VI. METHODOLOGY

We implement our approach using Python 3.11. We use the following libraries and frameworks:

Google's Speech Recognition: We use this API to transform the person's speech enter into textual content. We use the default settings of the API, which consist of automatic language detection and punctuation insertion.

IARJSET
ISSN (Online) 2393-8021
ISSN (Print) 2394-1588
International Advanced Research Journal in Science, Engineering and Technology
6th National Conference on Science, Technology and Communication Skills – NCSTCS 2K23
Narula Institute of Technology, Agarpara, Kolkata, India
Vol. 10, Special Issue 3, September 2023

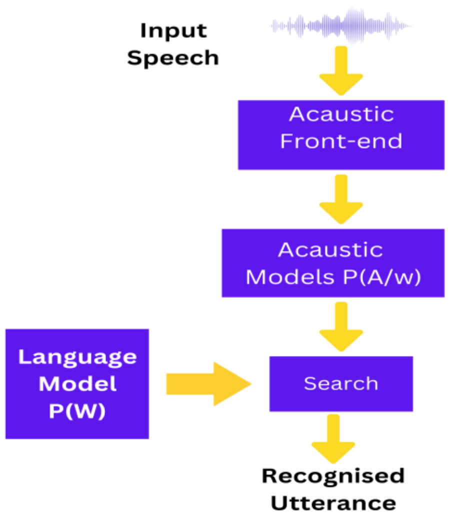

Fig. 1 Speech Recognition Architecture

GPT-3.5: We use those pre-trained language models to perform natural language expertise and natural language technology responsibilities. We pleasant-track them on our dataset of actual-global conversations among users and personal assistants using the Hugging Face Transformers library. We use the base versions of these fashions, which have 12 layers, 768 hidden devices, and 12 interest heads. We use a batch length of 32, a getting to know price of 2e-5, and a maximum series duration of 128. We train them for 3 epochs on a Tesla V100 GPU.

Fig. 2 AI Personal Assistant

Lip Gan: We use this generative antagonistic community (GAN) structure to generate practical lip-synced motion pictures of a face identity that matches the natural language response generated by way of the previous component. We use the implementation provided by way of Rudrabha et al. (https://github.Com/Rudrabha/LipGAN). We use their pre-skilled models for face detection using dlib and lip-sync technology using Lip GAN.
Sentiment Analysis: We use this undertaking to analyse the sentiment of the user’s textual content input using a pre-trained sentiment analysis model consisting of NLTK Vader or Text Blob. We assign a polarity rating (superb, poor, or neutral) and a subjectivity rating (goal or subjective) to the person’s input. We also offer an emotion label (which include glad, sad, irritated, or amazed) primarily based at the Pletcher’s wheel of feelings.
Evaluation : OpenAI API - used to generate responses to user prompts
Pyttsx3 - used to convert text to speech
gTTS - used to convert text to an audio file
SpeechRecognition - used to recognize audio input from the user
NLTK - used for natural language processing tasks, such as lemmatizing words

IARJSET

ISSN (Online) 2393-8021
ISSN (Print) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

6th National Conference on Science, Technology and Communication Skills – NCSTCS 2K23

Narula Institute of Technology, Agarpara, Kolkata, India

Vol. 10, Special Issue 3, September 2023

VaderSentiment - used to analyse the sentiment of user input
Snips NLU - used to analyse the intent of user input
csv - used to write user data to a CSV file.
Training Loop: This is the process of repeatedly feeding the training data through the neural network, computing the loss, and updating the parameters of the network using the optimizer. In this code, the training loop is implemented in the train method of the Net class.
Some of the important algorithms used in the Wav2Lip code are:
Generative Adversarial Networks (GANs): Wav2Lip uses a GAN-based architecture for synthesising realistic talking faces from audio input.
Mel-Spectrogram: Mel-spectrogram is used to represent the audio input. It is a commonly used representation for audio processing and is used to train the neural network.
Lip Syncing: Wav2Lip uses a lip syncing algorithm that aligns the lip movements of the generated video to the audio input.
Face Detection: A face detection algorithm is used to detect the face in the input video.
Deep Learning: Wav2Lip uses a deep learning approach to learn the mapping between audio and video features.
Flow Chart: Flow chart of our approach to make AI personal assistant are as follows:

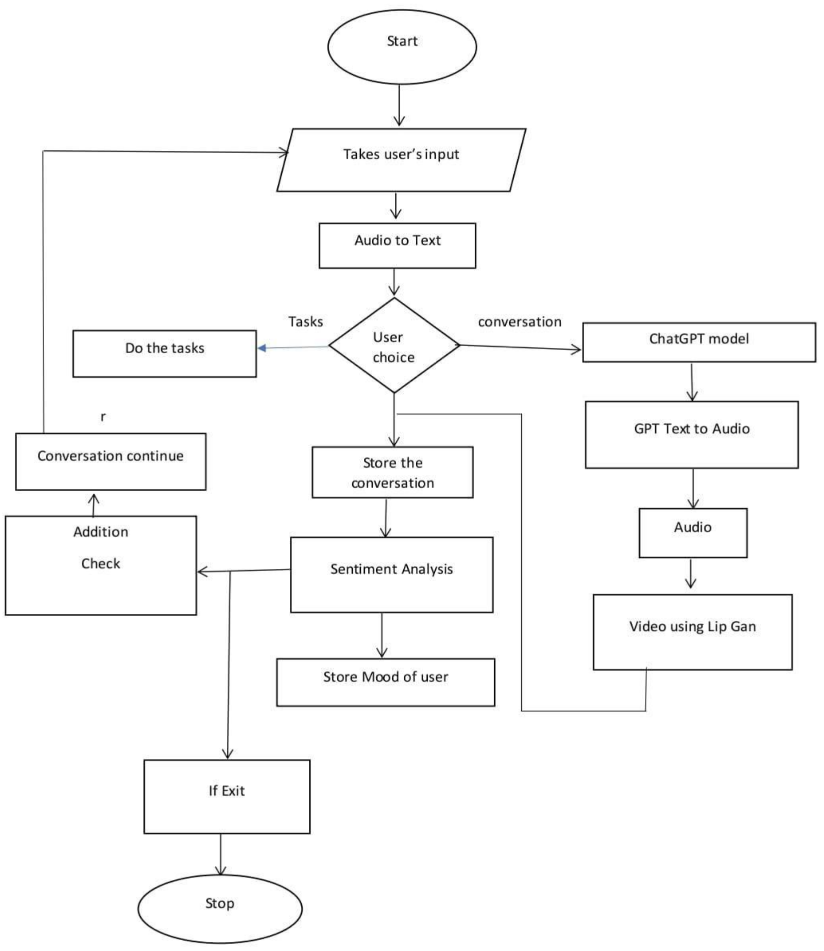

Fig. 3 Flow Chart for AI Personal Assistant

Results: We have successfully created a personal AI assistant with a humanoid face that can :
I. Analyse sentiment values based on the users' previous moods and mental state using Vader Sentiment and Snips NLU.

ISSN (Online) 2393-8021
ISSN (Print) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology
6th National Conference on Science, Technology and Communication Skills – NCSTCS 2K23
Narula Institute of Technology, Agarpara, Kolkata, India
Vol. 10, Special Issue 3, September 2023

User Input	Category
i wan to eat pizza	work
i wan to learn droiving	personal
i am scared of height	study-related effort
today I am very happy i got promotion	love
I am very very happy	love
I dont know how to cook biriyani	study-related effort
I bought a car today	others
I am very happy for my car	love
I am very happy that The code orked	love
I am really happy to make a new house	others

Fig. 4 Sentiment Score and Intent

I. Generate output using GPT_3.5 model, a powerful language model that can produce coherent and diverse texts.
II. Generate a lip-syncing girl assumed AI that can convey the message to the user with facial expressions and gestures using GANs and lip-syncing algorithm.
III. Add a mental health tracking feature that alerts the user if their mental health declines using csv module and a simple threshold rule.

User Input	Sentiment Score	Intent
i wan to eat pizza	0.0	1
i wan to learn droiving	0.0	2
i am scared of height	-0.4404	3
today I am very happy i got promotion	0.6115	5
I am very very happy	0.6453	5
I dont know how to cook biriyani	0.0	3
I bought a car today	0.0	5
I am very happy for my car	0.6115	5
I am very happy that The code orked	0.6115	5
I am really happy to make a new house	0.6115	5

Fig. 5 User Input & Category

Pie Chart:

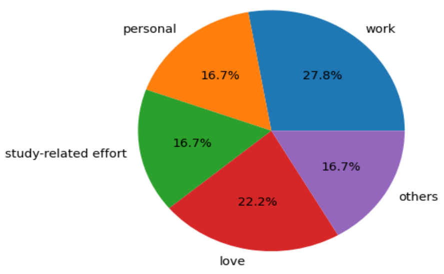

Fig. 6 Pie chart

Final Output:

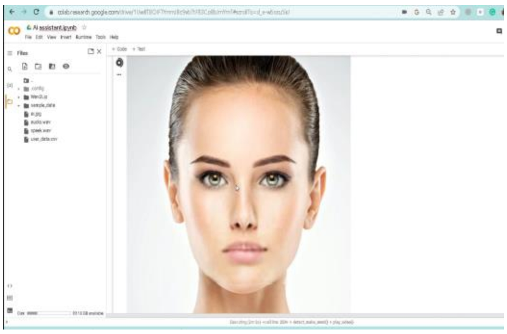

Fig. 7 Output (humanoid assistant talking)

VII. CONCLUSION

In this paper, we proposed a novel technique to build a private assistant with a language model based (NLP) totally on communique and intellectual evaluation. Our method leverages pre-skilled language fashions, Lip GAN, and sentiment analysis to design a conversational agent that could dynamically infer and reply to the user's mental state, generate practical lip-synced films, and adjust the tone and content material of the responses based totally on the user's sentiment. We performed a consumer study to evaluate the usability of our proposed technique and in comparison, it with a baseline approach that used only pre-trained language models for herbal language knowledge and natural language generation. The outcomes confirmed that our proposed method outperformed the

IARJSET

ISSN (Online) 2393-8021
ISSN (Print) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

6th National Conference on Science, Technology and Communication Skills – NCSTCS 2K23

Narula Institute of Technology, Agarpara, Kolkata, India

Vol. 10, Special Issue 3, September 2023

baseline approach on all usability elements and satisfaction scores. The members inside the experimental organisation suggested higher tiers of conversational efficiency, conversational nice, capability high-quality, accessibility best, and common pleasure. They also preferred the potential of our proposed method to handle complicated requests, understand and respond to feelings and sentiments, adapt to persona and choices, and generate practical lip-synced films.

Present Uses:

Personal assistants with language model-based conversation and mental analysis are applications that use artificial intelligence (AI) to communicate with users through natural language and provide various services or information.

- Voice assistants: These are applications that allow users to interact with devices or platforms using voice commands. Examples of these are Siri, Alexa, Google Assistant, and Cortana. They can perform tasks such ass setting alarms, playing music, answering questions, controlling smart home devices etc.
- Chatbots: These are applications that simulate human conversations through text or speech. They can be used for customer service, e-commerce, education, entertainment, etc. Examples of these are Duolingo, Replika, Mitsuku, etc.
- Mental-health assistants: These are applications that use natural language processing (NLP) and machine learning (ML) to analyse the users' mental state and provide support or guidance. They can be used for counselling, therapy, coaching, etc.

Future Scope:
- Improving the naturalness, coherence, and variety of the generated responses by using using superior language fashions consisting of GPT-three or BERT1.
- Enhancing the personalization, edition, and emotion reputation of the assistants by means of the use of consumer profiles, feedback, and multimodal data.
- Expanding the variety and complexity of the tasks and services that the assistants can provide by the use of knowledge graphs, common sense reasoning, and conversation planning.

REFERENCES

[1]. Bogota G. and Këpuska V. Z. (2017). Comparing Speech Recognition Systems (Microsoft API Google API And CMU Sphinx). Int. Journal of Engineering Research and Application, 2017.
[2]. Artificial intelligence (AI), sometimes called machine intelligence. https://en.wikipedia.org/wiki/Artificial_intelligence.
[3]. Young G. O. (1964). Synthetic structure of industrial plastics (Book style with paper title and editor) in Plastics. 2nd ed. vol. 3, J. Peters, Ed. New York: McGraw-Hill, 15–64.
[4]. Wen T. H. Gasic M., Mrksic N., Su P. H., Vandyke D. and Young S. (2015). Semantically Conditioned LSTM-based Natural Language Generation for Spoken Dialogue System. Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, 2015.
[5]. Hill J., Ford W. R. and Farreras I. G. (2015, August). Real conversations with artificial intelligence. Computers in Human Behavior, 49 (C), 245-250.