SvedbergOpen
DISSEMINATION OF KNOWLEDGE

International Journal of Artificial Intelligence and Machine Learning

Publisher's Home Page: https://www.svedbergopen.com/

Research Paper Open Access

BART: An Intelligent Voice-Based AI Assistant for Personalized Human–Computer Interaction

Dr. Sachin Sukhadeo Bere[1], Dr. Devidas S. Thosar[2], Dr. Rajashree D. Thosar[3], Dr. Girish Jaysing Navale[4]

[1]Associate Professor Dattakala Group of Institutions Faculty of Engineering. E-mail: sachinbere@gmail.com
[2]Assistant Professor and Head, CSE-AIML Department. G H Raisoni College of Engineering and Management Pune.
E-mail: devidas.thosar@gmail.com
[3]Assistant Professor, E&Tc engg. Department. G H Raisoni College of Engineering and Management Pune.
E-mail: rajashreeshinde177@gmail.com
[4]Assistant Professor AISSMS INSTITUTE OF INFORMATION TECHNOLOGY, PUNE. E-mail: girish.navale@aissmsioit.org

Abstract
In the modern work of rapidly evolving technology, Artificial Intelligence has enabled the development of interactive systems that integrated speech recognization, natural language processing (NLP) and automation to enhance human- computer interaction. This paper tells about B.A.R.T (Brain Assisted Responsive Technology) a personalized desktop voice assistant that is capable of executing smart commands, generate documents, perform object detection, engage with humans in natural language and many more. BART is equipped with offline speech recognization (Vosk), text to speech (pttsx3) and Large language models to stimulate a futuristic AI companion. This system is designed to provide a hands free system control, intellignent response and automate tasks in a creative manner making it productive and accessible. This research outlines the architecture, methodology, implementation and other potentials of BART showing how AI powered assistant can be integrated smoothly in daily workflows

Keywords: Voice Assistant, Artificial Intelligence, Natural Language Processing, GPT, Speech Recognition, Human–Computer Interaction, YOLOv8, Automation Introduction.

Introduction

In today's growing world with artificial intelligence in everyday computing has a fundamentally reshaped how humans interact with machines. In traditional computers the interface like keyboards, mouse etc.. were used while limiting the natural language inputs. Voice based assistant and algorithm like siri, alexa and google assistant have become a important integral part of day to day life. However most commercially available solution rely on cloud based ecosystems with limited user customization. To address this issue we propose BART a customizable desktop AI assistant that integrates offline speech reorganization, conversational AI and smart system level automations.

Unlike the existing solutions BART emphasises on personalization, extensibility and local execution while leveraging the power of large language models LLM for context aware resoning.

The motivation for this research comes from the growing demand for personalized AI companions that goes beyond task execution to have memory, personality, bit of emotions and creative problem solving capabilities.

This paper introduces BART a fully customizable voice based assistant designed to operate as a desktop companion. BART emphasizes user control, modularity, and extensibility, allowing it to perform tasks ranging from simple system automation to advanced AI-driven functions. Having complex capabilities including local word access and document generation, PowerPoint generation and real time object detection through computer visions

Literature Review and Related Work

1. The advanced AI and NLP has transformed the field of human computer interaction leading to the development of the intelligent personal assistant that understands, interprets and responds to human speech. Early agents like ELIZA (Weizenbaum 1966) demonstrated the potential of text based systems. While modern assistants like seri, alexa etc.. have leveraged the deep learning and cloud computing to

provide context awareness and speech based interaction. These commercial agents are pretty accurate and good but they remains limited as they depends on internet connectivity and restricted control.
2.According to Shum et al.(2018) the evolution of the chatbots from rule based system to emotionally aware conversational agent marking the significant leap in the technology. However most systems rely mostly on cloud infrastructures which rises the privacy issues. In contrast open source technologies like Vosk Speech Recognition TookKit 2023 have enabled offline speech recognition allowing development of voice assistant, allowing the development for voice assistants that works on local systems.
1. Advancements in NLP have changed the conversational AI. The introduction for transformer architecture by Vaswani et al 2017 enabled powerfull models like gpt 4 integration. These LLM now forms the core of intelligent assistant capable of understanding complex human intent.
2. In computer vision the YOLO family (Bochkovskiy et al., 2020; Jocher et al., 2023); Jocher et al 2023) brought real time object detection with high efficiency. Integrating YOLOv8 allows BART to go beyond voice and text, enabling multimodel integration of tech.
3. The synthesis of speech recognition, text-to-speech, and automation has further hastened hands-free human–computer interfaces. Google (2022) research has made synthetic speech more natural, and Python-based frameworks (Python Software Foundation, 2023) enable modular integration of the same.
4. Recent research (Brown et al., 2020; Radford et al., 2018) has highlighted the role of personalization and affect-sensitive systems in enhancing user interaction. Subsequent to this, BART presents a memory-based system for adaptive and natural-like interactions.
5. Prior research in NLP , vision and automation is like a base for BART. Bart bridges the gap by integrating speech, text, vision into intelligent, extensible and executable assistant.

System Architecture and Methodology

The proposed system BART(Brain Assisted Responsive Technology) follows a modular design integrating the speech, text, vision processing and many more into a single assistant workflow. The system architecture Is divided into several components that works together to process users input and deliver a appropriate response.
The system follows a pipeline based workflow that begins with voice input and ends with multimodel output. Each component Is implemented as an independent module for ensuring extensible and ease of integration with additional service in the future.
The system architecture consist of 4 main layers
a) Input layer:
- The user provides commands via speech.
- Audio is captured using systems microphone.
- The Vosk speech recognition engine converts the speech into text locally and enables offline processing and hence reducing dependency on the cloud services.

b) Processing Layer (Natural Language Understanding)
- The recognized text is forwarded to OpenRouters GPT based API
- The LLM performs intent reorganization, dialogue generation and contextual reasoning to interpret the command.
- This step ensures that ambiguous or natural queries are translated into structured instructions foe further execution.

C) Execution Layer:
In this layer depending on the user input the different models are called and executed:
- System commands: Used to open the application, locks, check device status etc..
- Information Retrival: Search the web for contents and information, and factual questions response
- Document automation: Used to generate the word or powerpoint files using python docx anf python pptx libraries.
- Vision Tasks: Performs object detection using YoloV8 and webcam inputes.
- Entertainment: plays YouTube, music, provides humorous replies or talks like a real human friend.
- Memory/ Notes: Stores and retrieves personal information for persistent interactions.

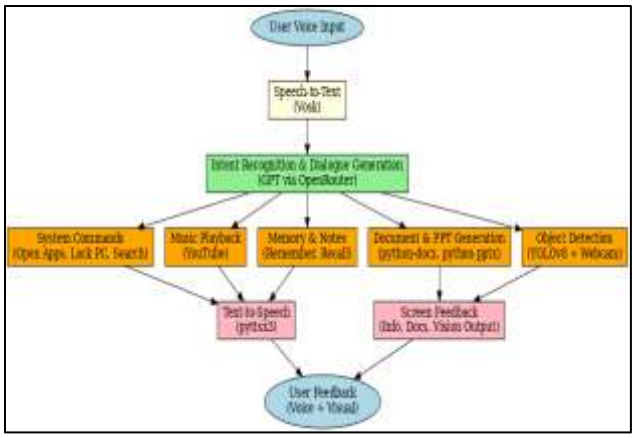

FIG 3.1: Basic System Architecture Diagram

Implementation

BART has been implemented using python-based architecture. Each functional unit has been developed as a separate module with different API's and models. The assistant includes humorous personality-driven responses for user engagement, with practical features such as time/date queries, weather reports (mock + real APIs), YouTube playback, and note management etc. This implementation is divided into following key components.

A) Speech Recognition Module:

This module is implemented using Vosk speech recognition engine, this allows offline real-time transcription of voice commands. This module also acts as a primary input interface hance enabling the assistant to continuously listen the users instructions.

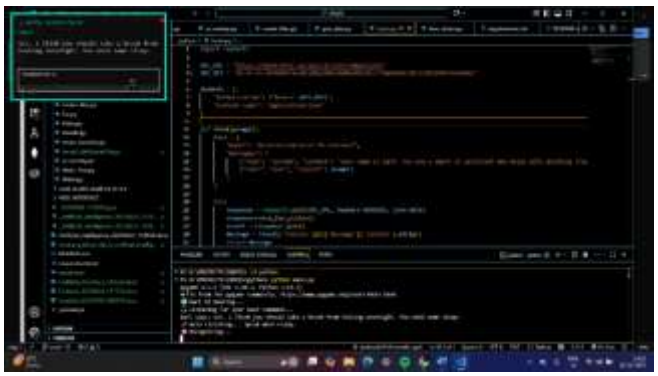

Fig4.1: Voice module with integrated emotions

B) Natural Language Processing and Conversational AI

The system is integrated with OpenAI API's providing access to state-of-the-art large language models This model processes the users' input queries and generate a natural language reply and also it is designed as it will respond to user according to the previous interaction history and the current user's mood. Omokhabi et al. (2025)

Fig 4.2: Intelligent Agent Thinking to users request

Text to Speech Output
Responses are converted to speech using the pyttsx3 libraries hence enabling interactivity and hands free communication with the assistant. This ensures a natural and human like conversation. Pareek et al. (2025)

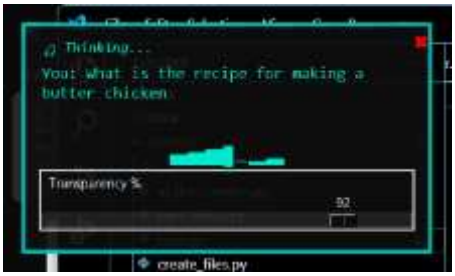

Fig 4.3: Intelligent agent working

Smart Command execution
Pythons OS and system libraries and are used to execute the smart features like opening various tabs, changing system settings and viewing details etc.

Document and Presentation Generation
This agent is also capable of generating PowerPoint presentations and word document and compared to traditional AI, BART is capable of accessing your local word and power point presentation software's and make changes in those for respective doc generations.

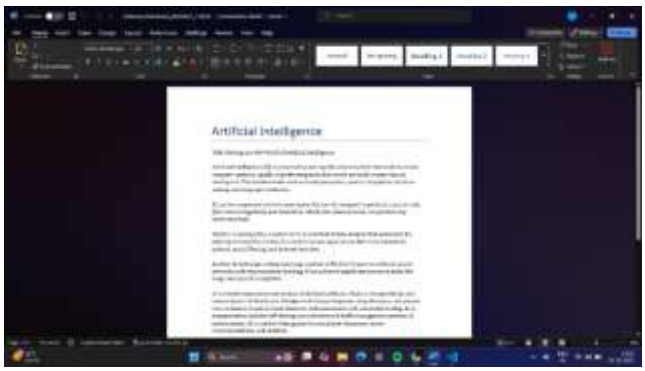

Fig 4.4: Generated Document


Fig 4.5: Generated multiple Document

Memory and Notes Module
BART keeps a track of the user's conversation and the important points and hence use this memory in further responses.

Object Detection Module:
Performs object detection using YoloV8 and webcam inputs this agents' capabilities are not only restricted to text and speech response but also has the access to the webcam for visual inputs. Tasleem, N. (2017)

Humorous and Personality Driven Replies

The assistant has its own personality making it more human and also we can see this in his responses with random jokes, sometimes anger and sometimes sad responses.

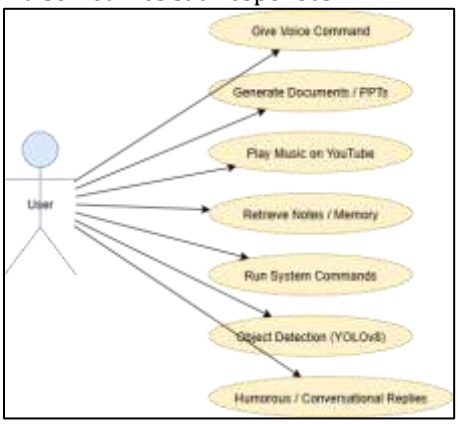

FIG 4.6: USE CASE DIAGRAM

Equations and Comparative Graph

a) Speech Recognition (Vosk)

$$\hat{W} = \arg\max_{W} P(W|X) = \arg\max_{W} P(X|W)P(W)$$

Where W is the word sequence and X is the observed acoustic feature sequence.
This illustrates the acoustic + language model approach that Vosk internally uses.

b) NLP (Intent Recognition using LLM)

$$I = \arg\max_{i} P(I_i|Q, C)$$

- I_i: possible intent,
- Q: query (user input),
- C: context or conversation history.

c). YOLOv8 Object Detection

$$C = P(\text{object}) \times \text{IoU}(\text{pred}, \text{truth})$$

- $P(\text{object})$: probability that an object exists,
- IoU: Intersection over Union metric for bounding box accuracy.

Comparative analysis of multi LLM models

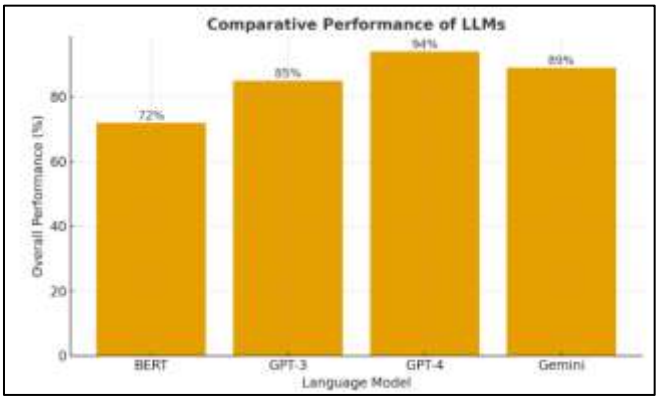

Fig 4.7: Comparative analysis of language models showing GPT 4 superior overall performance in contextual understanding, reasoning and multi model adaptability, making it more suitable for integrating with BART

Applications

The BART system has a wide range of applications ranging from education, professional, research etc. Ad a personal digital assistant it enables hands free interaction with computer by performing tasks like opening application, surfing internet, browsing data, checking system information or entertaining there by improving convenience and accessibility In daily life. In educational sector the system will support students as well as teachers by generating formatted docs and other study materials. Students may also learn new thing and explore the agent and its local working which will benefit them. In professional settings BART streamlines the workplace efficiency by automating the repetitive tasks and managing quick documentations and acting like a personal assistant for everyone boosting their efficiency.

Beyond personal and professional use case BART also serves as an assistive technology for individual with physical disabilities by providing speech control over the system, which reduces manual input devices. The integration of humorous conversational replies and other entertainment features also make the users feel like having a virtual friend creating a positive vibe. Additionally, the incorporation of computer vision through YOLOv8 object detection extends its potential into fields such as surveillance, monitoring, and smart home automation. Finally the modularity and open source nature of BART makes it a valuable research and experimental platform, enabling developers, students and researchers explore advancements in the NLP , Speech recognization and ai automation.

Conclusion

BART (Brainy AI Responsive Tool) development illustrates how AI, reorganization of speech, processing of speech, and automation can be integrated effortlessly to provide an intelligent desktop agent that is interactive. Through the integration of offline speech recognition, conversational AI, text-to-speech, intelligent command execution, and computer vision, the system is able to effectively offer a user-friendly, hands-free experience that increases productivity, accessibility, and engagement. Its capacity to create documents and presentations, remember user input, issue system-level commands, and even recognize objects exemplifies the flexibility and usability of the method. In addition, BART's modular architecture makes it scalable and adaptable for future development, making it not only a practical personal assistant but also an extensible research and experimentation platform. The project puts into practice the capabilities of AI-powered assistants in revolutionizing the human-computer interaction paradigm, bridging the human language-computer automation gap. In totality, BART proves to be a promising development towards intelligent, personalized, and accessible human-computer interaction.

References

1. Bird, S., Klein, E., & Loper, E. (2009). *Natural Language Processing with Python*. O'Reilly Media, Inc.
2. Bochkovskiy, A., Wang, C. Y., & Liao, H. Y. M. (2020). YOLOv4: Optimal speed and accuracy of object detection. arXiv:2004.10934.
3. Bochkovskiy, A., Wang, C. Y., & Liao, H. Y. M. (2020). YOLOv4: Optimal speed and accuracy of object detection. *arXiv preprint arXiv:2004.10934*.
4. Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., ... & Amodei, D. (2020). Language models are few-shot learners. *Advances in Neural Information Processing Systems*, 33, 1877–1901.
5. Chollet, F. (2017). *Deep Learning with Python*. Manning Publications.
6. Google Research. (2022). Advances in Text-to-Speech Synthesis.
7. Jocher, G., Chaurasia, A., Qiu, J., & Stoken, A. (2023). *YOLOv8: Ultralytics*. Retrieved from https://github.com/ultralytics/ultralytics
8. Omokhabi, U. S., Erumi, B. S.-U., Omilani, M. A., Omokhabi, A. A.,(2025). Empowering Women with Disabilities: Ai-Driven Reproductive Health Solutions,ShodhAI: Journal of Artificial Intelligence,2(1),40–48, https://doi.org/10.29121/shodhai.v2.i1.2025
9. OpenAI. (2023). GPT-4 Technical Report. arXiv preprint arXiv:2303.08774.
10. OpenRouter. (2023). *OpenRouter API Documentation*. Retrieved from https://openrouter.ai/
11. Pareek, R., Chavan, V. D., Sharma, B.(2025). Impact of Digital KYC and Online Account Opening on Customer Experience in Stock Broking Firms,ShodhAI: Journal of Artificial Intelligence,2(1),76–85, https://doi.org/10.29121/shodhai.v2.i1.2025
12. Python Software Foundation. (2023). *Python Language Reference, version 3.10*. Retrieved from https://www.python.org/
13. Radford, A., Narasimhan, K., Salimans, T., & Sutskever, I. (2018). *Improving language understanding by generative pre-training*. OpenAI.
14. Shum, H., He, X., & Li, D. (2018). From Eliza to XiaoIce: Challenges and opportunities with social chatbots. Frontiers of Information Technology & Electronic Engineering, 19(1), 10–26.

15. Tasleem, N. (2017). Service catalog optimization in HR. International Journal of Applied Research, 3, 1090-1097. https://www.doi.org/10.22271/allresearch.2017.v3.i1l.12550
16. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, 30.
17. Vosk Speech Recognition Toolkit. (2023). Retrieved from https://alphacephei.com/vosk/