International Journal of Scientific Research in Engineering & Technology
Volume6, Issue2 (March-April 2026), PP: 190-195.
www.ijsreat.com ISSN No: 2583-1240

Virtual Assistant for the Blind

Lakshman Madheshiya[1], Devansh Tyagi[2], Diwakar Yadav[3], Dr K. P. Jayant[4]
[1,2,3]UG Scholar, Department of CSE, Raj Kumar Goel Institute of Technology, Ghaziabad, Uttar Pradesh, India.
[4]Professor, Department of CSE, Raj Kumar Goel Institute of Technology, Ghaziabad, Uttar Pradesh, India.

To Cite this Article: *Lakshman Madheshiya[1], Devansh Tyagi[2], Diwakar Yadav[3], Dr K. P. Jayant[4], "Virtual Assistant for the Blind", International Journal of Scientific Research in Engineering & Technology, Volume 06, Issue 02, March-April 2026, PP: 190-195.*

Abstract: *The Virtual assistance for the blind Impowered the experience numerous and hurdles in maintaining effective communication with others and their knowledge of their environment to enhance growth in the society. Therefore, the virtual assistant for the blind people recently developed will help visually impaired individuals perform daily activities independently by utilizing the use of natural language processing (NLP), text to speech conversion, speech to text conversion, and object detection in real time and show the states of error which can be removed by time estimation. Natural Language Processing allows a computer or device to understand spoken commands while providing the optimal meaningful human computer interactions which helps in best utilization of AI. The assistant for the blind uses voice recognition technology for understanding user intentions and for carrying out user requests quickly and accurately. In addition, the assistant's ability is to visually recognize objects in the user's surrounding environment provides an audio representation of everything it sees. By processing audio input and utilizing natural language processing techniques, the assistant can understand the user's intent accurately, enabling it to produce clear speech and gives best benefits by the device efficiency. Index Terms AI Memory, Episodic Memory, Semantic Search, Hybrid Retrieval, Knowledge Graph, Multi Query Generation are the different parameters.*

I. INTRODUCTION

Disability presents a global challenge to millions of people from all over the world (WHO, 2021). There are two main categories of disability: the inability to participate in daily life activities and when the body does not perform in a way that is expected (WHO and WB, 2021). In addition, disability can come in many forms, including ASD, hearing or visual impairment, and physical disability (Miller et al. 2022). This types of disabilities, visual impairment is one of the most common disabilities, which influences special awareness (Martinez and others, 2017, Thinus-Blanc and Gaunet 2011). In 2010, WHO and WB issued a report with the following finding: - the total number of men and woman who have one or more disabilities was nearly 1 billion people globally.

However, by 2017 this number had been estimated to be 285 million with visual impairment i.e. 39 million who are blind recorded. This is alarming and indicates a rapid increase in Patient who are carring the disabilities.

Visually impaired individuals are unable to do everything they wish due to limitations imposed by blindness (the most serious form of visual impairment) when performing daily activities and moving about their environment independently, when little or no eyesight is able to perceive a physical world. Visual disabilities limits independence and access to information and therefore increases the dependency on others. However, with the rapid advancement in assistive technologies and intelligent digital systems, new opportunities for improving accessibility are now available everywhere and government is also working on it for the better future. AI assisted solutions are now serving as interactive platforms that connect users with the essential services required for navigation and environment recognition thereby increasing independence, increasing safety and allowing visually impaired individuals to engage more actively within the community to interact with such blind person.

They will listen to the user give them verbal instructions on where they would like to go, and then provide the user with real-time assistant.This system can perform two functions at simultaneously. First, provide scaling or map-based navigation and second, provide real time directional assistance based on changing conditions such as the user's current position and time of day.

Knowing the difficulties that visually weak users experience when accessing regular email program services, New Eyes has added functionality that allows for a voice mode virtual email system (i.e. users of this system are able to handle their email via voice commands). The new system includes additional features such as document sharing, converting voice to text, text to speech, etc.

New Eyes replaces blind sticks with advanced tools to provide navigation and communication for people who are blind. By using the advanced technology like AI guided obstacle detection and inform the email management, traditional tools like blind sticks have key limitations, which New Eyes addresses. With a multi faced application, it communicates to users about the existence of obstacles, but also continues to lead them through their environment until they reach their destination, which will help them to be independent, as well as improve the overall quality of their lives. New Eyes will allow visually impaired people to have full access to both the digital and physical worlds.

Problem Statement and Objectives:-

People with visual impairments often struggle to navigate their surroundings due to limited access to integrated and context aware assistive technologies. The majority of current products available provide limited functions using multiple independent devices, making them impractical for daily use which eventually become very costly. Smart assistant product that can effectively assist and guide visually weak users in completing multiple activities with one easy to use and integrated system.

Key points included:-

- Navigation tools that offer best directed route without considering environmental context.
- Traditional devices that complicate access and management of digital communication.
- Limited adaptive interfaces that lack usability and efficiency for voice command based control.

Objectives:-

- Creating a virtual assistant for visually weak individuals who use technology like Natural Language Processing (NLP), PyDub, and object recognition in day to day life.
- Providing real time audio based navigation assistance through advanced object recognition.
- Enabling easy communication by providing tools for converting speech to text and for managing email.
- Developing a software, voice interactive interface for users with visual impairments to make it easy to use and accessible.
- Testing the system's ability to perform accurately and software used in the actual world to be as a guardian to guide in the real time.

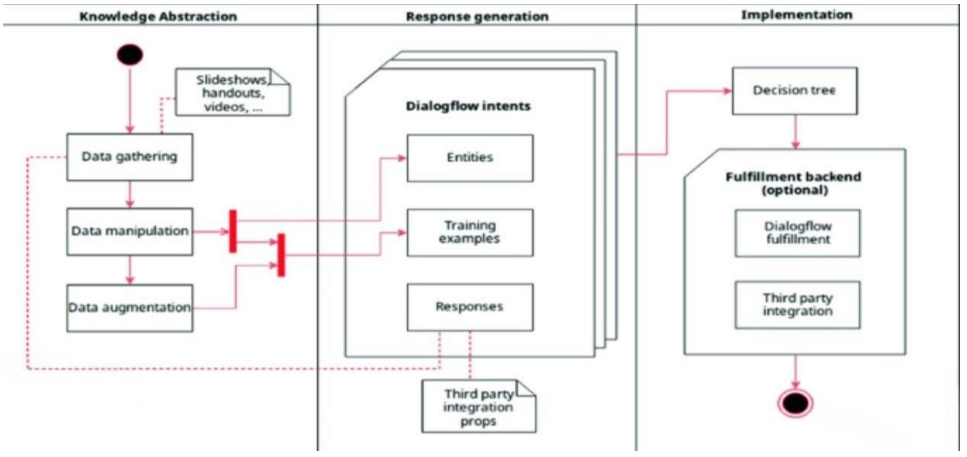
II.LITERATURE REVIEW

Summary of Virtual Assistant Research for the Visually Impaired:-

No.	Project Title	Authors	Key Features & Technology
1	Virtual Assistant for Visually Disable	Vinayak Iyer, Kshitij Shah, Sahil Sheth, Kailas Devadkar	Individuals who have visual disabilities are 31% less weak to access the internet than those who do not share that same disability. This system replaces typed input with spoken command input and incorporates using Voice to Text and Text to Voice applications facilitate the automation of the web browser. The system reads aloud to the user what they are seeing on the screen, it has the ability to help the user navigate via voice, and it provides a better overall experience by highlighting what the user sees on the page. In addition, the user's question will be processed through a BERT model which will ensure the individual receives an accurate response based on a summary of the content being viewed and distance from the obstacles which eventually helps the peoples to gets spacew in the normal society irrespective of their disabilities. This application enables to access for individuals with disabilities and sets the foundation for future integration into the Web 3.0 ecosystem and continuously improving the system for the better future.
2	Virtual Assistant for Blind People	Avanish Vijay Bahadur Yadav, Sanket Saheb Verma, Deepak Dinesh Singh	The project analyzes the development of an Android application to help blind people with the aid of AI, Machine Learning and Text or Image Recognition. Its features include a voice assistant, image and currency identification, virtual pdf book access and a chatbot. Python libraries (e.g. speech recognition) provide developers with tools for writing software that provides efficient text to speech functionality and efficient speech to text functionality, therefore enabling people to access these technologies.
3	Voice Controlled Virtual Assistants for Older People	Ho. D. Ho	The project is aimed to build an Android app to help blind people with AI, Machine Learning, and Image or Text Recognition to help in the critical situation. The app will feature a Voice Assistant that can greet users, tell them the time & date, open Apps, search

No.	Project Title	Authors	Key Features & Technology
			the web, and solve math. Images and currency recognition by the app. Digital Books will be accessible through the App and a Chatbot will be provided to communicate with the user. The Voice Assistant uses library software written in Python such as pyttsx3 and speech recognition to provide fast and accurate text to speech and speech to text functions. With these capabilities, the app will provide further access to technology and will make it easier for anyone in the blind community to use and make their life more easy and can have some better access to communicate.
4	Voice Based Email Assistant for Visually Impaired	Sunanda, Dr. B. Esther, Asritha, Mallajyosula Sai, Aruna Matcha	This project provides an email system based on Virtual Voice Technology for people with visual impairment. The email text will be converted to speech, and speech will be converted to email text using the VVE technology. This will allow for an accessible, effortless way to communicate digitally via email and will make the system much easier to use and more convenient for the user.
5	Voice Based Email Generating System using AI	Gaurav Kumar Rajput, Sachin Sharma, Bibhu Prasad Dash, et al.	A voice driven email system created and managed by combining Python programming and the application of Artificial Intelligence technologies has been designed to help so that visually disable individuals communicate openly and feel relaxed.
6	CNN Based Object Recognition and Tracking System	Fahad Ashiq, Muhammad Asif, Maaz Bin Ahmad, et al.	The system offers real time voice guided navigation and direction and a web based application for sharing locations with family, ensuring remote tracking for enhanced security and safety feature.
7	Guided System for Visually Disabled People	Kanchan Patil, Avinash Kharat, Pratik Chaudhary, et al.	Here we propose a hands free wearable device with an on board virtual assistant for use by visually impaired people to assist them in performing physical tasks like walking around, locating items, identifying people, and reading. This device will utilize deep learning and written in Python, offer an intuitive voice command interface and hardware buttons for ease in operation.

III.PROPOSED METHODOLOGY



Virtual Assistant System is designed in such a way that technology that allows blind people to use the system easily and

efficiently in order to provide assistance to them. Virtual assistant for the blind used the audio and speach recognition to enhance the user experience. these methods are widely used for the friendly behavior for the needed people.

1. Speech to Text and Text to Speech.

This module handles voice input and output enabling users to communicate naturally with the assistant.

Tools: Speech Recognition, PyDub, Pyttsx3, gTTS

- **Processes:-**

a. Capture uses command via a microphone, Convert speech to text using Speech Recognition system.
Analyze text with natural language processing for user intent.
b. Respond and voice with text to speech by using adjustments for tone, speed, and volume with pydub.
c. Natural language processing is a capacity of machine to understand and process language.
d. The assistant will accurately process user commands and complete tasks such as navigation, email management, calling facilities.

e. Process:-

a. Identify context for user commands which to process the problem statement.
b. Classify commands based upon category (e.g. navigation, email management, object identification).
c. Support complex multiple step commands (e.g. Read my latest email then respond to John's).

Object Recognition:-

Virtual assistant system uses real time model system to identify objects to provide important information about the surrounding to the user who are unable to know the real life activity in fast time as a normal person.
- Techniques used include OpenCV, YOLOv5 and Tensor flow.

- The steps would be as follows:-

a) Trained YOLOv5 model is used to helps in identifing objects in different patterns.
b) Provide audio feedback about the objects identified, their location relative to the user and their proximity to the user.

1. Email and Communication Management:-

Users can use voice commands to help them manage their emails in a more efficient way through this module.

Tools: - IMAP Client, Simple Mail Transfer Library, and Email.

Process:
a. Emails are read aloud through modules text to speach.
b. Email reply and compose drafts through speach to text.
c. Perform Email Functions mark emails as read or unread, delete and send emails.

2. Navigation Assistance :-

This module has been developed to help provide safe and guided mobility in both outdoor and indoor environments with multiple different types of tools including a GPS module and Object Recognition.

Tools – GPS Module & Object Recognition.

Processes:-

a. Use GPS to provide turn by turn navigation instructions while outside.
b. Use Object Recognition for Indoor Navigation to enable users to find obstacles and locate important objects which helps the visually disabled to communicate.
c. Provide step by step Navigation Assistance through Text to speach.

IV. EVALUATION

To enable a complete and accurate assessment of both how well memory operates within the confines of the typical conversation environment for an extended amount of time, Dynamic memory operation will be compared to many other large scale systems on multiple datasets using standard evaluation metrics for that task.

Datasets -

a) Single hop, multi hop and temporal are the best data sets to utilize and increase the efficiency.
b) The LoCoMo Dataset consists of long context dialogues and is therefore very different from the traditional datasets.
c) Open domain are questions with grounds to be assigned both in dialogue context and outside of that.
d) Adversarial is a misleading queries that test robustness.

Dialslim: - DialSim provides an additional way to test how well using memory has been utilized in real-world application we choose DialSim [5] as our secondary dataset. The DialSim is made up of real-world, multi-party long-term dialogue collected from popular television shows (i.e. Friends, Big Bang Theory, and The Office). There are approximately 1300 sessions heterogeneous in nature over a span of 5 years equating to over 350,000 tokens. Each session contains roughly 1000 unique questions derived from both well curated fan quiz data and temporal knowledge graphs. This results in a very challenging benchmark for testing episodic memory and temporal reasoning capabilities.

Baselines:-

We compare Dynamic memory against several state of the art conversational memory systems.
● LoCo Mo baseline [4], which is the reference system on which long context evaluation is running.
● Mem GPT [7], a recent system that integrates episodic memory mechanisms into GPT based agents.

Metrics:-

We use two metrics that are already popular in the field for assessing the quality of answers:
F1 Score: Measuring of accuracy by taking into account balancing of precision and recall, for fact based question answering.

Evaluation Goals:-

Using this assessment we measure:

- The capability of D-mem to be consistent throughout long dialogues.
- The impact of hybrid retrieval and multi query generation to strengthen the retrieval.
- Episodic reflection is a role in the dealing with complex Queries that are temporal and adversarial.

With the benchmark against LoCoMo, DialSim, and the strong baselines, we present the in depth analysis of the Dynamic memory capabilities in the real world and memory intensive conversational environments.

V. FUTURE SCOPE AND CONCLUSION

Modern era is the time where the review and the feedback of every indivituals are recognized equally where the virtual assistent play the best role in the development of new technologies to provide the plateform through which the disabled peoples can also contribute in the decision making and make the best possible conclusion at their end.

Contributions and Take aways :-

- Bi-temporal graph backbone logs the validity time (T) and transaction time (T′).
- Hybrid lexical, semantic retrieval using both BM25 and embedding similarity allows for balanced retrieval types for the 2 primary query types.
- Reflection drive by the large language model on episodic traces to create structured triplets and context summaries for potential promotion to semantic memory.
- Mechanisms used to promote temporally, resolve conflicts, and retain a history of events associated with the current state of knowledge.

Future research directions:-

- Improve the architecture layout to allow for multimodal episodic memory (e.g. visual, auditory) to be integrated within the system. Automate consolidation and forgetting through human in the loop feedback and reinforcement, enabling systems to learn optimal thresholds for promotion ranking over longer periods of time.
- Integrating additional cognitive modules such as procedural memory, a transient working memory layer to allow short term reasoning and planning.

We proposed a dual memory system called Duel memory Temporal. By combining Hybrid and Lexical, Semantic Memory Retrieval, Reflecting using the use of a Large Language Model, and Consolidating learned material, Duel memory Temporal prepares intelligent agents to have dynamically maintained, Correctly time stamped, and Interpretable memories.

Key Contributions and Findings:-

- A bi-temporal graph with Validity (T) and Transaction (T') Dates, this allows for temporal reasoning around recal, conflict resolution and source tracing.
- Language based Reflection translates episodic memory into structured Triplet and Context Summary so that it eligible for promotion to semantic memory.
- Mechanisms for promoting temporally and resolving Conflict record history, and maintain Knowledge at the present state.

Future Research Directions:-

- Incorporating more fusion techniques to generate dynamic weighting between language processing and Semantic Processing depending on the intent of the Query.
- Enhancing topology for architecture development to support Multimodal Episodic Memory.
- Automating consolidation and forgetting by means of Reinforcement or involving Humans in the Loop, thus enabling Learning of the best promotion Threshold configuration over Time.
- Adding additional Cognitive Modules (e.g. a procedural memory) to function as a temporary Working Memory layer to facilitate Short-Term Thinking and Planning.

All of the D-MEM T methods share common components such that time-varying neural memory systems provide a cognitive basis for prize-related learning through the symbolic representation, recall of knowledge using both external processing and continuous learning.

REFERENCES

1. Kuang-Huei Lee, Xinyun Chen, Hiroki Furuta, John Canny, Ian Fischer, “Generative agents: Interactive simulacra of human behavior,” in UIST, (2023).
2. Mathis Pink, Qinyuan Wu, Vy Ai Vo, Javier Turek, Jianing Mu, Alexander Huth, Mariya Toneva. “Episodic Memory is the Missing Piece for Long-Term LLM Agents.” arXiv: 2502.06975, 2025.
3. Wujiang Xu, Zujie Liang, Kai Mei, Hang Gao, Juntao Tan, Yongfeng Zhang. “A-Mem: Agentic Memory for LLM Agents.” arXiv: 2502.12110, 2025.
4. Adyasha Maharana, Dong-Ho Lee, Sergey Tulyakov, Mohit Bansal, Francesco Barbieri, Yuwei Fang. “Evaluating Very Long-Term Conversational Memory of LLM Agents.” arXiv preprint arXiv: 2402.17753, 2023.
5. Jiho Kim, Woosog Chay, Hyeonji Hwang, Daeun Kyung, Hyunseung Chung, Eunbyeol Cho, Yeonsu Kwon, Yohan Jo, Edward Choi. “DialSim: A Dialogue Simulator for Evaluating Long-Term Multi-Party Dialogue Understanding of Conversational Agents.” arXiv preprint arXiv: 2406.13144, 2024.
6. Kuang-Huei Lee, Xinyun Chen, Hiroki Furuta, John Canny, Ian Fischer. “A Human-Inspired Reading Agent with Gist Memory of Very Long Contexts.” arXiv preprint arXiv: 2402.09727, 2023.
7. Charles Packer, Sarah Wooders, Kevin Lin, Vivian Fang, Shishir G. Patil, Ion Stoica, Joseph E. Gonzalez. “MemGPT: Towards LLMs as Operating Systems.” arXiv preprint arXiv: 2310.08560, 2024.
8. Theodore R. Sumers, Shunyu Yao, Karthik Narasimhan, Thomas L. Griffiths. “Cognitive Architectures for Language Agents.” Transactions on Machine Learning Research, February 2024. (CoALA framework; modular memories, action/decision loop). Available at arXiv: 2309.02427.
9. Zeyu Zhang, Xiaohe Bo, Chen Ma, Rui Li, Xu Chen, Quanyu Dai, Jieming Zhu, Zhenhua Dong, Ji-Rong Wen. “A Survey on the Memory Mechanism of Large Language Model Based Agents.” arXiv preprint arXiv: 2404.13501, 2024. (Comprehensive taxonomy of memory sources, forms, operations, evaluations).
10. Kishore Papineni, Salim Roukos, Todd Ward, Wei-Jing Zhu. “BLEU: A Method for Automatic Evaluation of Machine Translation.” In Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL), 2002, pp. 311–318.
11. J. L. McClelland, B. L. McNaughton, R. C. O’Reilly. “Why There Are Complementary Learning Systems in the Hippocampus and Neocortex: Insights from the Successes and Failures of Connectionist Models.” Psychological Review, vol. 102, no. 3, pp. 419–457, 1995.
12. E. Tulving. “Episodic and Semantic Memory.” In Organization of Memory, Academic Press, 1972, pp. 381–403.
13. P. Lewis et al. “Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks.” In NeurIPS, 2020.
14. S. E. Robertson, S. Walker. “Some Simple Effective Approximations to the 2-Poisson Model.” In SIGIR, 1994.
15. N. Reimers, I. Gurevych. “Sentence-BERT.” In EMNLP, 2019.