STM JOURNALS

Journal of
Multimedia Technology &
Recent Advancements

ISSN: 2349-9028
Volume 12, Issue 3, 2025
September–December
DOI (Journal): 10.37591/JoMTRA

https://journals.stmjournals.com/jomtra

Review JoMTRA

Innovative Applications of AI and Machine Learning in Voice-Driven Systems: Friday AI 2.0

Raju Kumar[1], Navjot Kaur Cheema[1], Jaspreet Kaur[2,*], Pushpinder Singh[1], Sardar Randhir Singh[1]

Abstract

The development of artificial intelligence (AI) has enabled the development of increasingly sophisticated models that can understand and react to human questions. This study introduces Friday AI-2.0, a cutting-edge AI model that improves user engagement by using natural language processing (NLP) and voice synthesis technologies to produce text-based and audio-based responses. By increasing engagement through an intuitive and engaging interface, Friday AI-2.0 seeks to improve the user experience. By describing the architecture, design, and operation of the model, this study demonstrates its ability to mimic expressive, real-time communication. Numerous features are supported by the system, such as facial authentication, device information retrieval, object recognition using YOLO, action automation (scrolling, tab control, and YouTube playback), and interaction with the Gemini AI chatbot for intelligent chats. Additionally, it offers text-to-image generation, Wikipedia search, WhatsApp messaging and calling, real-time weather reports, and automation for accessing apps and websites. Friday AI-2.0 is an engaging assistant that also has entertaining features like joke generating and random advice. Tasks like launching desktop programs, playing particular YouTube videos, sending WhatsApp messages, and placing audio and video conversations may all be accomplished with voice-controlled interactions. This technology is especially useful for multitasking and assistive applications since it simplifies processes, encourages hands-free interaction, and increases accessibility by doing away with the need for human input. Through this research, the aim is to demonstrate Friday AI-2.0 as a noteworthy step toward more intelligent and engaging AI systems.

Keywords: Artificial intelligence (AI), Natural language processing (NLP), voice synthesis, Friday AI-2.0, YOLO

***Author for Correspondence**
Jaspreet Kaur
E-mail: jaspreet.kaur@bbsbec.ac.in

[1]Student, Department of Computer Science and Engineering, Baba Banda Singh Bahadur Engineering College, Fatehgarh Sahib, Punjab, India
[2]Assistant Professor, Department of Computer Science and Engineering, Baba Banda Singh Bahadur Engineering College, Fatehgarh Sahib, Punjab, India

Received Date: April 28, 2025
Accepted Date: June 16, 2025
Published Date: September 15, 2025

Citation: Raju Kumar, Navjot Kaur Cheema, Jaspreet Kaur, Pushpinder Singh, Sardar Randhir Singh. Innovative Applications of AI and Machine Learning in Voice-Driven Systems: Friday AI 2.0. Journal of Multimedia Technology & Recent Advancements. 2025; 12(3): 1–9p.

INTRODUCTION

Artificial intelligence (AI) has advanced so quickly that it has completely changed how people interact with technology. Artificial intelligence (AI) has evolved from being a passive tool for information retrieval to an active participant in human discussions, exemplified by virtual assistants such as Siri and Alexa. In a wide range of fields, AI models can now comprehend complex queries, participate in contextual discourse, and provide thorough answers. Friday AI-2.0 is an artificial intelligence model designed to deliver human-like responses in a cohesive multimedia format, combining text and audio.

Conversational agents, sometimes known as virtual personal assistants, can interact with humans

via text, speech, or image to complete everyday tasks such as setting alarms, and making calls (audio and video). Although conventional search engines show millions of outcomes in connection with a user-initiated search, contrasting outcomes and obtaining a file result, VPA is designed to provide the present projections as the only result. VPAs are rapidly developing into capable, reliable, and advantageous resources. Certain devices employ a deep learning method to facilitate speech with excellent clarity acknowledgment [1].

Friday AI-2.0 enhances task automation by integrating a variety of advanced features. It enables hands-free operation for tasks such as opening applications, browsing websites, playing YouTube videos, and sending WhatsApp messages. Additionally, it supports face authentication, real-time object detection using YOLO, Wikipedia search, weather updates, and text-to-image generation, making it a versatile tool for both personal and professional use. In addition to its task-oriented capabilities, Friday AI-2.0 incorporates natural language processing (NLP) to accurately read user commands, guaranteeing that the system can reply effectively to complex or context-dependent requests. This functionality goes beyond basic commands and allows for deeper conversational interactions with the user. The model's voice synthesis technology goes even farther, providing authentic responses that mirror human speech patterns, resulting in a more immersive experience.

Friday AI-2.0 is further distinguished by its capacity to function as a Large Language Model (LLM), similar to models such as ChatGPT and Google's Gemini by integrating Google Gemini API. This enables Friday AI-2.0 to not only execute voice commands, but also engage in complicated, natural, and context-sensitive discussions. It can provide extensive and insightful responses, explain concepts, and give content in real time, greatly expanding its application in a variety of settings such as education, customer service, and entertainment. With LLM capabilities, Friday AI-2.0 may function as a conversational agent, aiding users with both task-based and knowledge-driven responses, resulting in a more comprehensive AI-human connection.

This study will look into the technological improvements of Friday AI-2.0, namely the integration of voice control, speech synthesis, NLP, and LLM functions, as well as the possible applications in both personal and professional settings. Friday AI-2.0 marks a big step forward in the development of voice-controlled and AI-powered systems by boosting user ease, fostering hands-free interactions, increasing accessibility, and providing extensive conversational capabilities in AI-Human interaction.

LITERATURE SURVEY

Virtual personal assistants (VPAs) and speech recognition systems have undergone significant change as a result of the development of artificial intelligence (AI) and deep learning (DL). Natural language processing (NLP), automation, reinforcement learning (RL) for dialogue management, and deep learning-based speech recognition systems are some of the methods that researchers have investigated to enhance voice-based assistants. By improving usability and functionality, cutting-edge AI and machine learning applications in voice-driven systems are revolutionizing human-computer interaction. These developments make use of technology like machine learning algorithms, speech recognition, and natural language processing (NLP) to provide more responsive and user-friendly solutions.

The Development of AI-Powered Virtual Assistants

AI voice assistants improve user experience in desktop settings by using Python and open-source tools for task automation and speech recognition [2]. A Python-based desktop voice assistant that can process offline speech recognition to handle user actions including launching apps, sending emails, and conducting online searches was created by Gupta *et al*. [3].

The study emphasizes how speech-to-text technology is being used more and more in automation and human-computer interface (HCI). According to the study, the ease of use of speech recognition

technology in human-computer interaction has led to its growth in a variety of fields. These systems are flexible and customizable, enabling users to easily obtain information, make reminders, and manage apps [4].

Building on this, Mahesh investigated how to incorporate API-based automation and Natural Language Processing (NLP) into a personal AI assistant. In contrast to conventional virtual assistants, this model places a strong emphasis on customization, enabling users to alter its features to suit their own needs [5]. The study shows that by comprehending contextual inquiries and providing dynamic responses, NLP-driven assistants can improve user engagement. The process of speech recognition is shown in Figure 1.

The performance, capabilities, and drawbacks of the main Virtual Personal Assistants (VPAs), including Amazon Alexa, Microsoft Cortana, Google Assistant, Apple Siri, and Samsung Bixby, are thoroughly examined in a study by Jaber *et al*. [6]. Vendor platforms, device compatibility, speech recognition effectiveness, natural language processing (NLP) capabilities, and practical usefulness are some of the factors that the study considers when evaluating these assistants. With Alexa being the most popular smart home automation tool, Siri dominating in Apple ecosystem integration, and Google Assistant leading in voice recognition accuracy, it demonstrates how various VPAs apply unique AI-driven technologies to improve user interaction. But across all VPAs, privacy worries, data security problems, and device limits continue to be major obstacles. The variations between Alexa, Cortana, Google Assistant, Siri, and Bixby are shown in Table 1.

Reinforcement Learning for AI-Powered Dialogue Management

In order to improve user interactions, Den Hengst *et al*. investigated the use of reinforcement learning (RL) for personalized AI assistants and suggested a dialogue system based on Partially Observable Markov Decision Processes (POMDP) [7]. Two RL-based techniques are shown in their work:

1. *Single RL Learner:* Customizes replies by extending POMDP with user-context characteristics.
2. *Segmentation-Based RL Learner:* Develops distinct RL models for various patterns of user engagement.

Deep Learning for Speech Recognition: A Shift from Traditional Models

Hinton *et al*. made a major contribution to the field of speech recognition systems by proving that Deep Neural Networks (DNNs) perform better in acoustic modeling than conventional Gaussian Mixture Models (GMMs) [1].

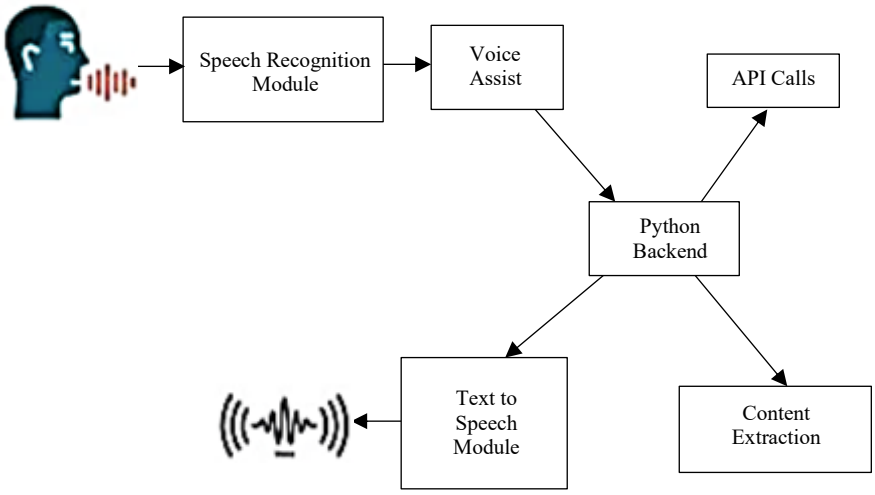

Figure 1. Process of speech recognition [5].

Table 1. The variations between Alexa, Cortana, Google Assistant, Siri, and Bixby [6].

Key Features	Alexa	Cortana	Google Assistant	Siri	Bixby
Vendor	Amazon	Microsoft	Google	Apple	Samsung
Devices	Echo Dot, Fire Tablet	Windows 10 Devices	Android, Google Home	iPhone, iPad, Mac, Apple Watch	Samsung Galaxy Phones and Tablets
Launch Date	2014	2009	2016	2011	2017
Wake Word	Alexa	Cortana	OK Google/Hey Google	Hey Siri	Hi Bixby
Voice Recognition and Human-Free Interaction Accuracy	7.91%	28.42%	59.80%	43.98%	37.9%
Technology Used	NLP + CI	NLP	NLP + CI	NLP	NLP
Social Games Usefulness	Success	Success	Success	Failure	Success
Social Media Integration	Success	Success	Failure	Success	Success
Email Notifications	Success	Failure	Failure	Possible	Success
Limitations	Lack of data security	Limited to one user	High data consumption, less user-friendly	Apple device exclusive	Data security concerns
Communication Mode	Voice-based	Text + Voice	Text + Voice	Text + Voice	Text + Voice

In order to simulate speech fluctuations, traditional speech recognition mostly relied on Hidden Markov Models (HMMs) in conjunction with GMMs. Nonetheless, the study shows that DNNs trained using contemporary methods can outperform GMMs across a number of metrics.

Key Findings

When compared to GMMs, DNN-based models dramatically lower word error rates (WERs). When it comes to feature extraction in speech recognition, deep learning frameworks do better than conventional statistical models. Pretraining and classification accuracy are enhanced by deep belief networks (DBNs) and restricted Boltzmann machines (RBMs). The paper states, *"Deep neural networks that have many hidden layers and are trained using new methods have been shown to outperform GMMs on a variety of speech recognition benchmarks, sometimes by a large margin."*

Xiong *et al*. at Microsoft AI and Research broke the previous record with a 5.1% Word Error Rate (WER) on the Switchboard conversational speech challenge, building on DNN-based voice recognition [8]. The conversational voice recognition system they developed in 2017 included: Acoustic processing CNN-BLSTM (Convolutional Neural Network-Bidirectional Long Short-Term Memory) models and Contextual language modeling using session-aware and character-based LSTMs. To increase the accuracy of voice transcription, a two-stage model combination is used at both the senone/frame and word levels. According to the study, their technology performed better than earlier standards and even surpassed human transcription accuracy on some jobs. According to the authors, "The resulting system yields a 5.1%-word error rate on the 2000 Switchboard evaluation set." With its demonstration of how CNN-BLSTM and hybrid architectures can maximize speech-to-text accuracy beyond human capabilities, this work represents a significant advancement in deep learning for speech recognition.

1. *Applications of Voice Analysis: Predicting Gender and Age:* Machine learning models are able to use voice characteristics to forecast demographic data, which is useful for user profiling and marketing. The identification of synthetic speech and the ability of sophisticated voice analysis algorithms to distinguish between synthetic and natural speech enhances the quality of interactions in a variety of applications [4].
2. *Security and Vulnerabilities: Obstacles in Healthcare:* Vulnerabilities in machine learning systems for voice disorder detection might compromise their dependability, requiring stronger security protocols [9].

3. *Avatar-Based AI Assistants:* In order to improve user interaction and experience, avatar-based AI assistants are being included into a growing number of industries, most notably healthcare and education. These artificial intelligence (AI)-powered avatars use cutting-edge tools like machine learning and natural language processing to offer interesting and customized interactions [10].

IMPLEMENTATION

Problem Statement

While AI assistants are becoming more common, their interaction modes are still limited to text and audio output. This reduces user engagement and makes interactions less immersive. Additionally, many AI assistants lack the ability to execute complex system commands efficiently, limiting their usefulness in real-world automation tasks.

Challenges

1. *Limited Interaction Modes:* Most AI assistants rely on either text or audio-based responses, lacking real-time visual feedback, which could enhance user experience.
2. *Real-Time System Commands Execution:* Many AI assistants cannot perform system-level tasks such as opening applications, adjusting settings, or automating workflows.
3. *Context Retention in Conversations:* Maintaining user context over multiple interactions is essential for an intelligent assistant but remains a major challenge.
4. *API Integration:* Efficiently integrating APIs like Gemini AI for better NLP understanding and command execution requires optimized implementation.

PROPOSED SOLUTION

To overcome these challenges, Friday AI 2.0 has been designed as an advanced AI-driven voice assistant that integrates voice commands, action automation and system controls to enhance user experience and productivity. Below are the key functionalities incorporated into the proposed system:

1. *Action Automation in Playback and Browser execution:* Enables hands-free navigation on web pages and applications.
2. *Device Information Retrieval:* Displays system status in real time. And allows users to modify screen brightness via voice control.
3. *Face Authentication*: Uses AI-based face recognition to verify users before executing commands.
4. *Database Management System Commands Storage*: Organizes and retrieves system-level commands i.e. Contacts and Web Commands and other necessary files.
5. *Gemini AI Integration*: Leverages Google's Gemini API to provide intelligent and context-aware responses and saves chat interactions for future reference.
6. *Real-Time Object Detection: YOLOv8 Integration*: Utilizes AI vision for detecting objects through webcam input.
7. *Enhanced User Interface and Interaction with Text and Voice Inputs:* Accepts commands via typing or speech with Chat History Management.
8. *Application and Website Access:* Opens installed applications on command. And Automated Website Navigation using voice input.
9. *Multimedia and Online Services:*
    - YouTube Video Search and Playback: Finds and plays videos via voice commands.
    - Text-to-Image Generation: Generates images based on text descriptions.
    - Weather Reports: Fetches real-time weather updates.
    - WhatsApp Automation: Supports text messaging, voice calls, and video calls.
    - Wikipedia Search: Retrieves and verifies information using Wikipedia.

RESULTS

Figures 2–7 show the step by step interfaces occurred after every interface.


```
Friday A.I. response for Prompt: using AI tell me about Hajipur Bihar
 ************************

I can definitely help you with that! Using artificial intelligence, here's what I found about Hajipur, Bihar:

Hajipur is a city located in the Vaishali district of Bihar, India. It is situated on the banks of the Gandak River,
a tributary of the Ganges. Hajipur is known for its historical significance and cultural heritage.

Here are some interesting facts about Hajipur:

* **History:** Hajipur has a rich history dating back to ancient times. It was once known as Hirawaapura or Hirapur,
and it was an important center of learning during the Gupta Empire. Hajipur was also a major center of Jainism
during the medieval period.

* **Culture:** Hajipur  is famous for its Chhath Puja celebrations, a four-day Hindu festival dedicated to the Sun god Surya.
The city is also known for its traditional crafts, such as handloom weaving and pottery.

* **Places of interest:** Some of the popular places of interest in Hajipur include the Hajipur Museum,
which houses a collection of artifacts from the region's history; the Maa Vaishno Devi Mandir, a Hindu temple
dedicated to the goddess Vaishno Devi; and the Sonepur Mela, one of the largest cattle fairs in Asia.

I hope this information is helpful!  Would you like to learn more about anything specific in Hajipur?
```

Figure 2. Gemini response.

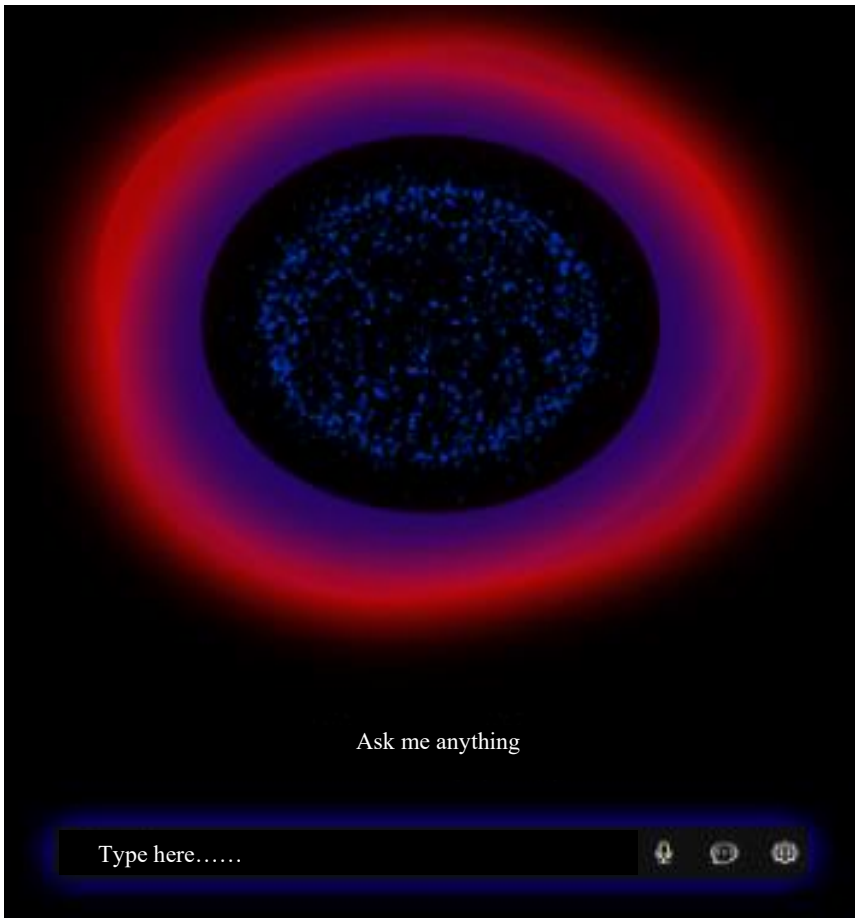

Figure 3. Friday AI 2.0 user interface.

Figure 4. Object detection command.

Figure 5. Object detection results.

Figure 6. Image generation command.

Figure 7. Generated image results.

CONCLUSION

The Friday AI 2.0 project marks a transformative advancement in AI-powered voice assistant technology by seamlessly integrating natural language processing (NLP), voice synthesis, real-time object detection and action automation into an intuitive and efficient system. Unlike conventional virtual assistants, Friday AI 2.0 offers enhanced interactivity, automation, and accessibility, making it a robust solution for both personal and professional use.

One of the key differentiators of Friday AI 2.0 is its ability to execute complex system commands, automate everyday tasks, and engage users through an interactive AI. The integration of face authentication, WhatsApp automation, multimedia handling, and real-time web interactions further solidifies its utility. By providing hands-free voice control, the system significantly enhances productivity and user convenience.

Moreover, Friday AI 2.0 leverages Google Gemini AI for intelligent conversations and YOLO-based object detection for real-world applications, bridging the gap between voice assistants and real-time interactive AI. With its advanced database management, automated browser control, and weather forecasting capabilities, the system caters to diverse user needs, from entertainment to task automation.

Future Scope

Together, these studies advance speech recognition and AI-powered virtual assistants.

Important topics requiring further study include:

1. Improving AI assistants for interactions that are context-sensitive and multilingual.

2. Enhancing voice recognition in loud settings.
3. Creating hybrid AI models for context-aware, adaptive VPAs by fusing on deep learning, natural language processing, and reinforcement learning.
4. Notwithstanding the substantial advantages of these developments, problems like security flaws and the requirement for reliable algorithms continue to be important subjects for more study and advancement. For voice-driven systems to continue to advance, these problems must be resolved.
5. Avatar-Based AI Assistants: Although the use of AI avatars has the potential to enhance healthcare and education, there are still issues with guaranteeing their efficacy and dependability in a variety of settings. To solve these problems and maximize its implementation, more research is necessary.

REFERENCES

1. Hinton G, Deng L, Yu D, Dahl GE, Mohamed AR, Jaitly N, Senior A, Vanhoucke V, Nguyen P, Sainath TN, Kingsbury B. Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. IEEE Signal Process Mag. 2012 Oct 18; 29(6): 82–97.
2. Gowthamy J, Senthilselvi A, Kumar A, Aakash S, Sreedhar G. Enhanced AI Voice Assistance using Machine Learning and NLP. In 2023 IEEE Third International Conference on Smart Technologies, Communication and Robotics (STCR). 2023 Dec 9; 1: 1–5.
3. Gupta U, Jindal U, Goel A, Malik V. Desktop Voice Assistant. Int J Res Appl Sci Eng Technol. 2022 May; 10(5): 901–905.
4. Tinge A, Saumya A, Nair VN, Biradar Y, Manjunath TC. Intelligenie-A Voice based Desktop Assistant. Grenze International Journal of Engineering & Technology (GIJET). 2024 Jun 5; 10:p1666-1672.
5. Mahesh TR. Personal ai desktop assistant. Int J Inf Technol Res Appl. 2023 Jun 22; 2(2): 54–60.
6. Jaber K, Lafi M, Alkhatib AA, AbedAlghafer AK, Jawad MA, Ahmad AQ. Comparative Study for Virtual Personal Assistants (VPA) and State-of-the-Art Speech Recognition Technology. Int J Comput Exp Sci Eng. 2024; 10(3): 427–433.
7. Den Hengst F, Hoogendoorn M, Van Harmelen F, Bosman J. Reinforcement learning for personalized dialogue management. In IEEE/WIC/ACM International Conference on Web Intelligence. 2019 Oct 14; 59–67.
8. Xiong W, Wu L, Alleva F, Droppo J, Huang X, Stolcke A. The Microsoft 2017 conversational speech recognition system. In 2018 IEEE international conference on acoustics, speech and signal processing (ICASSP). 2018 Apr 15; 5934–5938.
9. Perelli G, Panzino A, Casula R, Micheletto M, Orrù G, Marcialis GL. Vulnerabilities in Machine Learning-Based Voice Disorder Detection Systems. In 2024 IEEE International Workshop on Information Forensics and Security (WIFS). 2024 Dec 2; 1–6.
10. Halim NA, Azlan MH, Ismail AW, Fazli FE, Ahmad MA, Aladin MY. Edu-Metaverse Classroom with AI-Driven Virtual Avatar Assistant. In 2024 IEEE 5th International Conference on Smart Electronics and Communication (ICOSEC). 2024 Sep 18; 1548–1554.

