General Introduction

In recent years, the rapid advancement of artificial intelligence (AI) and embedded systems has paved the way for the development of intelligent virtual assistants capable of performing a wide array of tasks. This dissertation explores the integration of these technologies to create a versatile virtual assistant using the ESP32 microcontroller, OpenAI's ChatGPT, and Google APIs. The objective is to design a system that can execute office automation tasks such as scheduling, information retrieval, and smart device control efficiently and effectively.

The growing demand for smart and autonomous systems in various domains, from personal assistance to industrial applications, underscores the significance of this research. Embedded systems, characterized by their compact size, low power consumption, and dedicated functionality, offer an ideal platform for deploying intelligent edge devices. The ESP32 microcontroller, known for its robust performance and versatility, serves as the hardware foundation for our virtual assistant. Its integration with AI models and cloud-based services aims to demonstrate the potential of combining edge computing with powerful AI capabilities.

The use of OpenAI's ChatGPT, a state-of-the-art natural language processing (NLP) model, is central to the virtual assistant's ability to understand and generate human-like responses. This integration leverages the model's proficiency in handling complex language tasks, enhancing the assistant's interaction quality and user experience. Additionally, Google APIs provide access to a wide range of services, enabling the assistant to perform tasks such as fetching real-time data, managing schedules, and controlling smart devices with ease.

This dissertation is structured as follows: Chapter 1 introduces the research problem and objectives. Chapter 2 reviews relevant literature on embedded systems, AI-powered virtual assistants, and edge computing. Chapter 3 details the methodology, including system design, firmware development, API integration, and optimization techniques. Chapter 4 presents the results and discussion, focusing on performance metrics and testing outcomes. Finally, Chapter 5 concludes with a summary of findings, implications, and suggestions for future research.

Through this research, we aim to contribute to the growing body of knowledge on intelligent edge devices and their applications, demonstrating how combining embedded systems with advanced AI models can create efficient and powerful virtual assistants. This work not only highlights the technical challenges and solutions involved but also provides insights into the practical implementation and potential future directions for this technology.

Literature review

The ubiquitous presence of virtual assistants (VAs) in our daily lives is a testament to the transformative power of artificial intelligence (AI) and natural language processing (NLP). This dissertation delves into the state-of-the-art integration of these technologies within VAs, aiming to pave the way for the development of a novel AI-powered virtual assistant utilizing ESP32 microcontrollers and large language models like ChatGPT.

At the core of intelligent VAs lies the powerful synergy between NLP and AI. NLP empowers VAs to bridge the communication gap between humans and machines. Techniques like semantic analysis, explored by (Tröger, 2019), unlock the meaning behind user queries, allowing VAs to understand the intent and context of user requests. Sentiment analysis, investigated by (Anees, 2020)enables VAs to tailor responses based on user emotions, fostering a more natural and engaging interaction. Entity recognition, as studied by (Keraghel, 2024), further refines this understanding by identifying key elements within user requests, such as locations, people, or objects. This allows VAs to extract crucial information and perform actions accordingly.

AI, on the other hand, injects learning and adaptation capabilities into VAs. By analyzing user interactions and data, AI algorithms personalize responses, as demonstrated in the work of (Chang, 2017). This continuous learning process improves the user experience over time, allowing VAs to better anticipate user needs and preferences. Additionally, AI facilitates engaging conversations through multi-turn dialogues. These dialogues allow VAs to maintain context across conversations, generating responses that are relevant to the ongoing interaction. Proactive recommendations, another benefit powered by AI, enable VAs to anticipate user needs based on learned patterns, as investigated by (Huang, 2006, May). This proactive approach enhances user experience by suggesting actions or information that might be helpful even before users explicitly request them.

The integration of NLP and AI offers significant advantages for VAs. Advanced NLP techniques enable VAs to understand complex queries and respond with increased accuracy. AI fosters natural conversations through context-aware responses, personalized recommendations, and proactive assistance. However, limitations persist that need to be addressed for further advancements.

One limitation is the struggle of current NLP models with nuanced language and sarcasm, as identified by (Choubey, 2020 may). This can lead to misinterpretations and frustrating user experiences. Balancing personalization with user privacy remains another challenge. As highlighted by (Borna Kalhor, 2023), AI algorithms rely on user data for learning, raising concerns about data security and user privacy. Additionally, VAs often lack advanced reasoning capabilities, as discussed by (Chen Liang, 23 Dec 2021). This limits their ability to analyse situations and propose solutions beyond basic task completion.

Addressing these limitations is crucial for unlocking the full potential of VAs. Continued research in NLP, particularly with advancements in Transformers and other architectures, can enhance VAs' ability to grasp the subtleties of human language, including sarcasm and nuanced expressions. Integrating explainable AI techniques can empower users to understand how VAs arrive at their responses, fostering user trust and transparency. Additionally, developing advanced AI algorithms for enhanced reasoning and decision-making can significantly expand VA functionalities, allowing them to analyse situations, propose solutions, and engage in more complex interactions. Finally, research on privacy-preserving NLP and AI techniques, as emphasized by (Xu, 2021, May), is essential to ensure user data remains secure while still allowing for personalization and learning.

Building upon this foundation, this dissertation proposes a novel approach to VA development. It explores the potential of ESP32 microcontrollers, known for their processing power and connectivity, coupled with large language models like ChatGPT, capable of sophisticated language processing. By integrating these technologies, this research aims to develop a user-friendly voice assistant that facilitates seamless interaction through natural language commands. This exploration can offer valuable insights into the potential of this technology and pave the way for a new generation of VAs.

Chapter 1: Introduction to AI-Driven Virtual Assistants in Embedded Systems

1.1 Introduction

In today's rapidly evolving technological landscape, there is an increasing focus on the integration of artificial intelligence across various sectors. Embedded systems, which blend hardware and software to execute specific tasks, have become crucial in many applications such as smart homes, healthcare devices, and industrial automation. These systems frequently require user interfaces for interaction and control. Virtual assistants, leveraging AI to mimic human interactions and offer personalized support, represent a significant advancement in this field. This dissertation explores the development of a virtual assistant utilizing AI within the realm of embedded systems and electronics.

1.2 Virtual Assistants: Transforming Human-Computer Interaction

Virtual assistants are transforming human-computer interaction by replacing traditional input methods with a more intuitive, voice-controlled experience. This shift allows for hands-free operation and simplifies interactions by eliminating the need for complex keyboard or touchscreen inputs (Oulasvirta, 2018, 03). The advent of Conversational AI enables virtual assistants such as Amazon Alexa, Google Assistant, and Microsoft Cortana to comprehend and respond to natural language queries (Hoy, 2018). These assistants offer a variety of functionalities that enhance user experience, including setting reminders, managing schedules, controlling smart devices, and accessing information (Lopatovska, 2019). Beyond household applications, virtual assistants are being utilized in customer service, education, and healthcare, demonstrating potential for improving efficiency in office settings as well (Cowan, 2017). However, they face current limitations, such as challenges in understanding complex language or nuanced requests, as well as concerns related to privacy and security (Bentley, 2018).

1.2.1 Benefits and Applications of Virtual Assistants

Virtual assistants have become increasingly popular due to the wide range of benefits and applications they offer:

- Enhanced Productivity: Virtual assistants can streamline daily tasks by setting reminders, managing schedules, and handling appointments. This allows users to focus on more critical activities and prioritize their time effectively.

- Increased Convenience: Hands-free voice control provides a convenient way to interact with technology. Users can control smart home devices, access information, or play music without needing to pick up a phone or navigate through menus.

- Improved Accessibility: Virtual assistants remove barriers for users with physical limitations or visual impairments. Voice control allows them to interact with technology independently and perform tasks that might be challenging with traditional input methods.

- Personalized Experience: Many virtual assistants can learn user preferences and tailor responses accordingly. This personalization can provide a more intuitive and efficient user experience.

- Entertainment and Information Access: Virtual assistants can be used for entertainment purposes like playing music, reading audiobooks, or controlling smart TVs. They can also access and deliver information from the web, providing users with quick and easy answers to their questions.

These benefits translate into a wide range of applications for virtual assistants beyond personal use:

- Customer Service: Businesses are utilizing virtual assistants to provide 24/7 customer support, answer frequently asked questions, and automate simple tasks, improving customer service efficiency.

- Education: Virtual assistants can be integrated into educational settings to provide personalized learning experiences, answer student questions, and offer language learning support.

- Healthcare: Healthcare providers are exploring the use of virtual assistants for medication reminders, appointment scheduling, and basic health information access, potentially improving patient engagement and self-management.

- Office Automation: Virtual assistants can be used in office environments to manage schedules, control smart meeting room equipment, and even transcribe meetings, streamlining workflows and enhancing office productivity.

By offering these benefits and catering to diverse applications, virtual assistants are transforming the way we interact with technology and manage our daily tasks.

1.2.2 Limitations of Existing Virtual Assistants

While virtual assistants offer a plethora of benefits, they are not without limitations. Here are some key areas for improvement:

- Limited Understanding of Complex Language: Current virtual assistants excel at understanding basic commands and questions but can struggle with complex language structures, sarcasm, or nuanced requests. This can lead to misunderstandings and frustration for users.

- Accuracy and Recognition Rates: Speech recognition technology has made significant strides, but accuracy remains a challenge, especially in noisy environments or with non-standard accents. Frequent misinterpretations can disrupt the user experience.

- Limited Context Awareness: Virtual assistants cannot often understand the context of a conversation. expand more This can lead to irrelevant responses or an inability to follow up on previous requests effectively.

- Privacy Concerns: Voice-controlled interfaces raise concerns over data privacy and security. Users may be hesitant to share sensitive information through voice commands if robust security measures are not implemented.

- Limited Functionality: While functionalities are expanding rapidly, virtual assistants may not yet be able to handle all the tasks users are accustomed to performing with traditional input methods. This can limit their overall usefulness in certain situations.

Despite these limitations, significant research and development are ongoing to address them. Advancements in natural language processing, speaker identification, and context awareness will lead to more robust and user-friendly virtual assistants in the future.

Figure 1 Example of existing VAs in the market today.

1.3 Artificial Intelligence and the Rise of Conversational Interfaces

The rise of virtual assistants and their ability to interact with users through natural language is heavily driven by advancements in Artificial Intelligence (AI), particularly in the field of Natural Language Processing (NLP). AI plays a crucial role in several aspects of conversational interfaces, enhancing their ability to understand and respond to human language effectively.

Natural Language Understanding (NLU): At the core of any conversational interface lies the ability to understand user intent and the meaning behind spoken language. NLP techniques such as machine learning and deep learning algorithms are employed to train AI models on vast amounts of text and speech data. These models learn to identify patterns, analyze syntax, and extract semantic meaning from user utterances. This allows virtual assistants to interpret user requests and respond accordingly (Young, 2018). For instance, when a user asks a virtual assistant to "set a reminder for a meeting tomorrow at 10 AM," the NLU component interprets the intent (setting a reminder) and the details (time and date).

Speech Recognition: Converting spoken language into machine-readable text is crucial for understanding voice commands. Speech recognition models leverage deep learning techniques to identify phonemes (basic units of sound) within a spoken word and translate them into text. Advancements in speech recognition have significantly improved accuracy, especially in controlled environments. These improvements enable virtual assistants to better understand and process voice commands, even in noisy settings or with varied accents (Li, 2015).

Natural Language Generation (NLG): Once user intent is understood, virtual assistants need to formulate a natural language response. NLG techniques draw upon machine learning models trained on vast amounts of text data to generate coherent and grammatically correct responses that mimic human conversation. This capability allows virtual assistants to engage in back-and-forth conversations and provide informative or helpful answers. For example, if a user asks, "What's the weather like today?" the NLG component generates a response like, "Today, the weather is sunny with a high of 75 degrees" (Iqbal, 2022).

Dialogue Management: Maintaining coherence and context within a conversation is essential for a seamless user experience. Dialogue management systems leverage AI techniques to track conversation history, identify user goals, and determine the appropriate response strategy. This ensures virtual assistants stay on topic, follow up on previous requests, and adapt their responses based on the conversation flow. Effective dialogue management contributes to a more natural and engaging user interaction, allowing the assistant to handle multi-turn conversations and complex queries (Serban, 2015).

The integration of these AI techniques empowers virtual assistants to engage in natural language conversations, creating a more intuitive and user-friendly interaction experience. As AI research continues to evolve, we can expect even more sophisticated conversational interfaces capable of handling complex requests and adapting to diverse user preferences. This progress holds significant potential for enhancing the functionality and usability of virtual assistants in various domains, from personal use to professional settings.

1.3.1 Natural Language Processing (NLP) for Understanding User Intent

Natural Language Processing (NLP) is pivotal in enabling virtual assistants to comprehend user intent behind spoken language. NLP encompasses a range of techniques that allow AI models to interpret the meaning and context of human communication effectively:

Machine Learning and Deep Learning Techniques: NLP leverages machine learning algorithms such as Support Vector Machines (SVMs) and deep learning models like Recurrent Neural Networks (RNNs) and Transformers. These models are trained on extensive datasets of text and speech data, enabling them to learn the statistical patterns and relationships between words, phrases, and sentence structures. For instance, Transformers, introduced by (Vaswani, 2017), have revolutionized NLP with their ability to handle long-range dependencies in text efficiently.

Part-of-Speech Tagging (POS Tagging): NLP techniques like POS tagging identify the grammatical function of each word in a sentence, such as whether a word is a noun, verb, or adjective. This helps the model understand the role each word plays in conveying meaning and user intent. For example, POS tagging aids in parsing sentences correctly to distinguish between different types of actions or objects referred to by the user (Huang, 2006, May).

Named Entity Recognition (NER): Recognizing and classifying named entities within a sentence, such as people, locations, or organizations, is crucial for understanding the context of a user's request. NLP models are trained to identify these entities and extract relevant information, which is essential for determining user intent. For example, NER helps virtual assistants like Google Assistant to accurately identify and respond to queries about specific individuals or places (Wu, 2017).

Natural Language Inference (NLI): Techniques like NLI enable AI models to analyze the relationship between sentences and identify entailment, contradiction, or neutrality. This is vital for understanding the underlying meaning and intent behind a user's question or request, even if it is phrased indirectly. For instance, NLI can help a virtual assistant deduce the correct response to a user's statement that implies a request or need (Chen Liang, 23 Dec 2021).

Semantic Role Labeling (SRL): SRL goes beyond identifying the grammatical function of words by also identifying the semantic roles they play within a sentence, such as agent, patient, or instrument. This allows the model to understand the deeper meaning and relationships between words, leading to a more accurate interpretation of user intent. For example, SRL helps in distinguishing who is acting and who is receiving it, which is crucial for understanding complex user commands (He, 2017).

By leveraging these NLP techniques, virtual assistants can analyze user queries, identify keywords and name entities, understand the grammatical structure and semantic relationships between words, and ultimately determine the user's intended action or information need. This comprehensive understanding is fundamental for providing relevant and helpful responses within the context of a natural language conversation, significantly enhancing the user experience with virtual assistants.

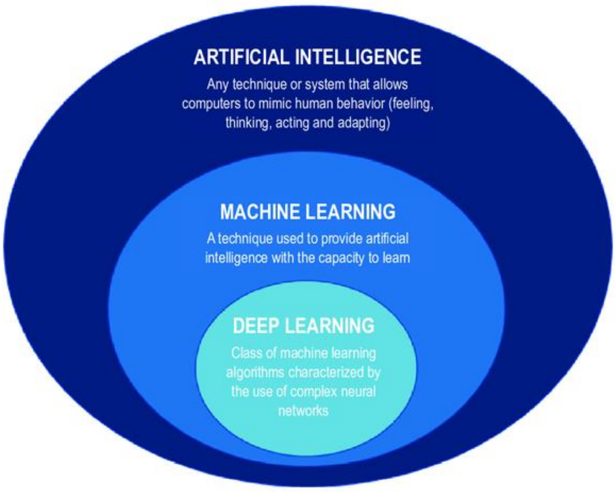

Figure 2 Illustration of the concept of AI

1.4 Embedded Systems: A Platform for Intelligent Edge Devices

The integration of artificial intelligence (AI) into virtual assistants necessitates a robust and efficient hardware platform for deployment. Embedded systems, known for their compact size, low power consumption, and dedicated functionality, provide an optimal solution for implementing intelligent edge devices that power virtual assistants.

This section examines why embedded systems are particularly well-suited for AI-powered virtual assistants and discusses the associated challenges and advancements in this domain.

Embedded systems offer several key advantages for the deployment of virtual assistants. Firstly, their compact size and low power consumption are critical features. These systems are designed to be small and energy-efficient, which is essential for developing virtual assistants that can be embedded into various devices, including smartphones, wearables, smart speakers, and home appliances. The integration of virtual assistants into such devices requires maintaining portability and battery life, both of which are supported by the inherent design of embedded systems (Ye, 2012).

Secondly, embedded systems are characterized by their dedicated functionality. Unlike general-purpose computers, embedded systems are optimized for specific tasks, allowing for hardware and software configurations tailored to particular applications. This optimization is particularly beneficial for executing complex tasks such as voice recognition, natural language processing (NLP), and response generation, which are central to the operation of virtual assistants. The ability to tailor the system precisely to these tasks enhances both performance and reliability (Merone, 2022).

Moreover, embedded systems often prioritize real-time performance, which is crucial for ensuring prompt responses and minimal latency during user interactions. Real-time operation is vital for delivering a seamless user experience when interacting with virtual assistants through voice commands, as users expect immediate and accurate feedback (Buttazzo, 2022). The ability of embedded systems to meet these real-time requirements further underscores their suitability for this application.

Cost-effectiveness is another significant advantage of embedded systems. Due to their specialized nature and the mass production of hardware components, embedded systems tend to be more cost-effective compared to traditional computer systems. This cost efficiency is a critical factor for the widespread adoption of virtual assistants in various consumer electronics and smart devices, making advanced AI technologies accessible to a broader audience (da Silva, 2019).

Despite these advantages, there are challenges associated with using embedded systems for AI-powered virtual assistants. One major challenge is the limited