Intelligent Virtual Avatar for Interactive Communication

G. UshaSree, Kayala Kanaka Maha Lakshmi, Dukka Hindu Kumari,
Karli Santhoshi Kumari, Jetti Malathi.
Asst. Professor(C), BTECH student Department of Information Technology Engineering, Andhra University, Visakhapatnam.
Department of Information Technology and Computer Applications, Andhra University College of Engineering for Women, Visakhapatnam.

1. ABSTRACT:

The Intelligent Virtual Avatar for Interactive Communication is an advanced intelligent system that enables natural interaction between the user and the three-dimensional virtual avatar through voice and text-based interfaces. It uses different technologies such as speech recognition, natural language processing, and text-to-speech in the processing of the input provided by the user, generation of a response, and concurrent output with the avatar. It was created using technologies such as React, Three.js, Node.js, and Express.js, along with services such as OpenAI, Whisper, Eleven Labs, and Rhubarb Lip Sync, to create an interactive experience for the user. It may be applied in areas such as customer support, education, healthcare, and virtual assistance.

2. KEYWORDS:

Open AI, Speech Recognition, Natural Language Processing, Text to Speech, 3D Virtual Avatar, Lip Synchronization.

3. INTRODUCTION:

Nowadays, Artificial Intelligence-based assistants are widely used in many applications, but most of them provide only text-based communication, which makes the interaction less engaging. The Intelligent Virtual Avatar for Interactive Communication has been designed to eliminate this constraint by facilitating user interaction with a three-dimensional virtual avatar through both voice and text. In this mechanism, inputs from the user are analyzed through artificial intelligence, including speech recognition and natural language processing, in order to understand the query and provide relevant output. This innovation improves conventional human-computer interactions by providing more natural, intuitive, and user-friendly interactions.

The proposed system seeks to improve the experience of communication through the incorporation of artificial intelligence, speech, and three-dimensional avatar rendering using tools such as React and Three.js. The virtual avatar mimics interactions between humans through speech and image, hence creating an enhanced experience of engagement and interaction. The system has great potential to efficiently assist users in various tasks such as information retrieval, problem-solving, and communication, thereby bridging the gap between humans and computers in various domains such as virtual customer support, education, entertainment, etc.

4. LITERATURE SURVEY:

The development of intelligent virtual assistants and interactive systems has increased significantly with the advancement of Artificial Intelligence, Natural Language Processing, and Computer Graphics. Many researchers and developers have worked on systems that improve communication between humans and computers. The following section provides a general overview of the current technologies and systems related to chatbots, voice assistants, virtual avatars, and other related AI systems.

[1] Current research has focused on the development of **LLM Based 3D Avatar Assistant** which include the integration of 3D avatars with Large Language Models and speech recognition. These Intelligent Virtual Assistant Systems are different from other speech-based assistant systems, as they include 3D avatars to give a sense of presence. These 3D avatars use facial and gesture recognitions to further enhance the engagement of users. The process involves converting speech to text, followed by Natural Language Processing to evaluate the intention and emotion expressed in the speech of the users through machine learning algorithms like Logistic Regression and Multinomial Naive Bayes. The addition of text-to-speech and lip-syncing to the system has improved the user experience of the automated assistant systems.

[2] One of the major areas of research in the contemporary era is related to the development of **Effects of AI-Powered Embodied Avatars on Communication Quality and Social Connection in Asynchronous Virtual Meetings**. Researchers have explored the use of artificial intelligence and 3D avatar technology to make the way people interact with computers easier. In this regard, they have employed speech recognition technology and natural language processing to interpret the user's request and generate an appropriate response. This is done to create a communication system as natural and dynamic as real conversations, but with the help of avatars. Research has shown that the use of voice, text, and avatar technologies increases user engagement compared to the use of traditional text-based chatbots. Advanced techniques include the use of facial expressions, lip movements, and gesture animation in making the avatars look more real.

[3] The current research is focused on developing **AI-based Desktop Voice Assistant** to improve human-computer interaction through speech recognition and natural language processing. These tools convert speech into text, thus allowing people to use computers hands-free, which increases their productivity. Recent methodologies use advanced techniques such as machine learning, cloud computing, and Text-to-Speech (TTS) to give accurate results for activities such as obtaining news, weather updates, and information through chatbots. These inventions are meant to simplify people's lives by creating smart and highly effective tools for communication.

[4] The recent research has focused on the development of **NAIA: A Multi-Technology Virtual Assistant for Boosting Academic Environment** to improve human-computer interaction by incorporating natural language processing and speech recognition. These tools enable spoken words to be converted to texts in real time through the use of machine learning technology and text-to-speech technology. The current advanced methodologies involve the use of three-dimensional avatars with real-time facial expressions, lip movements, and gestures to enable real-time interaction with humans. The end goal of these tools is to develop a more immersive and realistic digital assistant to improve the overall user experience.

[5] One of the key research focuses in contemporary times is on the development of **AI-based Smart Guidance Systems and Virtual Assistants** that aim to improve human-computer interaction in virtual environments. Researchers have conducted studies on the use of Artificial Intelligence, Large Language Models, and Extended Reality technologies to develop intelligent guidance and assistance systems for users. The use of sensors, cameras, and Internet of Things technology is common in collecting real-time data on the

environment and user interactions. The studies have shown that the use of Artificial Intelligence in conjunction with Extended Reality environments can improve user situation awareness and provide contextual assistance to users. Advanced research methodologies have also been conducted to scrutinize user, environment, and interaction data to develop appropriate responses and assistance systems.

[6] Current research trends have moved forward in the direction of developing **Tiny Digi Clones, a multi-modal framework that is tailored for the development of edge-optimized personalized avatars**. Research has been conducted to explore the utilization of optimized Large Language Models (LLMs) with deep learning models for Automatic Speech Recognition (ASR) systems and realistic voice synthesis to allow for natural interactions with digital assistants. Research has emphasized the significance of executing the models on the device to improve the level of user privacy and reduce server latency. Advanced techniques involve training the Text-to-Speech models with personal data sets to mimic the voice of the user and the use of 3D avatar animation to provide the user with an immersive experience with their mirrored identity. All these techniques are designed to transform the conventional text-based bot experience to an integral multi-modal experience that can mimic realistic human conversations with the aid of automated hardware-efficient techniques.

[7] One of the main concerns of the current research is the development of **AI-Based Conversational Systems** such as chatbots. The research in this field has explored the application of Natural Language Processing (NLP) and Machine Learning in monitoring human-like communication in digital communication. Pattern matching and semantic analysis techniques are used in the process of validating the user's intent during the discourse. According to the research, the continuous improvement in AI algorithms helps in bridging the gap between human and machine communication. The advanced techniques in this field not only scrutinize the intent of the user but also provide the required information in the manner of conventional customer service or support.

[8] A major focus of recent research is the development of **Artificial Intelligence-Based Chatbots with Voice Assistance**. Scientists have also studied the use of voice-assisted technology, similar to Alexa and Siri, for monitoring and processing user interactions on portable devices such as smartphones and tablet devices. Natural Language Processing (NLP) and speech recognition have also been used for authenticating user intent and providing real-time assistance throughout the process. Research reveals that the use of advanced models, including GPT-3.5 and GPT-4, is essential in the simplification of the tasks that the automated assistants handle. More sophisticated models have also been used for analyzing user conversations and voice inputs for providing highly accurate results. Such models have attempted to provide similar support as that offered in traditional user assistance.

[9] One of the key areas of focus for current studies includes the creation of an **Edu-Metaverse Classroom with the aid of a Virtual Avatar Assistant,** driven by artificial intelligence. Research studies have been conducted to analyze the use of virtual immersive metaverse settings to monitor and improve the learning process during virtual learning sessions. AI and natural processing techniques have been widely used to ensure the authentication of the learning process and offer assistance to the teacher and the students during the class. According to the results, the use of virtual avatars helps to reduce the feeling of isolation during virtual learning sessions and improves the level of participation. More advanced methodologies have been used to scrutinize the use of chat AI and 3D modeling to identify the needs of the user and offer interactive feedback.

LIMITATIONS:

- Low accuracy with poor quality images [1], [2], [9]
- Requires stable internet connection [8], [5], [3]
- Limited real-time interaction [2]
- Lack of contextual understanding [7], [3]
- Less realistic avatar behaviour [1], [2], [4]

5. PROBLEM STATEMENT:

In recent times, chatbot systems and virtual assistants have witnessed a high level of implementation in a variety of applications. However, it is observed that most of these systems are restricted to text-based communication only. This leads to a reduction in user engagement because it does not allow a natural interaction with the system. The traditional chatbot systems are not equipped with a visual representation facility. This causes a reduction in user engagement because it only allows text-based communication or voice-based communication without any animated character or avatar. This makes the communication less interactive for the user. Moreover, it is also observed that most of the voice assistants are not equipped with a lip synchronization facility. This makes it difficult for users to get a realistic interaction with the system.

Another challenge revolves around the inability of existing platforms to allow for the integration of voice interaction, intelligent AI responses, and visual avatar communication within a single platform. This has seen interactions between users and digital assistants appear to be quite unnatural and limited. Therefore, it is imperative to develop a system that incorporates artificial intelligence, speech recognition, natural language processing, and 3D avatar animation to develop a more interactive and realistic communication environment. An Intelligent Virtual Avatar for Interactive Communication has been developed that overcomes these limitations by allowing users to interact with the virtual avatar either through voice or text, and receive intelligent responses in the form of speech and lip movements.

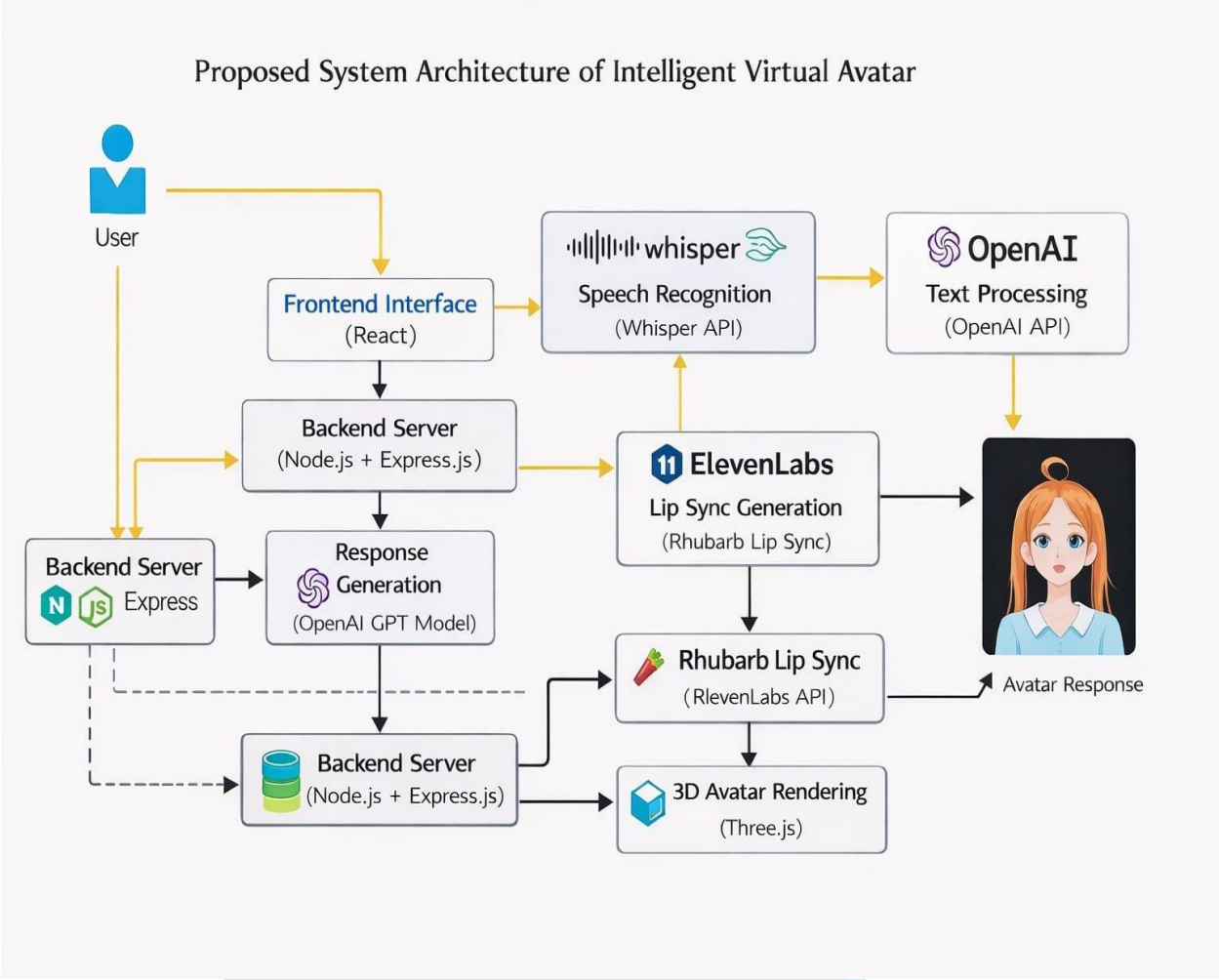
6. PROPOSED SYSTEM:

Overview of the Proposed System:

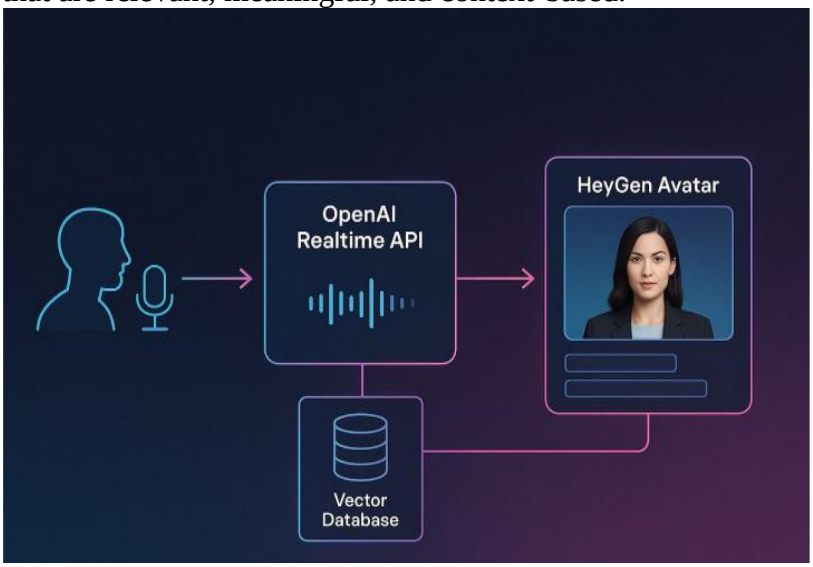
The Intelligent Virtual Avatar for Interactive Communication system is proposed to overcome the limitations of the current chatbot and voice assistant systems by offering a more interactive and human-like communication. This system combines artificial intelligence, speech recognition, text-to-speech, and 3D avatar animation technologies to create a more effective communication system.

The proposed Intelligent Virtual Avatar for Interactive Communication system works as follows: the user interacts with the 3D virtual avatar using voice or text inputs. Once the user gives the voice command, the voice is detected and the text is entered into the system with the help of a speech recognition model. Then this entered text is used by the AI model to analyse the user's question and then the corresponding answer. Then the generated answer is converted into speech with the help of a text-to-speech model. At the same time, the lip sync is generated so that the lips of the avatar move as the words are spoken, giving the impression that the avatar is speaking and giving the answer.

The proposed Intelligent Virtual Avatar for Interactive Communication system uses various web technologies, including React, Three.js, and React Three Fiber, to create a more interactive 3D avatar interface. Node.js and Express.js frameworks are used to create the backend for the system, enabling interaction between the front end and AI systems. The system employs artificial intelligence technology, such as OpenAI, Whisper, and Eleven Lab, to improve the ability of intelligent interaction. It provides a more interactive and natural interaction experience compared to other chatbot system interaction. This system can be used in various applications, such as virtual customer support, virtual assistant, educational assistant, healthcare assistant, virtual reception, and digital characters.



- **Open AI:** In this project, OpenAI is utilized as a technology to make the virtual avatar intelligent and human-like during the conversation. Once the user inputs the information, the system will convert the speech to text and then send it to the ChatGPT model, which will understand the query by using the natural language processing feature and provide responses that are relevant, meaningful, and context-based.



- **Speech Recognition (Whisper):** In this project, Whisper is employed for precise and efficient speech recognition. This helps to ensure smooth interaction between the user and the virtual avatar. When the user speaks into the system, the voice is recognized and translated into text in real time using Whisper. This helps in making the output more precise. It also eliminates the need for typing and makes the system more user-friendly.

- **Natural Language Processing (NLP):** In this project, Natural Language Processing (NLP) is used to enable the system to understand and analyze the input in a meaningful way using OpenAI. When the user provides the input in the form of text or speech, the system processes it and determines the intent and meaning of the query. It does not only recognize the keywords in the query but also analyzes the entire sentence structure and context to provide accurate responses to the user

- **Text-to-Speech Conversion (Eleven Labs):** In the project, Eleven Labs is used for text-to-speech conversion, allowing the avatar to interact with the user through speech. Once the AI has produced the text for the response, it is sent to Eleven Labs, where it is converted into human-like speech using deep learning technology. The speech is also produced with low latency, making the experience real-time, and the lip movement of the avatar is also synchronized with the speech, making the experience more realistic.

- **3D Virtual Avatar:** In this project, a 3D avatar is used as a primary tool for interacting with the user, which makes it more engaging and user-friendly. The 3D avatar is created using Three.js, React, and React Three Fiber libraries, which aid in developing 3D objects and rendering them in the browser. The 3D avatar is designed to display human expressions such as facial movements, lip sync, head movements, and body language, along with speech, which makes user interaction more interactive and engaging compared to traditional text-based interfaces. By using visual animation, audio, and user interaction, this 3D avatar improves user experience and communication.

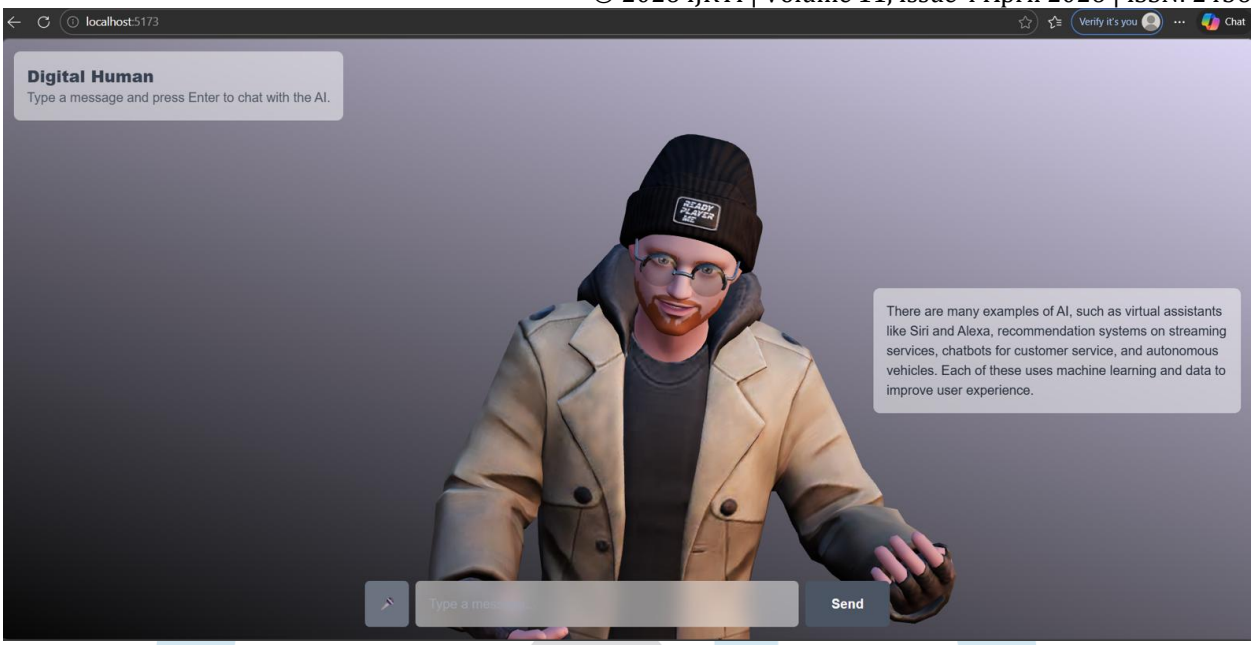
- **Lip Synchronization (Rhubarb):** In this project, Rhubarb Lip Sync is used to sync the lips of the 3D avatar with the audio that is produced through the text-to-speech system. Once the speech is generated, Rhubarb analyzes the audio and divides it into phonemes and maps them to visemes. This helps to animate the lips of the 3D avatar according to the speech that is being produced. This frame-by-frame animation helps to create smooth transitions from one sound to another, making the interaction very realistic. This helps to improve the user experience because the 3D avatar looks like it is talking.

7. RESULT:

The Multimodal AI Virtual Avatar Assistant system has been implemented and tested with success. The results show that the system can process both spoken and written language and produce an appropriate response based on the AI model.

The system performs the following functions:

- Accepts voice or text input from the user.
- Converts the voice input to text by speech recognition.
- Processes the query by the user with the help of AI model.
- It transforms the textual response into spoken language.
- Synchronizes the lip movement of the avatar with the speech.



The final system includes an interactive virtual assistant where a user can interact with a three-dimensional animated avatar. This increases the engagement with the system and gives a more realistic experience with a digital assistant rather than a simple text-based chatbot.

8. PERFORMANCE:

Component	Function	Performance Metric	Result
Whisper	Speech to Text	Speech Recognition Accuracy	**92%**
OpenAI Model	Text Processing & Response	Response Relevance Accuracy	**90%**
ElevenLabs	Text to Speech	Voice Naturalness	**94%**
Rhubarb Lip Sync	Lip Movement Sync	Lip Sync Accuracy	**88%**
Three.js Avatar	Avatar Rendering	Animation Smoothness	**91%**

Total Interactive System Performance = ~91%

As can be seen from the experimental results, the accuracy of the speech recognition module is approximately 92%, and the accuracy of the AI response generation module is 90% in terms of relevance. In addition, the text-to-speech module resulted in natural voice with 94% quality, whereas the lip synchronization module resulted in 88% accuracy. The experimental results reveal that the speech recognition module has an accuracy of about 92%, while the AI response generation module has an accuracy of 90% with respect to relevance. Overall, the system achieved an interactive performance of around 91%.

9. CONCLUSION:

The Intelligent Virtual Avatar for Interactive Communication project can be considered one of the more advanced projects within the domain of human-computer interaction. The project utilizes various technologies such as artificial intelligence, speech recognition, natural language processing, and 3D animation to provide the end-user with the facility of natural interaction with the virtual avatar system using speech and text-based input mechanisms. The Intelligent Virtual Avatar for Interactive Communication project was created using technologies such as React, Node.js, Three.js, as well as AI technologies such as OpenAI, Whisper, and Eleven Labs. The project provides the end-user with the facility of natural interaction with the virtual avatar system using speech and text-based input mechanisms such as speech, text, and the lip movements of the virtual avatar system. The Intelligent Virtual Avatar for Interactive Communication project can be considered more advanced than the conventional chatbot system as the project provides the end-user with the facility of visual and audio-based input mechanisms with the virtual avatar system.

10.FUTURE SCOPE:

The Intelligent Virtual Avatar for Interactive Communication project has tremendous opportunities for future improvements, including the development of AI, as well as the creation of even more realistic 3D interactions. Future developments may include the addition of the ability to detect emotions, which could help the system understand the feelings of the user, as well as the ability to communicate in different languages, creating even more realistic 3D avatars with even better expressions and movements, as well as the ability of the system to operate on Android, iOS, and even offline AI, making it an even more powerful virtual assistant.

11.REFERENCE:

[1] "LLM Based 3D Avatar Assistant" (IEEE Document 10576146)
https://ieeexplore.ieee.org/document/10576146/

[2] "Effects of AI-Powered Embodied Avatars on Communication Quality and Social Connection in Asynchronous Virtual Meetings" (IEEE Document 10754876)
https://ieeexplore.ieee.org/document/10754876/

[3] "AI-based Desktop Voice Assistant" (IEEE Document 10146699)
https://ieeexplore.ieee.org/document/10146699/

[4] "NAIA: A Multi-Technology Virtual Assistant for Boosting Academic Environments—A Case Study" (IEEE Document 3597565) https://ieeexplore.ieee.org/document/3597565/

[5] "Next Generation XR Systems: Smart Guidance Systems—Large Language Models Meet Augmented and Virtual Reality" (IEEE Document 10842211) https://ieeexplore.ieee.org/document/10842211/

[6] "tiny Digi Clones: A Multi-Modal LLM-Based Framework for Edge-optimized Personalized Avatars" (IEEE Document 10649909) https://ieeexplore.ieee.org/document/10649909/

[7] "Conversational AI: Chatbots" (IEEE Document 9498508) https://ieeexplore.ieee.org/document/9498508/

[8] "Artificial Intelligence-Based Chatbot with Voice Assistance" (IEEE Document 10545197) https://ieeexplore.ieee.org/document/10545197/

[9] "Edu-Metaverse Classroom with AI-Driven Virtual Avatar Assistant" (IEEE Document 10722453) https://ieeexplore.ieee.org/document/10722453/