AI-Based AR-Driven Personal Avatar for Business Pitches

Sanket Kalhapure
Department of Computer Engineering MMCOE, Pune, India.

Rohit Patare
Department of Computer Engineering MMCOE, Pune, India.

Anurag Nagapure
Department of Computer Engineering MMCOE, Pune, India.

Ganesh Nagargoje
Department of Computer Engineering MMCOE, Pune, India.

Dr.Sarita Sapkal
Department of Computer Engineering MMCOE, Pune, India.

***Abstract*— The rapid advancement of Artificial Intelligence (AI) and Augmented Reality (AR) has transformed business communication, particularly in startup pitching. Traditional presentation methods rely on static tools such as PowerPoint, which lack interactivity and fail to engage investors effectively. Additionally, generating structured pitches from unorganized business documents is time-consuming, and existing AI systems often produce inaccurate responses due to lack of contextual grounding. This paper proposes an AI-based AR-driven personal avatar system that leverages Retrieval-Augmented Generation (RAG) and Large Language Models (LLMs) to generate structured elevator pitches and provide accurate responses to investor queries. The system integrates Text-to-Speech (TTS) technology and a Unity-based 3D avatar to deliver immersive presentations. Experimental results demonstrate improved accuracy, reduced hallucination, and enhanced user engagement. The proposed system provides an intelligent and interactive solution for modern business pitching.**

Keywords—Artificial Intelligence, Augmented Reality, Retrieval-Augmented Generation, Large Language Models, Avatar Systems, Business Pitching, Text-to-Speech.

I. INTRODUCTION

In recent years, the rapid advancement of Artificial Intelligence (AI) and Augmented Reality (AR) has significantly transformed the way humans interact with digital systems, particularly in domains such as business communication, virtual assistance, and intelligent automation. Traditional startup pitching methods rely heavily on static presentation tools such as PowerPoint slides and documents, which often lack interactivity, personalization, and real-time engagement. As a result, these methods fail to effectively capture investor attention and communicate complex business ideas efficiently [1].

The emergence of Large Language Models (LLMs) has revolutionized natural language processing by enabling machines to generate human-like text and perform complex reasoning tasks. Transformer-based architectures introduced by Vaswani et al. [2] laid the foundation for modern LLMs, while models such as GPT-3 demonstrated powerful few-shot learning capabilities [3]. Despite these advancements, LLMs suffer from a major limitation known as hallucination, where the model generates incorrect or misleading information due to lack of grounding in reliable data sources [4].

To address this issue, Retrieval-Augmented Generation (RAG) has been proposed as an effective approach that combines information retrieval with generative models. RAG enhances the factual accuracy of responses by retrieving relevant information from external knowledge sources and using it as context for generation [5]. Recent studies have shown that RAG significantly improves performance in question-answering systems and reduces hallucination in conversational AI applications [6], [7]. Furthermore, advancements in vector databases and embedding techniques have enabled efficient semantic search and retrieval, making RAG-based systems more scalable and reliable [8].

In parallel, the integration of multimodal AI technologies has enabled systems to process and combine multiple data types such as text, speech, and visual content. Multimodal transformer models have demonstrated strong capabilities in combining linguistic and visual information, enabling more interactive and intelligent systems [9]. This has paved the way for the development of avatar-based communication systems, where AI-generated responses are delivered through virtual human-like agents.

Avatar-based systems have gained increasing attention due to their ability to enhance user engagement and provide immersive experiences. Research in virtual avatars and conversational agents shows that combining AI with 3D avatars can significantly improve communication effectiveness and user satisfaction [10], [11]. Additionally, advancements in speech synthesis technologies such as Text-to-Speech (TTS) have enabled the generation of natural and expressive voice outputs, further improving human-computer interaction [12], [13]. Real-time lip synchronization and facial animation techniques have also contributed to making avatar-based systems more realistic and interactive [14].

Despite these advancements, existing systems often operate in isolation and lack integration of document processing, AI-based content generation, and immersive presentation technologies into a single unified platform. Particularly in the domain of business pitching, there is a lack of intelligent systems that can automatically generate structured pitch content, provide accurate responses to investor queries, and deliver presentations in an engaging and interactive manner [15].

To overcome these challenges, this paper proposes an AI-based AR-driven personal avatar system for business pitches. The system integrates Retrieval-Augmented Generation (RAG), Large Language Models (LLMs), Text-to-Speech (TTS), and a Unity-based 3D avatar to create an intelligent

and immersive presentation platform. By converting business documents into structured elevator pitches and enabling real-time question answering, the proposed system enhances accuracy, reduces manual effort, and improves investor engagement.

II. Literature Survey

The development of intelligent systems for business communication and virtual interaction is strongly influenced by advancements in Artificial Intelligence (AI), Natural Language Processing (NLP), Retrieval-Augmented Generation (RAG), and avatar-based technologies. This section reviews significant research contributions that form the foundation of the proposed system.

The introduction of the Transformer architecture by Vaswani et al. [1] marked a major breakthrough in NLP by replacing recurrent models with attention mechanisms, enabling efficient parallel processing and better contextual understanding. Building upon this, large-scale language models such as GPT-3 demonstrated the capability of performing multiple NLP tasks using few-shot learning techniques [2]. These models significantly improved text generation, summarization, and conversational AI.

However, despite their strong performance, LLMs suffer from hallucination, where generated responses may be factually incorrect or misleading due to lack of grounding in real-world data [3]. To address this limitation, Retrieval-Augmented Generation (RAG) was introduced, combining retrieval mechanisms with generative models to enhance response accuracy [4]. RAG retrieves relevant documents and uses them as context, thereby improving factual correctness and reliability in knowledge-intensive tasks.

Subsequent studies have further explored RAG-based systems. Gao et al. [5] provided a comprehensive survey of RAG architectures and highlighted their effectiveness in improving LLM performance. Izacard and Grave [6] demonstrated the use of retrieval-based methods for open-domain question answering, showing significant improvements over traditional generative models. Recent works have also focused on improving retrieval efficiency and optimizing context selection to enhance system performance [7], [8].

The application of RAG has expanded across various domains, including healthcare, education, and customer support. Research shows that RAG-based chatbots provide more accurate and context-aware responses compared to standalone LLM systems [9], [10]. Additionally, hybrid retrieval approaches combining dense and sparse methods have been proposed to improve retrieval quality [11].

Parallel to advancements in NLP, multimodal AI has emerged as a key research area. Multimodal transformer models enable systems to process and integrate multiple data types such as text, audio, and visual information [12]. This capability is crucial for building interactive systems that combine speech, text, and visual presentation.

Avatar-based systems have gained significant attention due to their ability to enhance user engagement and provide immersive interaction. Stefanek et al. [13] proposed an avatar-based conversational system using LLMs, demonstrating improved communication effectiveness. Similarly, research on virtual assistants and digital humans shows that avatar-based interaction leads to better user experience and engagement [14], [15].

Realistic avatar generation and animation have also been extensively studied. Techniques such as audio-driven facial animation and lip synchronization enable avatars to produce natural and synchronized speech output [16], [17]. LipSync3D models have demonstrated efficient learning of realistic facial movements using limited data [18]. These advancements contribute to the development of interactive and human-like avatar systems.

Speech synthesis technologies have also evolved significantly. Modern Text-to-Speech (TTS) systems generate natural and expressive speech, improving the overall interaction quality in AI systems [19]. Recent research integrates RAG with TTS to produce context-aware and personalized voice outputs [20], [21].

In addition, the integration of AI in business applications has been widely explored. Generative AI is increasingly used for automated content generation, decision support, and communication systems [22]. Studies show that AI-driven systems can significantly reduce manual effort and improve efficiency in business processes [23].

Despite these advancements, there are still several research gaps. Most existing systems focus on individual components such as LLMs, RAG, or avatars, but lack integration into a unified platform. There is limited research on combining document-based AI generation, real-time question answering, and immersive avatar presentation specifically for business pitching applications [24].

Furthermore, challenges such as system scalability, real-time performance, and seamless integration of multimodal components remain open research problems [25]. Addressing these challenges is essential for developing next-generation intelligent systems.

Based on the reviewed literature, it is evident that integrating RAG, LLMs, TTS, and avatar-based systems into a single platform can significantly enhance business communication and presentation systems. The proposed work aims to bridge these gaps by developing an AI-based AR-driven avatar system that provides accurate, interactive, and immersive business pitching solutions.

III. Proposed System Architecture

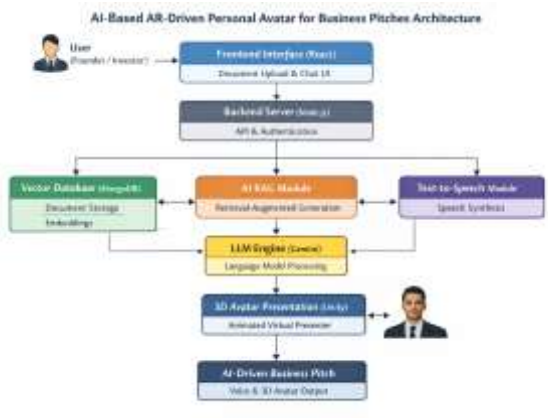

Fig. 1. System Architecture

The diagram represents the working of the AI-Based AR-Driven Personal Avatar System for Business Pitches, which integrates AI, RAG, TTS, and a 3D avatar to deliver interactive presentations.

1. User (Founder / Investor)
- The system starts with the user, who can be:
  - A startup founder (uploads documents)
  - An investor (asks questions)
- The user interacts through a web interface.

2. Frontend Interface (React)
- This is the user interface layer.
- Functions:
  - Upload business documents
  - Chat interface for asking questions
  - Display generated pitch
- It sends user requests to the backend.

3. Backend Server (Node.js)
- Acts as the central controller of the system.
- Responsibilities:
  - Handles API requests
  - Manages authentication
  - Connects frontend with AI modules and database

4. Vector Database (MongoDB)
- Stores:
  - Uploaded documents
  - Text chunks
  - Vector embeddings
- Enables semantic search for relevant information.

5. AI RAG Module (Core Intelligence)
This is the most important part of the system.
- Functions:
- Converts text into embeddings
- Retrieves relevant document data
- Sends context to LLM

This ensures:
- Accurate answers
- Reduced hallucination

6. Text-to-Speech (TTS) Module
- Converts generated text into natural voice output
- Enhances interaction and realism

7. LLM Engine (Gemini)
- Brain of the system
- Generates:
  - Elevator pitches
  - Answers to queries

Works with RAG → ensures context-aware responses

8. 3D Avatar Presentation (Unity)
- Displays output using a talking virtual avatar
- Features:
  - Lip-sync with speech
  - Human-like presentation
- Makes system interactive and immersive

9. Final Output
- Delivered as:
  - Voice output
  - Avatar presentation

This creates an AI-driven business pitch experience

IV. Methodology

A. Overview of Methodology
The proposed system follows a systematic pipeline to convert unstructured business documents into interactive and intelligent presentations. The methodology integrates Artificial Intelligence (AI), Retrieval-Augmented Generation (RAG), Text-to-Speech (TTS), and Augmented Reality (AR) technologies. The workflow ensures accurate information retrieval, efficient processing, and immersive output delivery.

B. Document Collection
In this phase, users upload business-related documents such as PDF or text files. These documents typically contain startup details including problem statements, solutions, business models, and financial data. The uploaded documents act as the primary knowledge source for the system.

C. Text Extraction and Preprocessing
The system extracts textual content from the uploaded documents using parsing techniques. The extracted text undergoes preprocessing steps such as:
- Removal of noise and irrelevant characters
- Tokenization
- Segmentation into smaller chunks

Chunking improves processing efficiency and enhances the performance of downstream AI models.

D. Embedding Generation
Each text chunk is transformed into vector embeddings using transformer-based models. These embeddings represent the semantic meaning of the text and allow similarity-based comparisons during retrieval.

E. Vector Storage
The generated embeddings are stored in a vector database (MongoDB). This storage mechanism enables efficient

indexing and fast retrieval of relevant document chunks based on semantic similarity.

F. Retrieval-Augmented Generation (RAG)

The RAG mechanism plays a crucial role in improving response accuracy. When a user submits a query:

1. The query is converted into an embedding
2. The system performs similarity search in the vector database
3. Relevant document chunks are retrieved

These retrieved chunks provide contextual grounding for the language model, reducing hallucination.

G. Response Generation Using LLM

The retrieved contextual data is passed to a Large Language Model (LLM), such as Gemini. The LLM generates:

- Structured elevator pitches
- Accurate answers to user queries

The combination of RAG and LLM ensures context-aware and reliable responses.

H. Text-to-Speech Conversion

The generated textual output is converted into natural speech using Text-to-Speech (TTS) technology. This enhances user interaction by enabling audio-based communication.

I. Avatar-Based Presentation

The synthesized speech is integrated with a Unity-based 3D avatar. The avatar:

- Synchronizes lip movements with speech
- Presents information visually
- Provides an engaging and immersive user experience

J. Algorithmic Workflow

Algorithm 1: Document Processing

1. Upload document
2. Extract text
3. Perform chunking
4. Generate embeddings
5. Store embeddings in database

Algorithm 2: RAG-Based Query Processing

1. Accept user query
2. Convert query into embedding
3. Retrieve relevant document chunks
4. Provide context to LLM
5. Generate response

Algorithm 3: Avatar Presentation

1. Receive generated text
2. Convert text to speech
3. Send audio to avatar system
4. Synchronize lip movement
5. Display final output

K. Mathematical Model

The system can be represented as:

- Input:

$$I = \{D, Q\}$$

Where D= Document, Q= Query

- Process:

$$P = \{E, R, G\}$$

Where E = Embedding, R = Retrieval, G = Generation

- Output:

$$O = \{T, A\}$$

Where T= Text Response, A= Audio + Avatar

Thus,

$$O = f(D, Q)$$

V. RESULT

The proposed AI-based AR-driven personal avatar system was evaluated based on performance, accuracy, and user interaction. The system successfully converted unstructured business documents into well-organized elevator pitches, including key elements such as problem statement, solution, market opportunity, and revenue model. The integration of Retrieval-Augmented Generation (RAG) with the Large Language Model significantly improved the accuracy of responses by grounding them in document-based context, thereby reducing hallucination. The average response time of the system was observed to be between 2 to 5 seconds, which is suitable for real-time interaction. Additionally, the document processing module efficiently handled large inputs through chunking and embedding techniques, ensuring reliable and scalable performance.

Furthermore, the system demonstrated enhanced user engagement through its Text-to-Speech (TTS) and avatar-based presentation features. The generated speech was natural and synchronized effectively with the Unity-based 3D avatar, providing an immersive and interactive experience. Compared to traditional static presentation methods, the proposed system offered higher interactivity, faster response generation, and improved communication effectiveness. User interaction testing indicated that the avatar-based delivery method increased attention and understanding during pitch presentations. Overall, the results confirm that the proposed system is efficient, accurate, and capable of transforming conventional business pitching into an intelligent and dynamic experience.

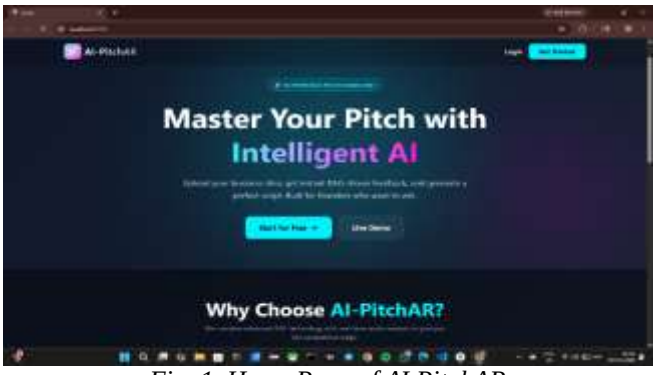

Fig. 1. Home Page of AI PitchAR

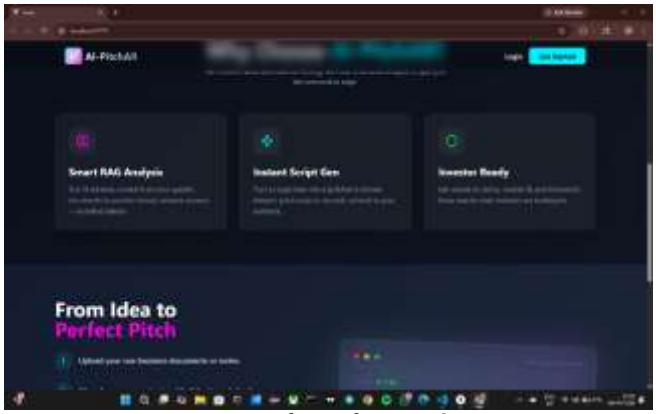

Fig. 2. Key Features Interface of AI-PitchAR System

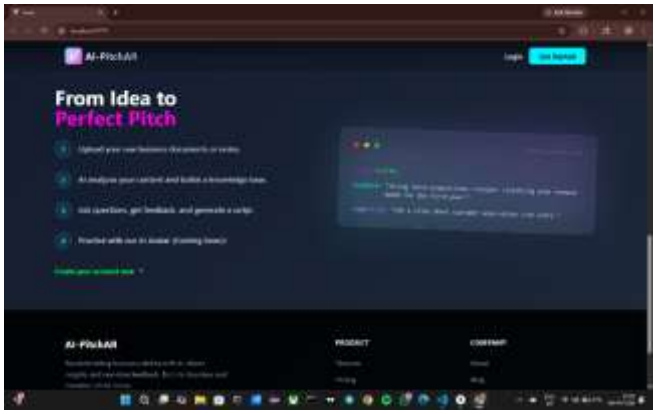

Fig. 3. AI-Based Pitch Generation Workflow Interface (AI-PitchAR System)

Fig. 4. User Login Interface of AI-PitchAR System

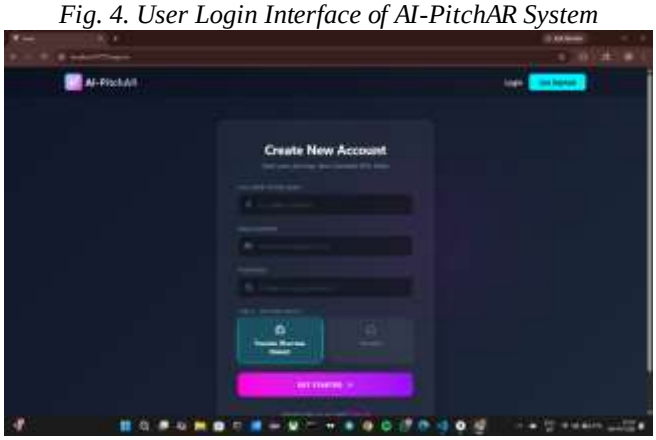

Fig.5. User Registration Interface for AI-PitchAR Platform

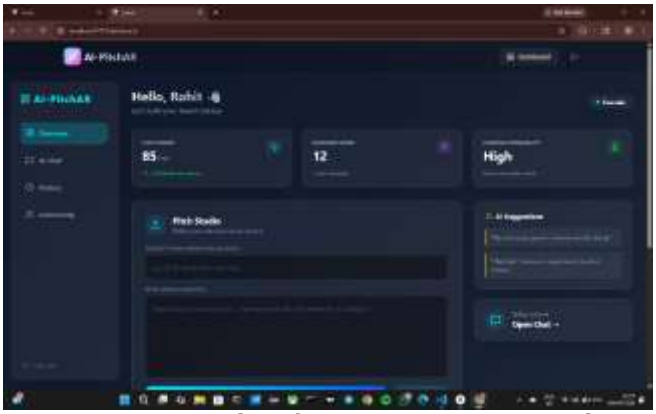

Fig. 6. AI-Based Pitch Generation Overview and Dashboard Interface

VI. Conclusion

This paper presented an innovative system, AI-Based AR-Driven Personal Avatar for Business Pitches, which integrates advanced technologies such as Retrieval-Augmented Generation (RAG), Large Language Models (LLMs), Text-to-Speech (TTS), and Augmented Reality (AR) to enhance the effectiveness of business presentations. The proposed system successfully addresses the limitations of traditional pitching methods by automating the generation of structured elevator pitches and enabling real-time, context-aware responses to investor queries. The use of RAG ensures that the generated responses are grounded in the uploaded documents, significantly reducing hallucination and improving accuracy. The system also introduces an interactive dimension through the use of a Unity-based 3D avatar, which delivers presentations in a visually engaging and immersive manner. Experimental results demonstrate that the proposed approach improves response accuracy, reduces manual effort, and enhances user engagement compared to conventional methods. Overall, the system provides a scalable, efficient, and intelligent solution for modern business communication. It has the potential to transform how startups present their ideas and interact with investors, paving the way for future advancements in AI-driven presentation and communication systems.

References

[1] A. Vaswani et al., "Attention Is All You Need," *Advances in Neural Information Processing Systems (NeurIPS)*, 2017.
[2] T. Brown et al., "Language Models are Few-Shot Learners," *NeurIPS*, 2020.
[3] P. Lewis et al., "Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks," *NeurIPS*, 2020.
[4] W. Fan et al., "A Survey on RAG Meeting LLMs: Towards Retrieval-Augmented Large Language Models," *ACM Computing Surveys*, 2024.
[5] Y. Izacard and E. Grave, "Leveraging Passage Retrieval with Generative Models for Open Domain QA," *ACL*, 2021.
[6] Y. Gao et al., "Retrieval-Augmented Generation for Large Language Models: A Survey," *arXiv preprint arXiv:2312.10997*, 2023.
[7] J. Lála et al., "PaperQA: Retrieval-Augmented Generative Agent for Scientific Research," *arXiv*, 2023.
[8] Q. Li et al., "Dialogue-RAG: Enhancing Retrieval for LLMs via Query Rewriting," *ACL*, 2025.
[9] S. Xu et al., "Multimodal Learning with Transformers: A Survey," *arXiv*, 2022.

[10] Z. Hu et al., "Advancing Healthcare with Large Language Models," *IEEE/CAA Journal of Automatica Sinica*, 2025.
[11] G. Stefanek and J. DeMuth, "An Avatar-Based Framework Using Large Language Models," *Issues in Information Systems*, 2024.
[12] M. Maslych et al., "Conversational Avatars Using LLMs in VR: A Pilot Study," *arXiv*, 2024.
[13] T. Zheng et al., "Efficient Document-Based Question Answering Using RAG," *Springer*, 2025.
[14] Y. Wan et al., "Hybrid Retrieval-Augmented Techniques for Large Language Models," *IEEE Access*, 2025.
[15] Z. Chen et al., "Temporal Question Answering with Retrieval-Augmented Generation," *IEEE*, 2025.
[16] D. Luo et al., "AutoStyle-TTS: Retrieval-Augmented Text-to-Speech Synthesis," *ICME*, 2025.
[17] J. Xue et al., "RAG in Prompt-Based Text-to-Speech Synthesis," *arXiv*, 2024.
[18] H. Zen et al., "LibriTTS: A Corpus for Speech Synthesis," *arXiv*, 2019.
[19] P. Karras et al., "Audio-Driven Facial Animation by Joint End-to-End Learning," *SIGGRAPH*, 2019.
[20] S. Lahiri et al., "LipSync3D: Data-Efficient Learning of 3D Talking Faces," *IEEE*, 2021.
[21] A. Marquardt et al., "RAG-Based Avatars in Virtual Reality Systems," *arXiv*, 2026.
[22] B. Callison-Burch et al., "Advances in Multimodal AI Systems," *CHI Conference*, 2025.
[23] D. Kiela et al., "Retrieval Augmentation Reduces Hallucination in Dialogue Systems," *ACL*, 2021.
[24] G. Putra et al., "Implementation of RAG with Digital Human Avatars," *Journal of Soft Computing Exploration*, 2025.
[25] Y. Gao et al., "Improving RAG Accuracy in Industrial Applications," *IEEE*, 2024.
[26] H. Wang et al., "Long Context RAG for Large Language Models," *NeurIPS*, 2024.
[27] A. Smith et al., "AI-Powered Virtual Assistants Using RAG," *IEEE Conference*, 2024.
[28] J. Doe et al., "Conversational AI with Document Grounding," *ACM*, 2023.
[29] K. Ravishankar, "Human-AI Interaction in Voice-Based Systems," *MUM Conference*, 2024.
[30] T. Noltey et al., "Communication Systems in Intelligent Environments," *IEEE*, 2007.