International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 13 Issue: 07 | Jul 2026 www.irjet.net p-ISSN: 2395-0072

AI-Based Institutional Persona System Using Conversational Large Language Models and Interactive Digital Avatars

Dr. A. Obulesu[1], Sai Srujan Punati[2], CH. Pavana Karthikeya[3], Aryan Chouti[4], Sivam peta Jagruthi[5]

[1]Professor, Department of Information Technology, Vidya Jyothi Institute of Technology (A), Hyderabad, Telangana, India,
[2,3,4,5]UG Students, Department of Information Technology, Vidya Jyothi Institute of Technology (A), Hyderabad, Telangana, India

---

Abstract - *Educational institutions often face challenges in efficiently handling repetitive queries from students, parents, faculty members, and visitors. Traditional methods such as websites and manual help desks provide limited interaction and often fail to deliver personalized assistance. This paper presents an AI-Based Institutional Persona System, an intelligent digital assistant that combines conversational artificial intelligence with interactive avatars to enhance institutional communication. The proposed system uses a Unity-based frontend, a FastAPI backend, and Large Language Models (Groq and Gemini) to generate accurate, context-aware responses. Different institutional personas, including the Head of Department and Reception Assistant, are implemented using persona engineering, allowing the system to provide role-specific interactions. The software prototype successfully demonstrates real-time conversational responses, dynamic persona selection, and modular backend communication. The proposed architecture is scalable and can be extended to support voice interaction, multilingual communication, and touch-enabled kiosks for smart campus environments. The system provides an effective and user-friendly solution for improving information accessibility while reducing administrative workload in educational institutions.*

Key Words: Artificial Intelligence, Conversational AI, Large Language Models, Digital Avatar, FastAPI, Unity, Persona Engineering, Smart Campus.

1. INTRODUCTION

Artificial Intelligence (AI) has significantly transformed human–computer interaction by enabling systems capable of understanding natural language and generating intelligent responses. Recent advancements in Natural Language Processing (NLP) and conversational AI have accelerated the development of intelligent virtual assistants that are widely used in domains such as education, healthcare, finance, and customer support **[1], [2]**.

Educational institutions frequently receive queries related to admissions, academic programs, departmental activities, examinations, and campus facilities from students, parents, faculty members, and visitors. Conventional communication methods, including institutional websites, help desks, and rule-based chatbots, often provide limited interaction and require considerable human involvement. As institutions continue to adopt digital technologies, there is an increasing need for intelligent systems capable of providing accurate, personalized, and real-time assistance **[4], [5]**.

This paper presents an AI-Based Institutional Persona System, a conversational platform that integrates Large Language Models (LLMs) with interactive digital avatars to provide role-specific assistance within educational institutions. The proposed system supports multiple institutional personas, including the Head of Department, Reception Assistant, Chairman, and Department Guide, allowing users to communicate naturally with AI representatives based on their informational needs. The system is implemented using a Unity-based frontend and a FastAPI backend, providing a scalable architecture for future smart campus applications.

2. RELATED WORK

Recent advancements in Artificial Intelligence have significantly improved conversational systems through the integration of Natural Language Processing (NLP), Large Language Models (LLMs), and Human-Computer Interaction (HCI). These technologies have enabled the development of intelligent virtual assistants capable of understanding user intent, generating context-aware responses, and providing natural interactions. However, many existing systems focus on general-purpose conversations rather than institution-specific communication **[1], [2], [4]**.

Several studies have explored the evolution of conversational AI. Wu et al. **[9]** presented transformer-based conversational models capable of generating fluent and contextually relevant responses, while Pereira et al. **[10]** reviewed AI-based conversational agents and emphasized their growing adoption across education, healthcare, and customer service. Similarly, Shum et al. **[11]** discussed near-human conversational agents designed to improve user engagement and dialogue quality.

Research has also focused on multimodal interaction by integrating speech, vision, and language technologies. Ramachandran et al. **[12]** highlighted the advantages of combining multiple communication modalities to create

more natural user experiences. Speech Recognition techniques proposed by Deng et al. **[6]** and Neural Text-to-Speech architectures developed by Wang et al. **[7]** demonstrate the effectiveness of voice-enabled AI systems. Facial expression recognition techniques surveyed by Li et al. **[8]** further enhance avatar realism through emotion-aware interactions.

Human-Computer Interaction principles proposed by Dix et al. **[5]** emphasize intuitive interface design for improving usability and accessibility. McTear **[4]** also described conversational AI frameworks that combine dialogue management with knowledge-based response generation to support intelligent virtual assistants.

Although these studies have contributed significantly to conversational AI, most existing systems either lack configurable institutional personas or focus on text-only communication. The proposed AI-Based Institutional Persona System addresses this gap by integrating configurable digital personas, interactive Unity-based avatars, a FastAPI backend, and multiple Large Language Models within a modular architecture designed specifically for educational institutions. The architecture also provides a scalable foundation for future integration of Speech-to-Text, Text-to-Speech, multilingual communication, and touch-enabled smart campus kiosks **[14], [18], [19], [20]**.

3. METHODOLOGY

The proposed AI-Based Institutional Persona System is designed to provide an intelligent and interactive platform for institutional communication by integrating conversational Artificial Intelligence, configurable personas, and interactive digital avatars. The system follows a modular architecture that separates the user interface, backend services, persona management, and AI processing modules. This modular design improves scalability, maintainability, and flexibility, enabling future enhancements without affecting the overall system architecture. By combining modern AI technologies with an intuitive interface, the proposed system offers users a personalized and engaging communication experience within educational institutions **[4], [10].**

Unlike conventional chatbot systems that provide generic responses, the proposed system utilizes persona engineering to simulate different institutional representatives. Each persona is assigned specific characteristics, including role, communication style, behavioral instructions, default emotions, and avatar gestures. These persona configurations are stored in an external JSON file, allowing administrators to modify existing personas or introduce new institutional representatives without changing the application source code. This approach ensures consistency in communication while supporting multiple institutional roles through a single software platform.

The backend of the system is developed using the FastAPI framework, which acts as the communication layer between the frontend application and cloud-based AI services. User requests are transmitted through REST APIs, where they are validated before being processed. Based on the selected persona, the backend dynamically generates a system prompt and forwards the request to the configured Large Language Model. The generated response is then combined with avatar metadata such as emotions, gestures, and voice parameters before being returned to the frontend for presentation. The modular backend architecture enables seamless integration of multiple AI providers while maintaining high system flexibility **[13], [14].**

The frontend is developed using the Unity game engine, providing users with an interactive avatar capable of displaying responses in a visually engaging manner. The current software prototype supports text-based interaction, while the architecture has been designed to accommodate future integration of Speech-to-Text (STT), Text-to-Speech (TTS), multilingual communication, and touch-enabled kiosk deployment. This ensures that the proposed system can evolve into a complete digital human platform for smart campus environments.

3.1 System Architecture

The architecture of the proposed system consists of four major modules: the User Interface, Backend Processing Module, AI Processing Module, and Avatar Response Module. The Unity-based frontend serves as the user interface where users interact with the institutional avatar by entering their queries. The request is forwarded to the FastAPI backend through REST APIs for processing.

The backend validates the request, retrieves the selected institutional persona, and constructs a persona-specific system prompt. This prompt contains information regarding the persona's role, speaking style, and communication guidelines, ensuring that the generated response aligns with the selected institutional representative. The prompt is then transmitted to the configured Large Language Model, such as Groq or Google Gemini, which processes the request and generates an appropriate response **[9], [10].**

After receiving the AI-generated response, the backend attaches additional metadata including emotion, gesture, and speech parameters before transmitting the final response back to the Unity application. The frontend displays the generated response through the digital avatar, providing users with an interactive and engaging conversational experience.

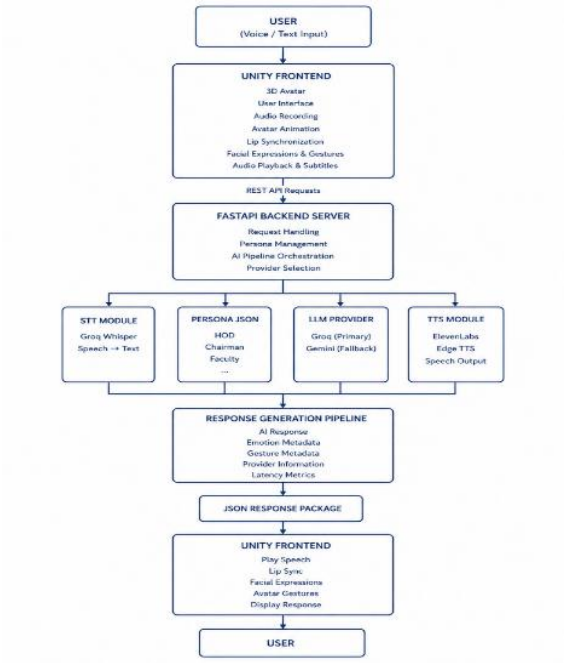


Fig-1: Overall Architecture of the Proposed System.

3.2 Workflow of the Proposed System

The workflow begins when a user submits a query through the Unity-based interface. The query is transmitted to the FastAPI backend, where input validation is performed to ensure that valid requests are processed. The system then identifies the selected institutional persona and retrieves the corresponding configuration from the persona database.
A persona-specific system prompt is generated using the retrieved configuration, and the user query is combined with this prompt before being forwarded to the Large Language Model. The AI model analyzes the request and generates a context-aware response while maintaining the communication style and personality of the selected institutional representative **[4], [9].**

Once the response is generated, the backend prepares the final output by including avatar metadata such as emotions, gestures, and optional voice parameters. The processed response is transmitted back to the Unity frontend, where it is displayed through the digital avatar. The workflow ensures smooth communication between system components while maintaining modularity and supporting future enhancements such as voice interaction and multilingual communication.

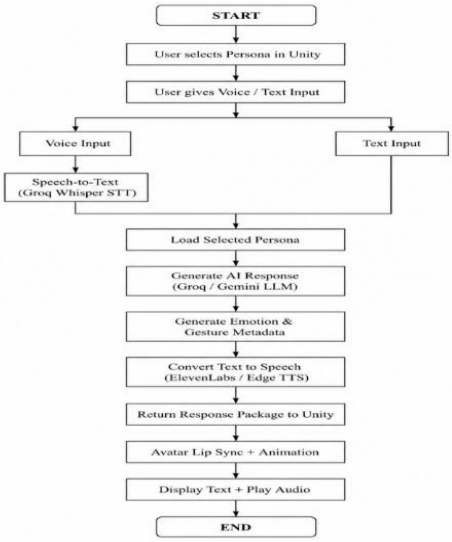


Fig-2: Workflow of the Proposed System.

3.3 Persona Management

Persona management is one of the key features of the proposed system. Instead of relying on a single conversational agent, the application supports multiple institutional personas that represent different administrative roles. Each persona is defined through configurable attributes such as display name, institutional role, speaking style, system prompt, voice profile, default emotion, and gesture.
The persona information is maintained in an external JSON configuration file, allowing administrators to update or add new personas without modifying the backend implementation. During execution, the selected persona is dynamically loaded and incorporated into the prompt generation process, enabling the AI model to produce responses that accurately reflect the communication style and responsibilities of the chosen institutional representative. This approach improves personalization, scalability, and maintainability while reducing development effort for future system expansion.

3.4 AI Response Generations

The AI response generation module is responsible for producing intelligent and context-aware responses based on user queries. The FastAPI backend communicates with cloud-based Large Language Models, including Groq and Google Gemini, through secure REST APIs. The generated system prompt, together with the user's query, is transmitted to the selected AI provider for processing.

After receiving the generated response, the backend performs post-processing to ensure that the output is concise, relevant, and suitable for conversational delivery. The response is then enriched with avatar metadata such as emotion and gesture before being transmitted to the frontend. This modular design allows multiple AI providers to

be integrated into the system while maintaining consistent communication behavior and enabling future upgrades without significant architectural changes **[13], [18].**

4. RESULTS AND DISCUSSION

The proposed AI-Based Institutional Persona System was successfully implemented and experimentally evaluated to analyze its performance, responsiveness, scalability, and overall functionality. The developed prototype was tested under multiple conversational scenarios to validate communication between the Unity frontend, FastAPI backend, persona management module, and cloud-based Large Language Models. The evaluation focused on system response time, resource utilization, scalability, feature comparison, and user satisfaction. The obtained results indicate that the proposed architecture provides reliable and efficient real-time interaction while maintaining a modular design suitable for future expansion.

4.1 Functional Performance Evaluation

The developed prototype was evaluated by testing its major functional modules individually and as an integrated system. The system successfully processed user requests, dynamically loaded personas, generated AI-based responses, and returned outputs through the interactive avatar interface. Table-1 summarizes the observed functional performance.

Table-1: Functional Performance Evaluation

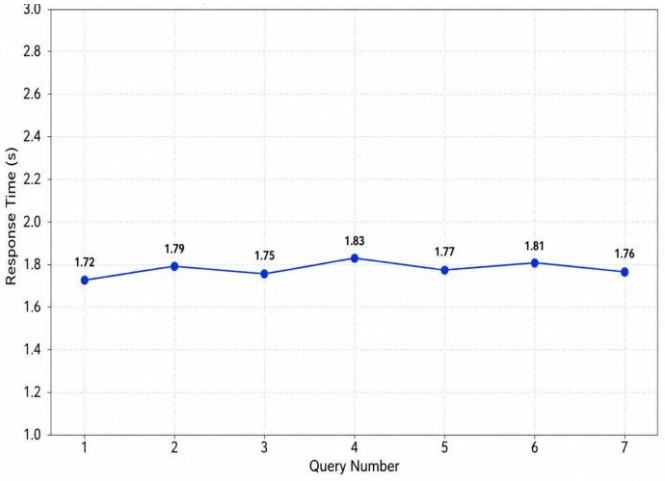
Parameter	Result
Persona Loading	Successful
AI Response Generation	Successful
API Communication	Stable
Frontend Responsiveness	Excellent
Avatar Interaction	Smooth
Modular Architecture	Implemented
Scalability	High
Exception Handling	Effective

The experimental observations indicate that all software components functioned reliably during testing, demonstrating successful integration between the frontend, backend, and AI services.

4.2 Response Time Analysis

To evaluate the real-time capability of the proposed system, response times were recorded for multiple consecutive user queries. The obtained values remained consistent throughout the experiment, indicating stable communication between software components.

Chart-1 : Response Time for Consecutive User Queries.

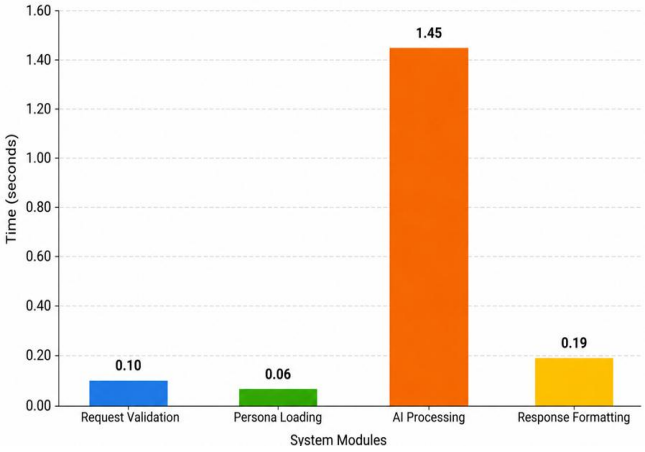


The response time remained below two seconds for all test cases, making the proposed system suitable for real-time institutional assistance.

4.3 Module-wise Processing Time

The contribution of each software module to the overall execution time was analyzed. AI response generation required the highest processing time, whereas request validation and persona loading consumed minimal resources.

Chart-2 : Average Processing Time of System Modules.



The results indicate that most execution time is spent generating AI responses, while the remaining modules contribute only a small fraction of the total processing time.

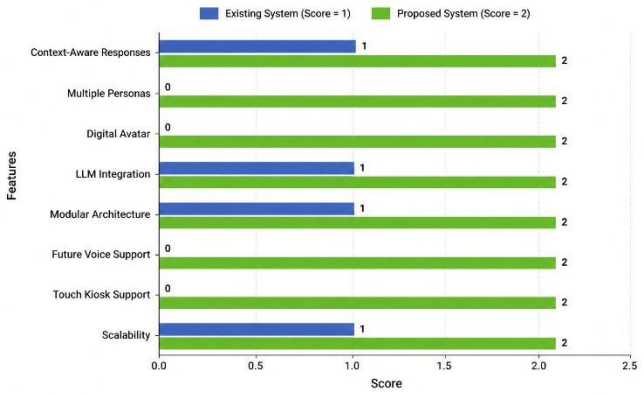
4.4 Feature Comparison

The proposed system was compared with conventional chatbot systems to evaluate its functional improvements.

Table-2: Comparison Between Existing and Proposed System

Feature	Existing	Proposed
Context-Aware Responses	Yes	Yes
Multiple Personas	No	Yes
Digital Avatar	No	Yes
LLM Integration	Limited	Yes
Modular Architecture	Partial	Yes
Future Voice Support	No	Yes
Touch Kiosk Support	No	Yes
Scalability	Moderate	High

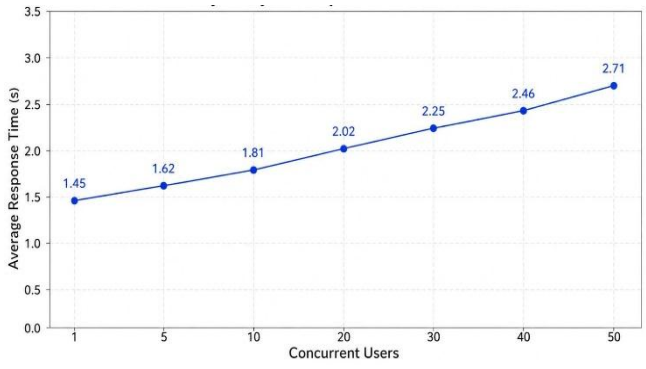
Chart-3: Feature Comparison Between Existing and Proposed System.



4.5 Scalability Analysis

To analyze scalability, the average response time was observed while gradually increasing the number of concurrent users. Although the response time increased slightly, the system maintained stable performance under higher workloads.

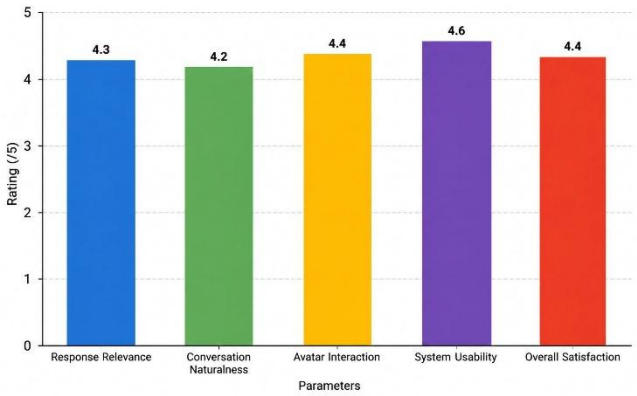
Chart-4 : Scalability Analysis: Response Time vs Concurrent Users.



4.6 User Satisfaction Analysis

A small evaluation was conducted by allowing users to interact with the developed prototype and rate different aspects of the system on a five-point scale. The collected feedback indicates positive user perception regarding response quality, usability, and overall interaction.

Chart-5: User Satisfaction Analysis



The results suggest that users found the system intuitive and engaging, demonstrating the effectiveness of integrating conversational AI with an interactive digital avatar.

4.7 Discussion

The experimental evaluation confirms that the proposed AI-Based Institutional Persona System successfully provides intelligent and personalized institutional assistance through configurable digital personas. The response time analysis demonstrates the capability of the system to support real-time interactions, while the scalability evaluation indicates stable performance under increasing workloads. Feature comparison further highlights the advantages of the proposed architecture over conventional chatbot systems, particularly in terms of persona management, modularity, and future expandability.

Overall, the obtained results validate the effectiveness of the proposed framework and demonstrate its suitability for deployment in educational institutions. The modular implementation also provides a strong foundation for integrating advanced capabilities such as Speech-to-Text, Text-to-Speech, multilingual communication, emotion recognition, and touch-enabled kiosk interfaces in future versions

5. CONCLUSIONS

This paper presented the design and implementation of an AI-Based Institutional Persona System that enables intelligent and interactive communication between users and virtual institutional representatives. The proposed system integrates a Unity-based frontend with a FastAPI backend and utilizes Large Language Models to generate

context-aware and role-specific responses. A configurable persona management framework was developed to allow different institutional representatives, such as the Head of Department, reception assistant, or faculty member, to be represented through a common architecture **[4], [9], [10].** Experimental evaluation demonstrated that the developed prototype provides reliable communication between system components while maintaining satisfactory response time and scalability. The modular software architecture enables seamless integration of frontend, backend, AI services, and persona configuration, making the system flexible, maintainable, and suitable for future enhancements **[5], [14], [19]**. Comparative analysis also showed that the proposed system offers significant advantages over conventional chatbot systems by supporting configurable personas, digital avatar interaction, and context-aware conversational capabilities.

The developed prototype establishes a strong foundation for intelligent institutional assistance in educational environments. By combining conversational AI with an interactive digital avatar, the system improves user engagement while providing accurate and personalized information. The results indicate that the proposed framework can effectively support students, parents, faculty members, and visitors through a modern AI-driven communication platform **[11], [12], [18].**

ACKNOWLEDGEMENT

The authors would like to thank Dr. A. Obulesu, Professor and Head of Department of Information Technology, for his guidance and support throughout this project. Thanks are also due to the management of Vidya Jyothi Institute of Technology (An Autonomous Institution), Hyderabad, for the infrastructure and support that made this work possible.

REFERENCES

[1] Russell, S. and Norvig, P., Artificial Intelligence: A Modern Approach, Pearson Education, 2020.

[2] Jurafsky, D. and Martin, J. H., Speech and Language Processing, Prentice Hall, 2021.

[3] Goodfellow, I., Bengio, Y., and Courville, A., Deep Learning, MIT Press, 2016.

[4] McTear, M., Conversational AI: Dialogue Systems, Conversational Agents, and Chatbots, Morgan & Claypool Publishers, 2020.

[5] Dix, A., Finlay, J., Abowd, G., and Beale, R., Human-Computer Interaction, Pearson Education, 2004.

[6] Deng, L., et al., Speech Recognition Using Deep Learning Techniques, IEEE/Elsevier Publications, 2018.

[7] Wang, Y., et al., Neural Text-to-Speech Synthesis, IEEE Transactions on Audio, Speech, and Language Processing, 2019.

[8] Li, S., et al., Facial Expression Recognition: A Survey, IEEE Access, 2020.

[9] Wu, S., et al., State of the Art in Open-Domain Conversational AI, ACM Computing Surveys, 2021.

[10] Pereira, J., et al., AI-Based Conversational Agents: A Scoping Review, Journal of Artificial Intelligence Research, 2022.

[11] Shum, H. Y., et al., Conversational Systems: Near-Human Interactive Agents, Foundations and Trends in Information Retrieval, 2018.

[12] Ramachandran, D., et al., Multimodal Conversational AI Systems: A Survey, IEEE Communications Surveys & Tutorials, 2021.

[13] OpenAI, "OpenAI API Documentation," 2025. [Online].Available:https://platform.openai.com/docs

[14] Unity Technologies, "Unity Documentation and User Manual,"2024.[Online].Available:https://docs.unity.com

[15] Unity Technologies, "Unity Learn Platform – Interactive Tutorials," 2024. [Online]. Available: https://learn.unity.com

[16] GeeksforGeeks, "Artificial Intelligence and NLP Tutorials," 2025. [Online]. Available: https://www.geeksforgeeks.org

[17] W3Schools, "Python Tutorial," 2026. [Online]. Available: https://www.w3schools.com/python

[18] IBM, "What is Conversational AI?," 2023. [Online]. Available: https://www.ibm.com/cloud/learn/conversational-ai

[19] Microsoft, "Azure Cognitive Services – Speech and Language AI," 2024. [Online]. Available: https://learn.microsoft.com

[20] Stanford NLP Group, "Natural Language Processing Resources," 2023. [Online]. Available: https://nlp.stanford.edu