

## Article

# NAIA: A Robust Artificial Intelligence Framework for Multi-Role Virtual Academic Assistance

Adrián F. Pabón M. , Kenneth J. Barrios Q. , Samuel D. Solano C.  and Christian G. Quintero M. \* 

Department of Electrical and Electronics Engineering, Universidad del Norte, Barranquilla 081007, Colombia; afpabon@uninorte.edu.co (A.F.P.M.); bkenneth@uninorte.edu.co (K.J.B.Q.); sdsolano@uninorte.edu.co (S.D.S.C.)

\* Correspondence: christianq@uninorte.edu.co

## Abstract

Virtual assistants in academic environments often lack comprehensive multimodal integration and specialized role-based architecture. This paper presents NAIA (Nimble Artificial Intelligence Assistant), a robust artificial intelligence framework designed for multi-role virtual academic assistance through a modular monolithic approach. The system integrates Large Language Models (LLMs), Computer Vision, voice processing, and animated digital avatars within five specialized roles: researcher, receptionist, personal skills trainer, personal assistant, and university guide. NAIA's architecture implements simultaneous voice, vision, and text processing through a three-model LLM system for optimized response quality, Redis-based conversation state management for context-aware interactions, and strategic third-party service integration with OpenAI, Backblaze B2, and SerpAPI. The framework seamlessly connects with the institutional ecosystem through Microsoft Graph API integration, while the frontend delivers immersive experiences via 3D avatar rendering using Ready Player Me and Mixamo. System effectiveness is evaluated through a comprehensive mixed-methods approach involving 30 participants from Universidad del Norte, employing Technology Acceptance Model (TAM2/TAM3) constructs and System Usability Scale (SUS) assessments. Results demonstrate strong user acceptance: 93.3% consider NAIA useful overall, 93.3% find it easy to use and learn, 100% intend to continue using and recommend it, and 90% report confident independent operation. Qualitative analysis reveals high satisfaction with role specialization, intuitive interface design, and institutional integration. The comparative analysis positions NAIA's distinctive contributions through its synthesis of institutional knowledge integration with enhanced multimodal capabilities and specialized role architecture, establishing a comprehensive framework for intelligent human-AI interaction in modern educational environments.



Academic Editors: Gila Kurtz and Antonella Poce

Received: 15 October 2025

Revised: 27 November 2025

Accepted: 30 November 2025

Published: 3 December 2025

**Citation:** M., A.F.P.; Barrios Q., K.J.; Solano C., S.D.; Quintero M., C.G. NAIA: A Robust Artificial Intelligence Framework for Multi-Role Virtual Academic Assistance. *Systems* **2025**, *13*, 1091. <https://doi.org/10.3390/systems13121091>

**Copyright:** © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

**Keywords:** artificial intelligence; large language models; virtual assistant; computer vision; digital avatars; human-AI interaction

## 1. Introduction

The integration of artificial intelligence (AI) in higher education has experienced unprecedented growth in recent years, particularly following the emergence of advanced language models and multimodal AI systems [1]. AI is seen as a transformative force in education, with generative AI setting the mission for Artificial General Intelligence (AGI) [1]. The application of AI in education (AIED) is growing exponentially and is prominently featured in reports as an important development in educational technology [2]. While AIED has a substantial history as a research domain, spanning about thirty (30) years

or even around fifty (50) years and has seen significant investments from tech giant like Amazon, Google and Facebook, its recent renaissance and exponential growth are primarily due to the abundance of data, economic access to computing power and advances in machine learning [2–4].

Higher education institutions face multiple challenges that demand innovative technological solutions. Faculty members struggle with mounting administrative burdens, such as making assignments and answering frequently asked questions [5]. This reduces the time available for core educational activities. AI-enables systems are being developed to support educators and reduce their workload by automating tasks like administration, assessment, feedback, and plagiarism detection, thereby freeing them to focus on creative, empathetic, and inspirational aspects of their profession [2,4].

The development of AI-powered virtual assistants in educational settings represents a significant evolution from early intelligent tutoring systems (ITS) to contemporary multimodal platforms. ITS are among the most common and likely the oldest applications of AI in education [3]. A meta-analysis of 39 studies found that ITS improves students' learning, though not as effectively as one-on-one human tutoring. However, ITS outperformed all other kinds of instruction methods such as standard classroom lessons, reading books (digital or print), or assignments [6].

Currently, AI, especially generative AI, is seen as a transformative force in education [1–3]. The vision of Artificial General Intelligence (AGI) includes machines being able to understand and integrate diverse input and output modalities, such as visual, auditory, and textual data [1].

Multimodal AI approaches are paving the way towards AGI in educational contexts, emphasizing auditory, visual, kinesthetic, and linguistic modes of learning [1].

This paper presents NAIA (Nimble Artificial Intelligence Assistant), a multi-role virtual assistant architecture designed to address the complex challenges facing academic environments through optimized multimodal integration. NAIA implements five specialized roles: researcher, receptionist, personal skills trainer, personal assistant, and university guide within a unified platform that leverages large language models, computer vision, and voice processing technologies. Figure 1 presents an overview of the proposed system.



**Figure 1.** An illustration of the general idea of NAIA.

It is important to highlight that the technologies integrated into NAIA were not incorporated arbitrarily but had already been explored and tested in a prior prototype [7].

The earlier iteration of NAIA provides initial evidence that directly informs the current integration strategy. Early deployments indicate a clear user preference for multimodal interaction, sustained engagement with avatar-based communication, and practical benefits derived from role-specific assistance in academic scenarios. The prototype further shows the relevance of combining speech processing, large language models, and university-

related information retrieval, as users consistently relied on these elements to complete informational tasks. These observations validate the importance of multimodality, RAG-based access to institutional content, and agent specialization, establishing an empirical foundation for the technology stack consolidated in the present version of NAIA. Additionally, the preliminary stage highlighted the need to streamline response delivery and strengthen the connection with official academic resources, informing the enhanced role definitions and the deeper integration with institutional services implemented in the current system.

Within this context, the present work focuses primarily on consolidating NAIA as a robust, deployable framework for multi-role academic assistance, while the user study described in Sections 3.5 and 4 provides an initial evaluation of usability and perceived effectiveness rather than a comprehensive effectiveness trial.

## 2. Related Works

### 2.1. Theoretical Framework

The theoretical components underlying this assistant, such as Large Language Model (LLMs), virtual assistants, computer vision, digital avatar, voice generation, and Retrieval-Augmented Generation (RAG), are presented in detail below. This section provides a comprehensive overview of these foundations.

#### 2.1.1. Large Language Models (LLMs)

The core of NAIA's capabilities stems from the implementation of Large Language Models (LLMs), which are neural network architectures based on Transformers that enable natural language understanding and generation through training on vast datasets [8]. These models achieve significant results in natural language processing (NLP) tasks such as translation, summarization, question answering, and classification [9]. Making them valuable for applications in content creation, customer support, data analysis, and automation of language-based tasks.

Beyond text-based LLMs, recent advancements have introduced multimodal models, which extend the Transformer architecture. The development of Multimodal Large Language Models (MLLMs) arose from the need to overcome the limitations of text-only Large Language Models (LLMs). Despite their strong capabilities in processing textual data, traditional LLMs had a limited scope of application, particularly in domains that are inherently multimodal. In response to these limitations, MLLMs were developed to process and fuse diverse modalities of information, not only text but also visual data such as images [10].

#### 2.1.2. Retrieval Augmented Generation (RAG)

RAG is a technique that grounds the responses of a Large Language Model (LLM) in external, verifiable knowledge sources [11,12]. This approach enhances search capabilities by integrating data retrieval strategies with LLM text generation, overcoming limitations of traditional search engines that may struggle with complex queries or providing relevant contextual insight [13]. By retrieving relevant document chunks from a domain-specific corpus (e.g., university policy documents or uploaded research papers) and providing them to the LLM context, RAG reduces the problem of generating factually incorrect content [11].

This process often involves converting text into high-dimensional vectors for semantic similarity calculation, enabling the system to perform search based on conceptual meaning rather than just keyword matching [11,13].

Ultimately, RAG mitigates the risk of factual inaccuracies or “hallucinations”, enhances the accuracy and credibility of generated answers, and ensures they are both relevant and factually precise for knowledge-intensive tasks [11–13].

#### 2.1.3. Virtual Assistants (VA)

A virtual assistant (VA) is a software-based agent designed to carry out tasks or services in response to user commands or questions. These systems can replicate human conversation through natural language processing (NLP) and are generally organized in three (3) main modules: a user interface for interaction, a core responsible for processing user information, and external tools for information retrieval and task execution [14].

#### 2.1.4. Digital Avatars

A 3D avatar can be described as a digital representation of a human operated by artificial intelligence. It is designed to interact with users through natural language, simulating realistic conversations and behaviors. Three-dimensional avatars are increasingly applied across industries such as education, healthcare, social media, and banking, where they enhance customer interaction. Their ability to deliver immersive and engaging experiences positions them as a transformative tool [15].

#### 2.1.5. Text-to-Speech (TTS)

Text-To-Speech (TTS) is a technology whose primary purpose is to generate human-like speech from text. This allows for control characteristics such as emotion, speaking style, language, and has found applications in contexts like assistants, robotics, and natural human-computer interaction [16].

#### 2.1.6. Speech-to-Text (STT)

Speech-to-Text (STT) is defined as the task of converting speech signals in one language to text. This technology finds applications in various contexts, including transcription, tourism, and hands-free communication [17].

### 2.2. State-of-the-Art

In this section, the State-of-the-Art related to virtual assistants is explored within the NAIA framework, focusing on user support across different types of interaction, whether through text, avatars, or other modalities, and examining how these assistants perform within their respective contexts.

First of all, the study presented in [18] presented AIIA (Artificial Intelligence-Enabled Intelligent Assistant), a virtual teaching assistant (VirtualTA) designed for personalized and adaptive learning in higher education. The system integrates GPT-3.5 for natural language processing, OpenAI embeddings, and Whisper + pyannote for transcription and speaker diarization of recorded lectures. Built with a NodeJS backend, PostgreSQL, React, and Canvas LMS integration, AIIA provides a wide range of student services: flashcards, automated quiz generation and grading, coding sandbox execution, summarization, and context-aware conversations. For instructors, it supports auto-evaluation of assignments, homework detection, and automated question generation. Evaluations demonstrated the system’s potential to reduce cognitive load, improve engagement, and support both students and instructors. Future directions include adaptive learning algorithms, multimodal integration, gamification, and ethical safeguards.

The research presented in [19] proposed DataliVR, a virtual reality (VR) application designed to improve data literacy education for university students. The system integrates immersive VR scenes with a ChatGPT-powered virtual avatar with speech-to-speech (Whisper) and text-to-speech (Oculus Voice SDK) for conversational assistance. The chatbot

improved user experience and usability but unexpectedly led to slightly lower learning performance, likely due to distraction or reliance on assistance. The work highlights the promise of VR+LLM integration for education but stresses the need for careful design to balance guidance with self-directed learning.

The study in [20] introduced EverydAI, a multimodal virtual assistant that supports everyday decision-making in cooking, fashion, and fitness. The system integrates GPT-4o mini for language reasoning, YOLO + Roboflow for real-time object detection, augmented reality, 3D avatar (Ready Player Me + Mixamo), voice interaction (Web Speech API + ElevenLabs), web scraping, image generation, and deployment on AWS with Flask and Three.js. Users interact via text, voice, or avatars to receive personalized, context-aware recommendations based on the available resources (ingredients, clothing, equipment). The assistant effectively reduced decision fatigue and improved task efficiency.

Work developed in [21] proposed a Virtual Twin framework that integrates 3D avatar generation, real-time voice cloning, and conversational AI to enhance virtual meeting experiences. The system employs Neural Radiance Fields (NeRF) + Triplane neural representations for photorealistic avatars, Tacotron-2 + WaveRNN for natural voice cloning, and a context-aware LLaMA 3.1 Instruct 8B model (fine-tuned on the AMI Meeting Corpus) for coherent, multi-turn dialogue. User feedback indicated the system is 85% more engaging than conventional assistants. Limitations remain in gesture fluidity, including latency and data privacy. The framework has applications in business, education, and healthcare, aiming to provide more immersive, human-like virtual collaboration.

Reference [22] introduced HIVA, a holographic 3D voice assistant to improve human-computer interaction in higher education. HIVA uses a Pepper's Ghost-style pseudo-holographic projection combined with an animated 3D mascot and Russian-language NLP models for natural dialogue. The assistant provides information about admissions, departments, fees, student life, and university services, functioning as an alternative to website or staff inquiries. Its architecture integrates speech-to-text, suggestion classification (Multinomial Naïve Bayes), short-answer subsystems, and a Telegram chatbot. HIVA has been deployed since July 2021, handling over 7000 user requests, with NLP classification accuracy ranging from 74–97%. Future work focuses on expanding datasets and enhancing speech recognition for noisy conditions.

The study in [23] proposed a 3D avatar-based voice assistant powered by large language models (LLMs) to enhance human-computer interaction beyond traditional assistants like Alexa. The system integrates speech recognition, emotion analysis, intent classification, text-to-speech, and Unity-based 3D avatar rendering. Results show a 40% rise in task completion and a 25% reduction in user frustration, confirming improved engagement through natural language dialogue and a lifelike 3D avatar. Future directions include adding gesture/emotion recognition, multimodal interactions, and AR integration for richer experiences.

Reference [24] proposed MAGI, a system of embodied AI-guided interactive digital teachers that combines large language models (LLMs) with Retrieval-Augmented Generation (RAG) and 3D avatars to improve educational accessibility and engagement. MAGI introduces a hybrid RAG paradigm that organizes educational content into a hierarchical tree structure for accurate knowledge retrieval, mitigating LLM hallucinations. The system pipeline integrates Llama 3.8 B as the backbone LLM and a text-to-speech (TTS) module. A web interface allows learners to interact with customizable 3D avatars. MAGI shows promise in bridging educational gaps and providing high-quality, personalized digital teaching experiences at scale.

The research presented in [25] shows ELLMA-T, a GPT-4-powered embodied conversational agent implemented in VRChat to support English language learning through

situated, immersive interactions. The system integrates Whisper speech-to-text, OpenAI TTS, Unity avatars, and OSC-based animation control for real-time dialogue, role-play, and adaptive scaffolding. ELLMA-T generates contextual role-play scenarios (supermarket, café, interview) and provides verbal + textual feedback.

Reference [26] introduced a speech-to-speech AI tutor framework that integrates noise reduction, Whisper ASR, Llama3 8B Instruct LLM, Retrieval-Augmented Generation (RAG), Piper-TTS, and Wav2Lip avatar into a fully edge-deployed system running on AI PCs. The system supports dynamic conversational learning with an animated avatar while ensuring low-latency, multimodal interaction. Results highlight effective integration across modules, with strengths in response accuracy and avatar throughput, though latency optimization remains a challenge.

The research in [27] developed UBOT, a virtual assistant that achieved significant improvements in institutional response times. The system implements Retrieval-Augmented Generation (RAG) technology to provide contextually relevant responses using institution-specific information, demonstrating the effectiveness of domain-focused AI systems in educational settings.

The development of domain-specific virtual assistants across various sectors has provided crucial insights into role-based specialization approaches. In museum environments, DEUSENS HYPERXPERIENCE SL [28] developed the Alice Assistant, emphasizing natural communication through Speech-To-Text technology and digital avatar delivery. The system creates intuitive visitor interactions through voice-based communication and focuses on enhancing museum experiences.

The work in [29] addressed an intelligent conversational agent for museum environments using Google Cloud Speech-To-Text, RASA for Natural Language Understanding, and SitePal for avatar-based speech synthesis. Their implementation demonstrates the technical integration of speech processing with avatar animation systems.

The research in [30] created a context-aware virtual assistant designed to adapt across various usage scenarios while maintaining natural dialogues. The system demonstrates versatility in handling multiple domains without requiring specialized configurations for different use cases.

The contribution in [31] on the ALICE Chatbot demonstrated how a single assistant architecture could handle multiple specialized knowledge domains (tourism, healthcare, sports) while maintaining a consistent user experience through visual representation. Their implementation maintains a consistent user experience through visual representation across different domains.

Several implementations demonstrated the importance of extending virtual assistants beyond conversational interaction to functional automation. Garibay Ornelas [32] implemented a virtual assistant for a Mexican airline operating through WhatsApp and web interfaces, specifically addressing customer service bottlenecks through automated interactions.

The research [33] developed healthcare-focused digital avatar systems for patient follow-up care and appointment management. The system demonstrates role-specific functionalities in healthcare contexts, focusing on scheduling and communication capabilities.

The contribution in [34] is an AI-based hospital assistant that improves patient information access, addressing critical delays in healthcare settings by providing timely responses about medical processes and patient status updates. Their focus centers on reducing information access delays and improving patient information accessibility.

The study in [35] implemented a tourism-focused virtual assistant that enhances visitor experiences by providing precise information about operating hours, local events, points

of interest, and gastronomy options. The system demonstrates success in domain-specific information delivery.

The work in [36] developed Edith, a university-focused virtual assistant using IBM Watson Assistant that efficiently resolves academic process queries across multiple interfaces. The system demonstrates multi-interface deployment capabilities and focuses on streamlining university-related information delivery for students and faculty.

The research in [37] proposed a Computer Vision-enhanced virtual assistant for higher education students in specialized disciplines. Their system integrates visual processing capabilities to provide specialized support that extends beyond traditional text-based interactions.

The authors in [38] implemented an AI assistant supporting engineering students' thesis development with specialized tools for enhancing research methodology. Their system combines conversational AI with specialized academic functions.

The [39] developed an AI-powered career guidance system that analyzes prospective students' profiles to recommend suitable higher education paths. This work demonstrated the value of personalized AI assistance in academic contexts.

The research in [40] proposed an AI-based instructor for motor skill learning in virtual co-embodiment, demonstrating that virtual instructors can effectively replace human instructors while maintaining or improving learning efficiency. Their research validated the potential for AI-driven personalized instruction.

Recent advances in knowledge-enhanced conversational systems provided crucial insights for NAIA's information processing capabilities. Mlouk and Jiang [41] developed KBot, a chatbot leveraging knowledge graphs and machine learning to improve natural language understanding over linked data. The system provides multilingual query support and demonstrates improved accuracy through structured knowledge integration.

Finally, the reference [42] developed AIDA-Bot, which integrates knowledge graphs to improve natural understanding within the scholarly domain. The system demonstrates the capability to answer natural language queries about research papers, authors, and scientific conferences, focusing specifically on academic research assistance.

Table 1 lists all the studies mentioned above, summarizing all the main characteristics such as work, impact, and area.

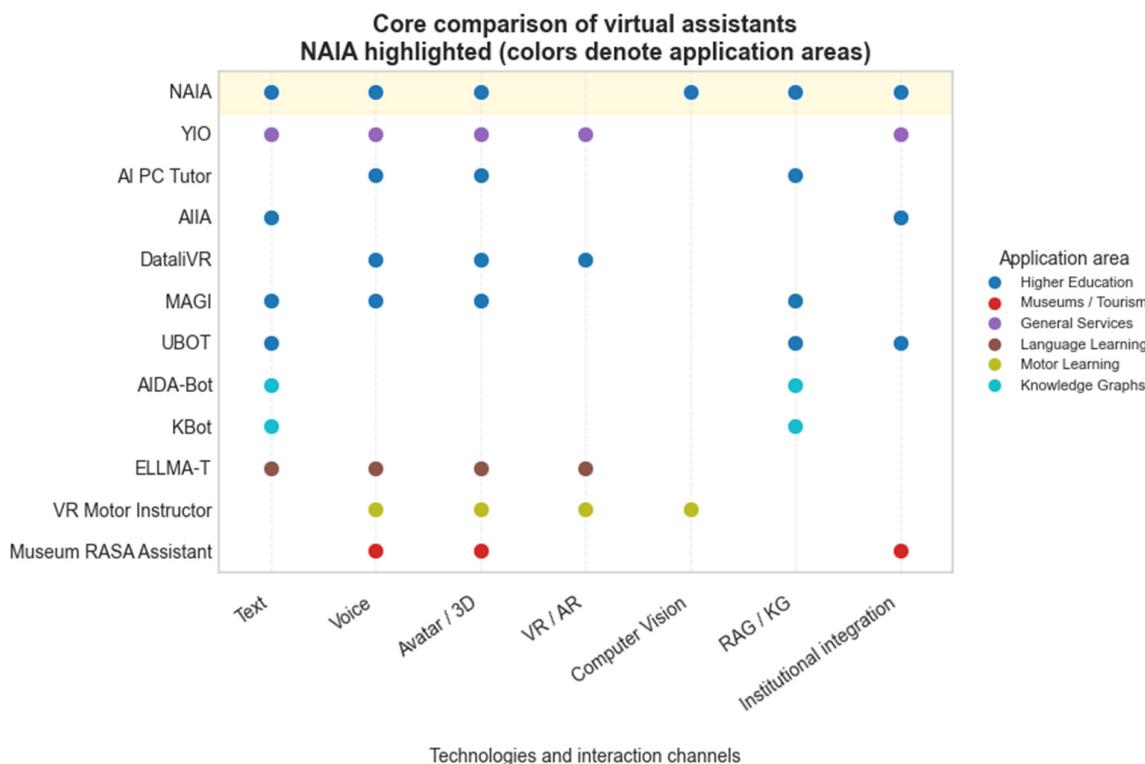
**Table 1.** Overview of State-of-the-Art Embodied and AI-Driven Virtual Assistants.

Reference	Work	Impact	Area
[18]	AI-enabled Virtual Teaching Assistant (AIIA)	Enhanced personalization and adaptive learning	Higher Education
[19]	VR application for data literacy (DataliVR)	Improved engagement and usability of the learning process	Higher Education
[20]	Multimodal assistant (EverydAI)	Reduced decision fatigue in daily routines	Lifestyle
[21]	Virtual Twin for meetings	Enhanced realism and interactivity in virtual collaboration	Business/Education
[22]	Holographic 3D voice assistant (HIVA)	Provided immersive university info access	Higher Education
[23]	LLM-based 3D avatar assistant	Increased task completion and reduced frustration	Human–Computer Interaction
[24]	MAGI embodied digital teachers	Improved education accessibility and engagement	Higher Education

**Table 1.** *Cont.*

Reference	Work	Impact	Area
[25]	ELLMA-T, an embodied LLM tutor in VRChat	Provided immersive, contextualized English practice	Language Learning
[26]	Real-time TTS AI Tutor on AI PCs	Delivered private, multimodal learning at the edge	Higher Education
[27]	Virtual assistant for university response times	Improved response times for institution-related queries	Higher Education
[28,29]	Museum virtual assistant with speech interface	Enhanced visitor interaction through natural communication	Museums/Cultural
[30]	YIO: Context-aware virtual assistant	Versatile dialogue system for multiple service contexts	General Services
[31]	Alice: Digital avatar assistant	Multi-domain query resolution with visual representation	Tourism/Healthcare/Sports
[32]	Airline customer service virtual assistant	Reduced waiting times for customer requests	Aviation Industry
[33]	Medical patient follow-up avatar	Improved patient care management and appointment scheduling	Healthcare
[34]	Hospital patient care AI assistant	Enhanced response times to patient information requests	Healthcare
[35]	Tourist information virtual assistant	Improved tourist access to local information and services	Tourism
[36]	Edith: University query assistant	Streamlined university-related information delivery	Higher Education
[37]	Computer Vision-enabled educational assistant	Specialized support for medical and engineering students	Higher Education
[38]	Scientific production assistant	Enhanced thesis development support for engineering students	Academic Research
[39]	Career guidance virtual system	Improved career selection process for prospective students	Higher Education
[40]	AI instructor for motor skill learning in virtual co-embodiment	Improved learning efficiency and skill retention without requiring a human instructor	Virtual Learning and Motor Learning
[41]	KBot: Knowledge Graph-Based Chatbot	Improved natural language understanding and structured data retrieval using linked data	Conversational AI and Knowledge Graphs
[42]	AIDA-Bot: Conversational Agent with Knowledge Graphs	Enhanced academic research assistance with structured, accurate, and verifiable scholarly data	Conversational AI and Knowledge Graphs in Academia

To complement the tabular overview of the reviewed systems, a concise visual comparison is introduced in Figure 2. This representation groups a selected subset of assistants according to their main interaction channels and underlying technological capabilities, providing a clearer perspective on how different approaches relate to one another and how NAIA is positioned within this landscape.



**Figure 2.** Comparative visualization of NAIA and representative virtual assistants.

The visualization highlights the diversity of technological directions explored across recent virtual assistant platforms. Several systems combine multiple interaction channels such as voice, avatar-based interfaces, VR or AR environments, computer vision analysis, or structured knowledge sources, which reflects a broader trend toward increasingly multimodal and context-aware solutions. Although these assistants originate from distinct application categories, including education, tourism, healthcare, and knowledge management, they share a common interest in integrating richer modes of interaction with more capable reasoning mechanisms. Within this landscape, NAIA aligns with the direction observed in current developments and distinguishes itself through its simultaneous support for text, voice, and avatar-based interaction, together with retrieval augmented access to institutional information and full integration with university systems. This configuration positions NAIA as a comprehensive academic platform built upon the same technological tendencies identified across the reviewed work.

### 3. Proposed Approach

This paper proposes NAIA (Nimble Artificial Intelligence Assistant), a comprehensive virtual assistant architecture specifically designed for academic environments. Accordingly, the technical description of this study, and the accompanying evaluation, is intended to illustrate how users experience and perceive this framework in a realistic institutional deployment. The intelligent assistant is developed with the objective of supporting academic community users through five distinct specialized roles: Researcher, Receptionist, Personal Skill Trainer, Personal Assistant, and University Guide, integrating State-of-the-Art technologies previously mentioned.

Each role provides unique characteristics that make it valuable within specific contexts of the academic environment. The Researcher role specializes in document creation, academic information retrieval, text analysis, content generation, and professional research writing support. The Receptionist serves as a comprehensive information hub, managing

campus navigation, local information access, and visitor support services. The Personal Skills Trainer enhances soft skills through simulations, language practice, appearance analysis, and personalized training reports. It also includes interview practice sessions for both job applications and university admissions. The Personal Assistant manages daily administrative tasks, including email management, scheduling, and calendar integration. The University Guide provides institution-specific information through a RAG (Retrieval-Augmented Generation) system built with official university documents.

NAIA enables user interaction through a user-friendly platform featuring a dedicated 3D avatar for each selectable role. Through this platform, NAIA can interact with users via user image analysis to provide positive environmental feedback, supports both text and voice-based interaction, and responds contextually based on user-specific stored information, creating a personalized experience for each user.

NAIA aims to enhance the academic environment experience through integration with various tools from Universidad del Norte's technological ecosystem. The system operates through Microsoft's tenant authentication, ensuring secure access for the academic community while maintaining institutional privacy standards.

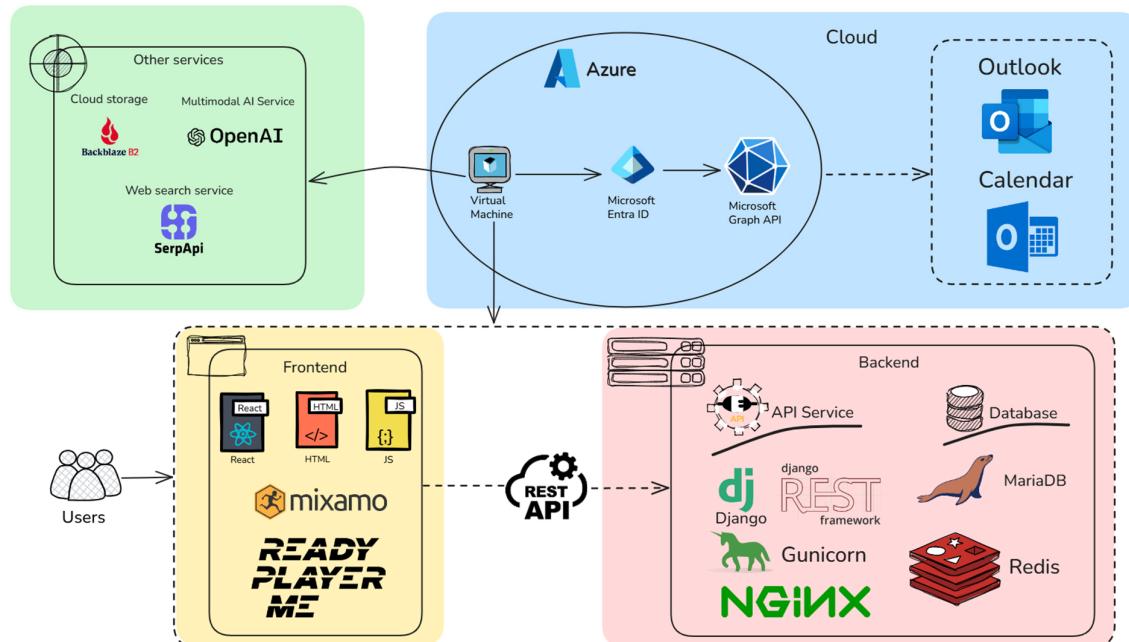
This section explains NAIA's general workflow, technological integration, and detailed role-specific implementations.

### 3.1. Architecture

#### 3.1.1. System Architecture

The NAIA platform has evolved from an initial prototype [7] into a scalable, decoupled, multi-tier system designed for deployment within an enterprise academic environment. The architecture prioritizes modularity, maintainability, and real-time performance to support its diverse functional requirements.

Figure 3 illustrates the comprehensive integration architecture of NAIA within the Universidad del Norte ecosystem. The system integrates multiple technological components through a unified web platform that serves as the primary interface for all user interactions.



**Figure 3.** NAIA's Architecture and Technologies Integration.

The platform described in this work builds upon the lessons learned from the initial prototype, which successfully demonstrated the technical viability of the multimodal assistant concept [7].

The following subsections analyze these components in detail, explaining how each element integrates into the complete system.

### 3.1.2. Frontend and Platform Design

The frontend architecture implements a sophisticated multi-technology stack that orchestrates visual rendering, avatar animation, and real-time speech processing within the client environment. The platform employs a hybrid approach combining React.js for complex 3D avatar rendering with traditional HTML/CSS/JavaScript for core interface elements, optimizing both performance and development flexibility. This architectural decision enables specialized handling of computationally intensive avatar animations while maintaining lightweight delivery for standard interface components.

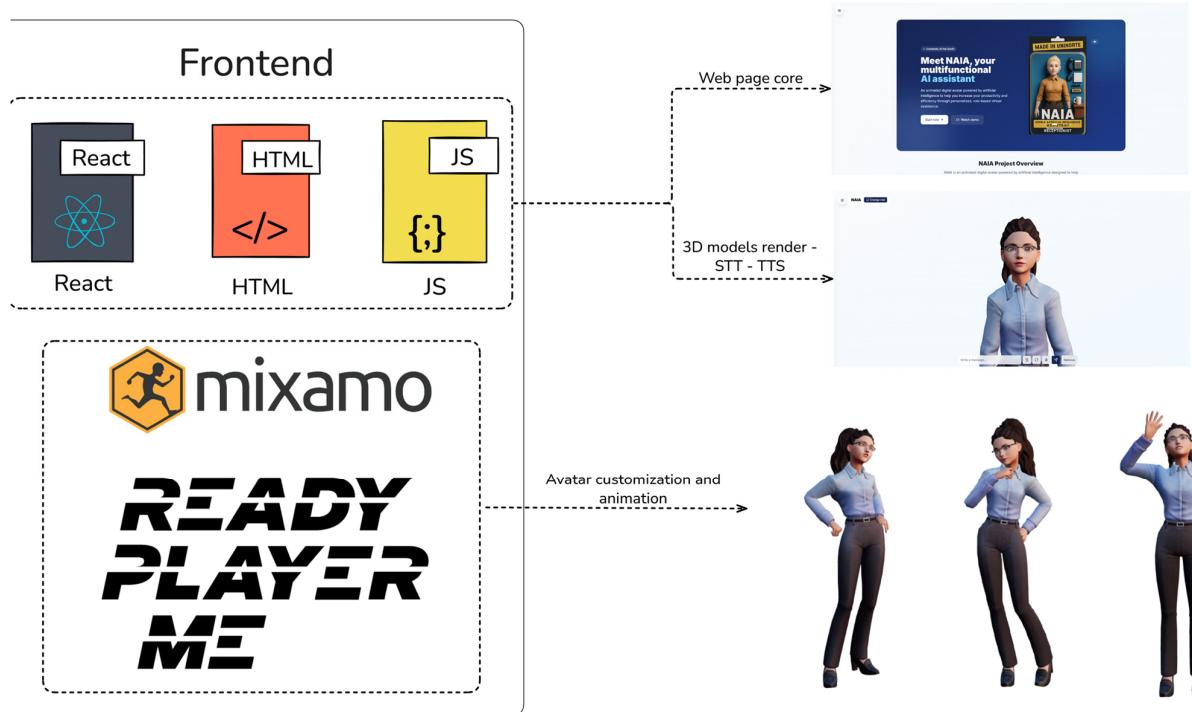
The digital avatar system represents a critical frontend component, integrating Ready Player Me for initial avatar creation with Mixamo for professional animation libraries. Ready Player Me provides a streamlined avatar generation pipeline that produces customizable 3D models without requiring specialized 3D modeling expertise, ensuring consistent visual quality across all five NAIA roles. Mixamo contributes an extensive animation repository including gestures, facial expressions, and body movements that create natural, context-appropriate responses during conversations. These avatar components are processed through Blender for format optimization before integration into the React-based rendering pipeline using Three.js and React Three Fiber, which enables real-time 3D visualization directly within web browsers without plugins.

Speech processing capabilities are strategically implemented on the frontend to minimize latency and enhance conversational fluidity. The Speech-to-Text (STT) functionality leverages the Web Speech API, a browser-native technology that eliminates server roundtrips for voice recognition, providing immediate transcription with automatic silence detection and multilingual support. Similarly, Text-to-Speech (TTS) processing occurs on the frontend through direct API calls to voice synthesis services, with audio streaming beginning before complete synthesis to minimize perceived latency. The frontend coordinates TTS output with avatar mouth movements and gestures, synchronizing visual and auditory elements to create cohesive multimodal interactions.

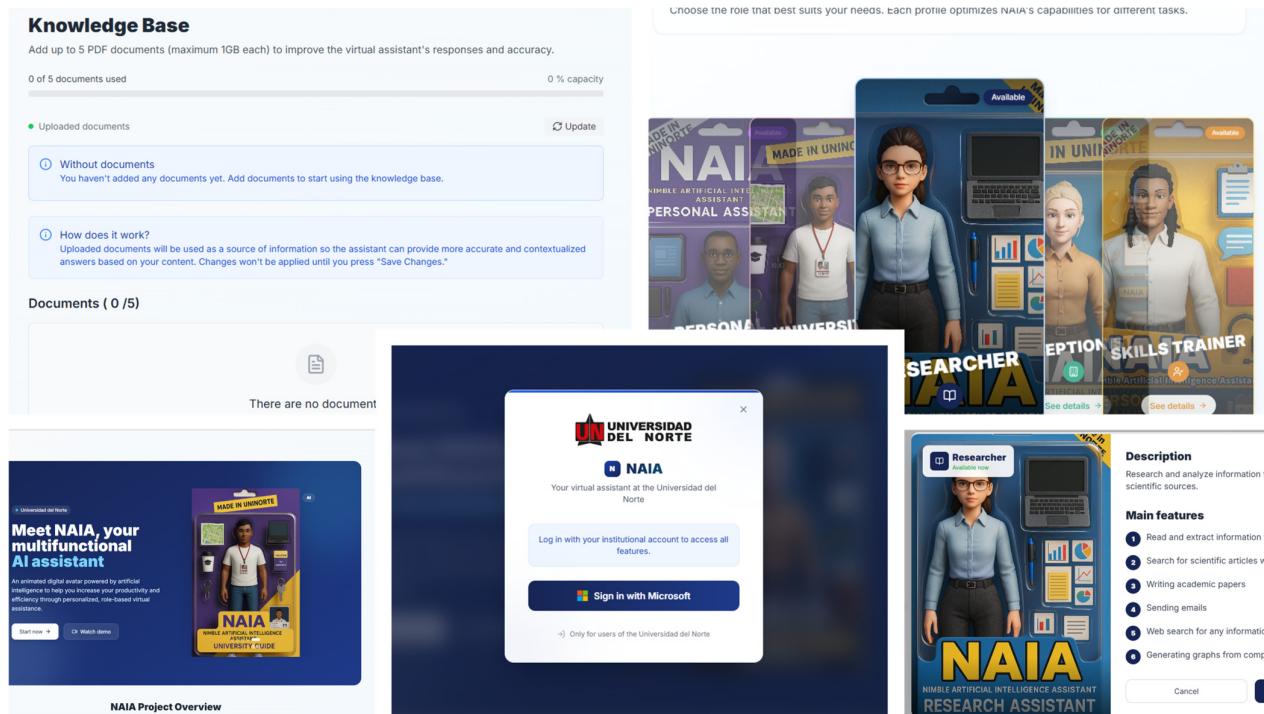
The platform design implements responsive layouts that adapt seamlessly across devices while maintaining full functionality of avatar rendering and speech processing. The interface architecture supports both voice-first and text-based interactions, accommodating user preferences and environmental constraints while maintaining consistent conversation state across modalities.

Figure 4 illustrates the comprehensive frontend architecture that orchestrates NAIA's client-side technologies. The diagram demonstrates how React.js, HTML, and JavaScript work in concert to deliver the web interface, while Ready Player Me and Mixamo integration enable sophisticated 3D avatar creation and animation. This architectural visualization highlights the critical role of client-side speech processing in achieving real-time conversational interactions.

The platform is accessible through the following link <https://naia.uninorte.edu.co/naia> (accessed on 4 October 2024). In addition, Figure 5 graphically illustrates the different aspects of the NAIA platform, including the roles, the access pathways, and the overall system design. Several demonstration videos are also available to showcase the system's functionality, allowing users to gain a general understanding of how each aspect of the platform operates.



**Figure 4.** Frontend architecture and technology stack.



**Figure 5.** NAIA platform general design.

### 3.1.3. Backend Architecture

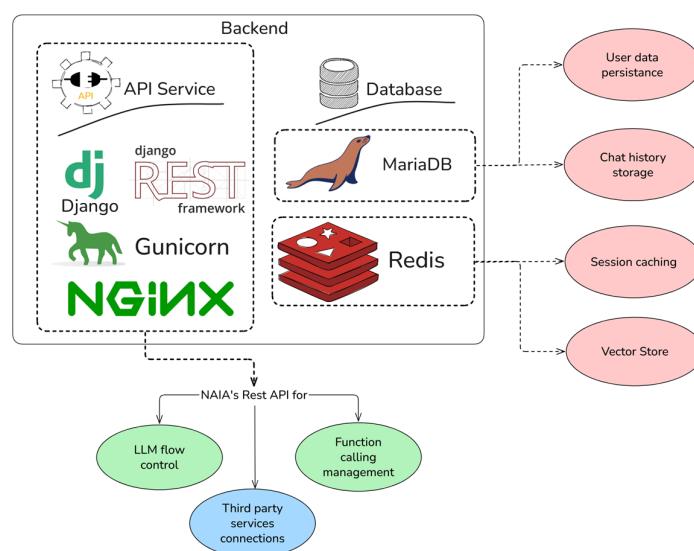
The backend architecture implements a robust, scalable infrastructure built on Django REST Framework that orchestrates complex AI interactions while maintaining high performance and reliability. The system employs a multi-layered architecture with Django serving as the core framework, deployed through Gunicorn as the WSGI HTTP server and Nginx as the reverse proxy, ensuring efficient request handling and load distribution.

The API service layer exposes RESTful endpoints that manage all client-server communications through versioned interfaces, enabling backward compatibility while supporting continuous platform evolution. The Django REST Framework facilitates sophisticated request validation, serialization, and response formatting while implementing authentication and authorization mechanisms that integrate with institutional identity providers. The API architecture implements three critical management systems that coordinate NAIA's intelligent capabilities: LLM flow control for managing conversation context and response generation, function calling management for orchestrating role-specific capabilities and external service integrations, and third-party service connections for seamless integration with external APIs and institutional systems.

Database architecture implements a dual-storage strategy combining MariaDB for relational data persistence with Redis for high-performance caching and vector operations. MariaDB manages structured data, including user profiles and conversation histories.

Redis serves multiple critical functions within the backend architecture, operating as an in-memory data store that dramatically reduces latency for real-time operations. The implementation includes session caching that maintains user state across requests without database overhead, enabling sub-millisecond response times for session validation and retrieval. Redis also functions as a vector store for the Retrieval-Augmented Generation (RAG) system, utilizing its native vector similarity search capabilities to quickly locate relevant documents and information from institutional knowledge bases.

Figure 6 presents the multi-layered backend architecture that powers NAIA's server-side operations. The diagram showcases the Django REST Framework implementation with Gunicorn and Nginx, alongside the dual-database strategy employing MariaDB for persistent storage and Redis for high-performance caching and vector operations. This infrastructure design enables the platform to manage LLM flow control, function calling, and third-party service connections through a unified API service layer.



**Figure 6.** Backend infrastructure and database architecture.

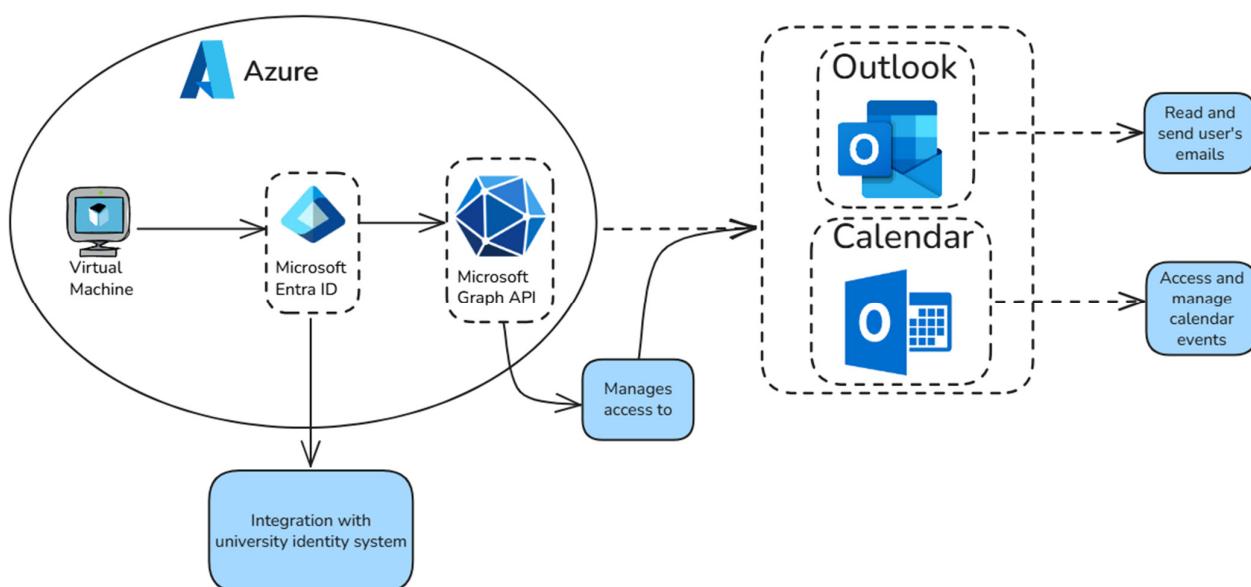
### 3.1.4. Cloud Infrastructure

NAIA operates on a virtual machine deployed within Universidad del Norte's Microsoft Azure tenant, ensuring seamless integration with institutional cloud infrastructure and governance frameworks. This deployment strategy places all security configurations, compliance policies, network access control, and data governance protocols under direct management of the university's IT security department, guaranteeing adherence to institutional standards and regulatory requirements.

Microsoft Entra ID (formerly Azure Active Directory) serves as the foundational identity and access management layer, providing secure authentication and authorization for all NAIA users. This integration enables Single Sign-On (SSO) capabilities that allow students, faculty, and staff to access NAIA using their existing institutional credentials without creating separate accounts. This identity integration ensures that NAIA inherits the university's existing security posture and compliance certifications while providing seamless user experiences.

Microsoft Graph API represents the critical integration layer that transforms NAIA from an isolated assistant into a deeply integrated component of users' academic workflows. Through Graph API, NAIA gains authorized access to Microsoft 365 services within the institutional tenant, enabling rich interactions with productivity tools that users already depend upon. The Outlook integration provides bidirectional email capabilities, allowing NAIA to read users' inboxes for context awareness, compose and send messages on users' behalf, and manage email organization through intelligent filtering and categorization. Calendar integration through Graph API enables NAIA to access users' academic schedules, create and modify events with appropriate attendee management, identify scheduling conflicts and suggest optimal meeting times, and synchronize reminders across the Microsoft 365 ecosystem.

Figure 7 depicts NAIA's cloud infrastructure within the Microsoft Azure ecosystem. The diagram illustrates how the virtual machine deployment integrates with Microsoft Entra ID for identity management and leverages Microsoft Graph API to access Outlook and Calendar services. This architectural representation emphasizes the seamless integration with the university's existing Microsoft 365 infrastructure, enabling NAIA to become an integral part of users' institutional workflows.



**Figure 7.** Cloud infrastructure integration within the Microsoft Azure ecosystem.

### 3.1.5. Third-Party Services Integration

The platform extends its core conversational capabilities through strategic integration with specialized third-party services that provide essential functionality across document management, multimodal AI processing, and information retrieval.

Backblaze B2 serves as the primary cloud storage solution for document and multimedia asset management, providing cost-effective, scalable storage for various file types critical to NAIA's multimodal operations. The platform supports multiple document formats, including PDF, DOCX, PNG, and TXT files, enabling users to upload, process,

and reference academic materials during conversations. The Backblaze integration implements secure access controls through time-limited, pre-signed URLs that ensure documents remain protected while enabling authorized access for processing. This storage system maintains conversation-specific document associations, allowing NAIA to reference previously uploaded materials across sessions while preserving user privacy through isolated storage namespaces.

OpenAI's suite of services represents the cognitive core of NAIA's intelligence, providing both language understanding and multimodal processing capabilities through a unified API interface. The GPT-4.1 integration enables sophisticated natural language processing for context-aware conversations, role-specific response generation, and complex reasoning across academic domains. The multimodal capabilities extend beyond text processing to include vision analysis for uploaded images, enabling NAIA to interpret charts, diagrams, and visual content relevant to academic discussions. The platform leverages OpenAI's embedding models to generate semantic representations of documents. Text-to-speech and speech-to-text services through OpenAI ensure consistent quality across voice interactions while maintaining low latency through optimized API calls and response streaming.

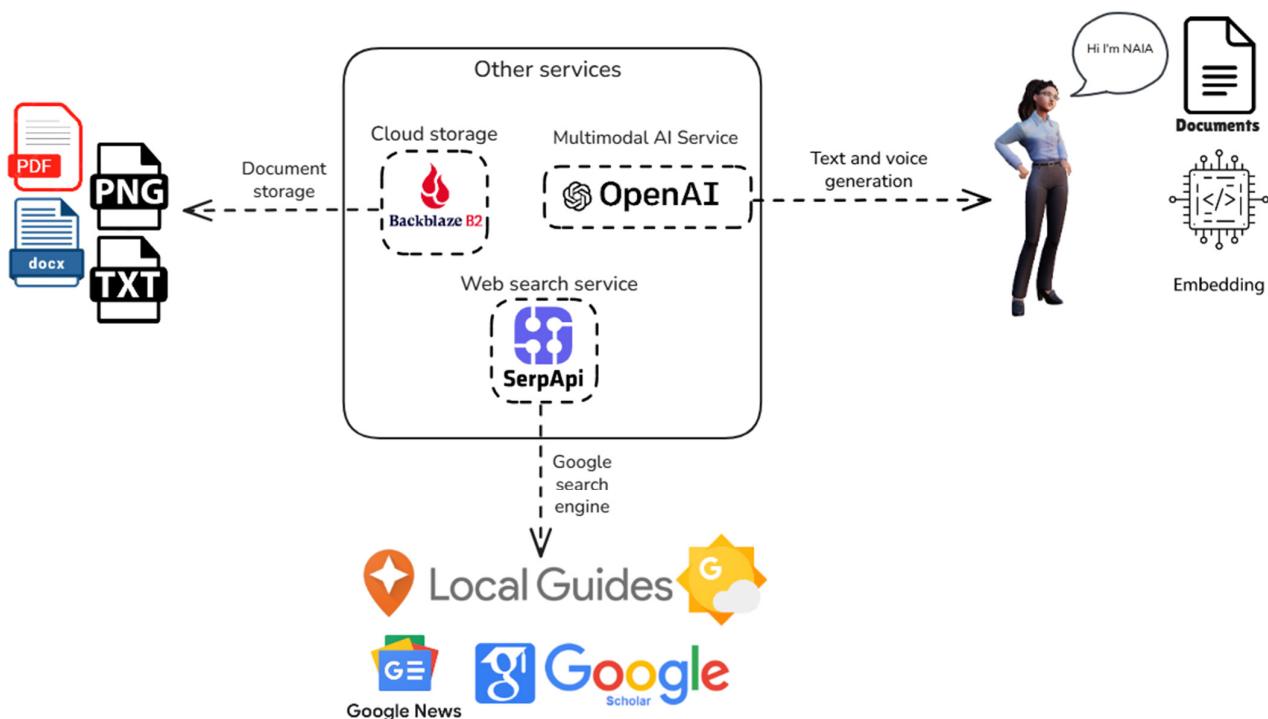
SerpAPI provides comprehensive web search capabilities that enable NAIA to access current information beyond its training data. The integration implements intelligent query construction that translates natural language requests into optimized search parameters, maximizing result relevance while minimizing API calls. Through SerpAPI's Google search engine integration, NAIA accesses multiple specialized search verticals, including Google Scholar for academic publications, Google News for current events relevant to university activities, Google Local Guides for restaurant and venue recommendations, and standard web search for general information retrieval. The search integration implements filtering and ranking algorithms that prioritize authoritative sources, academic publications, and institutionally relevant content.

Document processing workflows leverage the combination of these services to create powerful academic capabilities. When users upload documents, the system stores files in Backblaze B2, generates embeddings through OpenAI for semantic search, extracts text content for analysis and citation, and maintains metadata for efficient retrieval. This integrated approach enables sophisticated document-based interactions where NAIA can answer questions about uploaded materials, compare multiple documents, generate summaries with proper citations, and maintain context across conversation sessions.

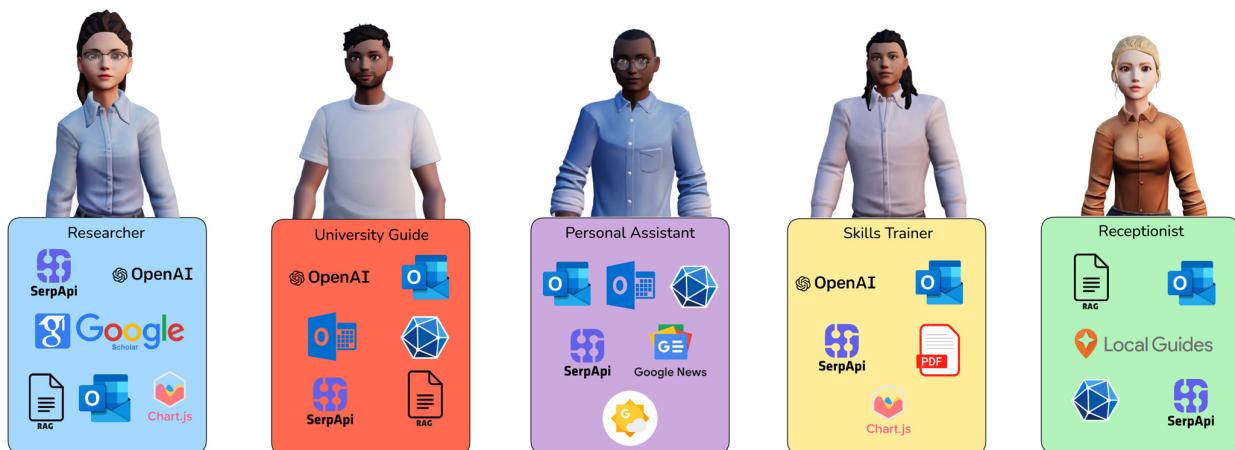
Figure 8 demonstrates the extensive third-party service integration architecture that extends NAIA's capabilities. The diagram shows how Backblaze B2 provides document storage for multiple file formats, OpenAI delivers multimodal AI processing, including embedding generation, and SerpAPI enables comprehensive web search through various Google services. This visual representation highlights how these external services are orchestrated to create a cohesive ecosystem for document processing, intelligent responses, and real-time information retrieval.

### 3.2. The Five Specialized Roles of NAIA

The NAIA platform operationalizes its capabilities through five specialized agents, or roles, each designed to address a distinct set of needs within the academic community. Each role integrates multiple AI technologies and external service integrations to provide comprehensive functionality that extends beyond simple conversational interactions. Figure 9 relates the different roles of NAIA with the technologies that each role incorporates, in accordance with the descriptions provided in the previous subsections.



**Figure 8.** Third-party service integration architecture.



**Figure 9.** Five Specialized Roles of NAIA.

The selection and design of the avatars associated with each role are based on a psychological analysis conducted by professionals from Universidad del Norte. Their guidance ensured that the visual and behavioral traits of the avatars reflected principles of diversity and inclusion, thereby aligning the system with institutional values and fostering equitable representation within the academic community. The following subsections present each role in detail.

### 3.2.1. Researcher Role

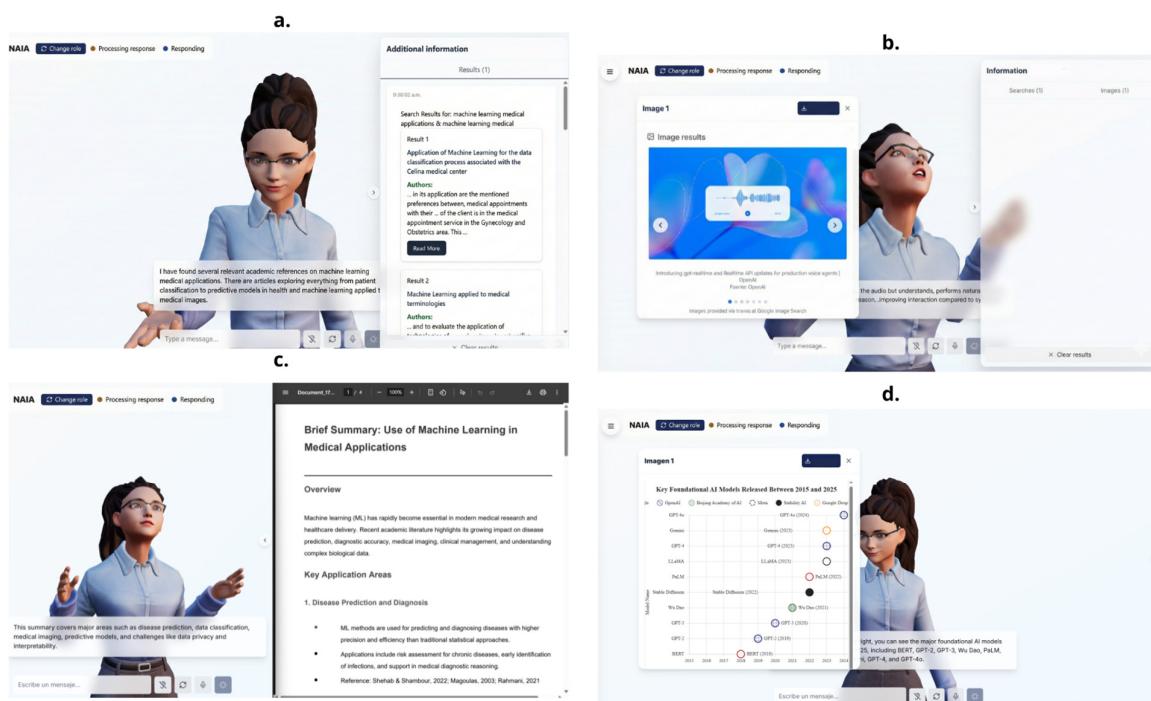
The Researcher role serves as a comprehensive research assistant, integrating seven specialized functions that support the complete academic workflow. This role leverages the OpenAI platform's diverse model ecosystem to deploy specialized assistants for specific tasks, demonstrating the strategic advantage of the platform's multi-model architecture.

The academic writing function employs a dedicated writing-focused assistant that can autonomously search Google Scholar references via SerpAPI integration and access the user's personal document repository through RAG (Retrieval-Augmented Generation)

technology to complement text generation. The document RAG system allows users to upload up to fifty (50) PDF documents, creating a personalized knowledge base that the role can query to provide contextually accurate responses based on the user's specific research materials.

For information retrieval, the role implements bibliographic search capabilities through Google Scholar integration via SerpAPI, enabling the discovery of relevant academic articles and research papers. The internet search function provides access to current information through Google's search engine via SerpAPI, with responses that include spoken narration, relevant images, and source links for verification. Similarly, the news access function leverages Google News through SerpAPI to provide recent news updates with the same multimodal response format.

The role's email functionality integrates with the institutional system, allowing users to send research-related information through the project's official email system, with intelligent redirection capabilities that can interpret commands like "send it to my email" and automatically route content to the logged-in user's institutional account. Finally, the chart creation function employs a specialized graphing assistant that can generate visualizations using data from user inputs, RAG documents, or internet searches, creating customized charts and graphs according to user specifications. Figure 10 shows examples of some of the functions previously explained.



**Figure 10.** Researcher role function examples ((a). Bibliographic search, (b). Internet search, (c). Academic writing, (d). Chart creation).

### 3.2.2. University Guide Role

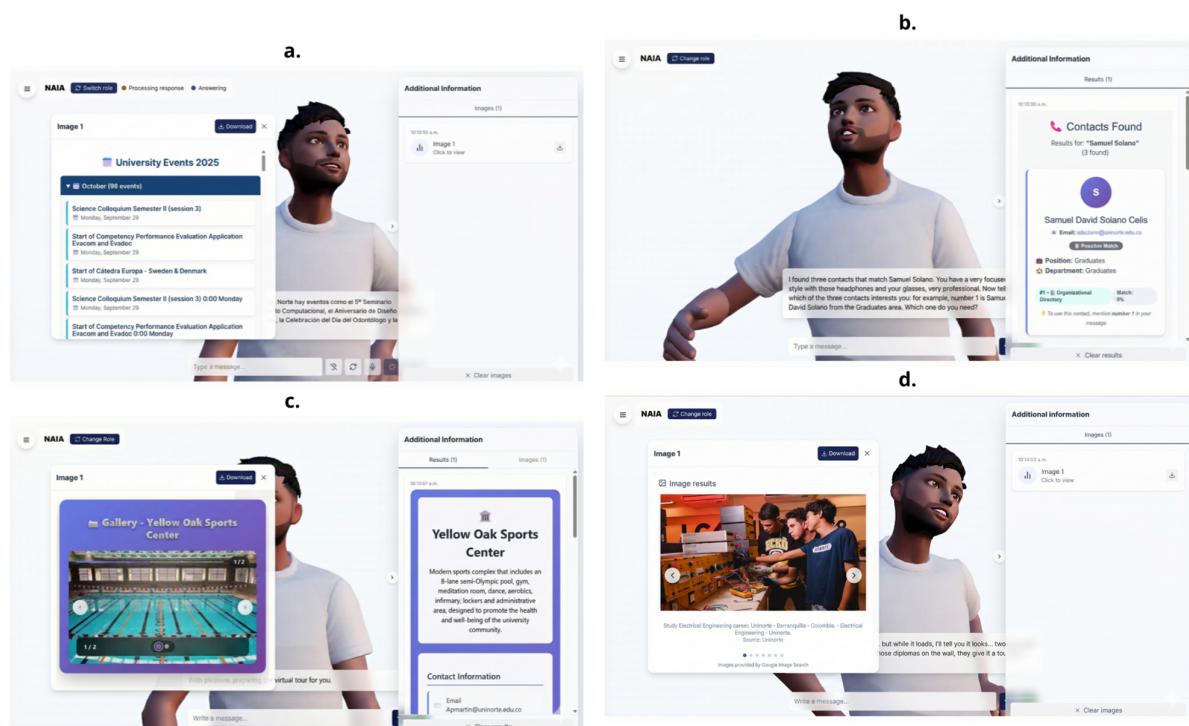
The University Guide role functions as a comprehensive institutional assistant, combining university-specific knowledge with broader informational capabilities through seven integrated functions. This role emphasizes institutional integration and personalized assistance through Microsoft Graph API connectivity.

The role's foundation is a specialized RAG system populated with official university documents, policies, and procedures, enabling accurate responses to institution-specific queries. Email functionality mirrors that of the Researcher role, providing communication capabilities for university-related information dissemination.

Direct calendar integration with the university's academic calendar system allows the role to provide real-time information about academic events, important dates, and institutional schedules. The university tour function offers interactive campus guidance, presenting visual tours of key university locations with accompanying images and detailed information about facilities and services.

The role implements prioritized internet search capabilities that focus on university-related queries, serving as an effective backup when the institutional RAG system cannot provide complete answers to specific questions. Personal calendar integration through the user's institutional email account enables the creation of personalized reminders and schedule management.

The university contact search function leverages Microsoft Graph API integration to access the official university directory, providing comprehensive contact information for faculty, staff, and departmental resources. Figure 11 shows examples of some of the functions previously explained.



**Figure 11.** University guide role function examples ((a). University's academic calendar, (b). Contact search, (c). Virtual tour, (d). University's web search).

### 3.2.3. Personal Assistant Role

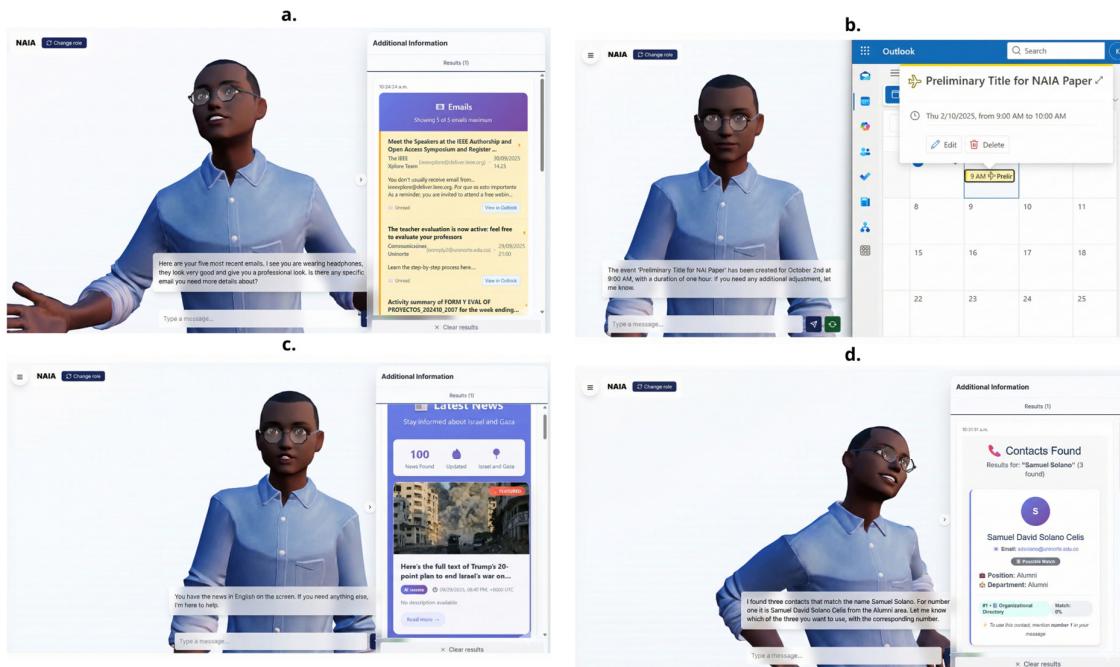
The Personal Assistant role represents the most extensively integrated functionality within NAIA's ecosystem, with seven functions that demonstrate deep integration with Microsoft Graph API infrastructure. This role's capabilities are predominantly enabled through the institutional Microsoft 365 environment, creating a seamless bridge between NAIA and the user's existing digital workflow.

The email management system operates through the user's institutional account rather than the project's official email, enabling direct communication from the user's personal institutional identity. The inbox reading function provides comprehensive access to the user's email system, allowing the assistant to search, summarize recent messages, highlight unread emails, and provide a detailed analysis of email content and patterns.

Calendar reading capabilities offer complete access to the user's institutional calendar within specified date ranges, enabling comprehensive schedule analysis and information

provision. The complementary calendar writing function allows for the creation of reminders, appointments, and events directly within the user's institutional calendar system.

The role incorporates university contact search functionality identical to the University Guide role, global news access like the Researcher role, and worldwide weather information capabilities, creating a comprehensive personal productivity assistant that bridges institutional and personal information needs. Figure 12 illustrates some of the functions discussed earlier.



**Figure 12.** Personal assistant role function examples ((a). Inbox reading, (b). Calendar writing, (c). Global news, (d). Contact search).

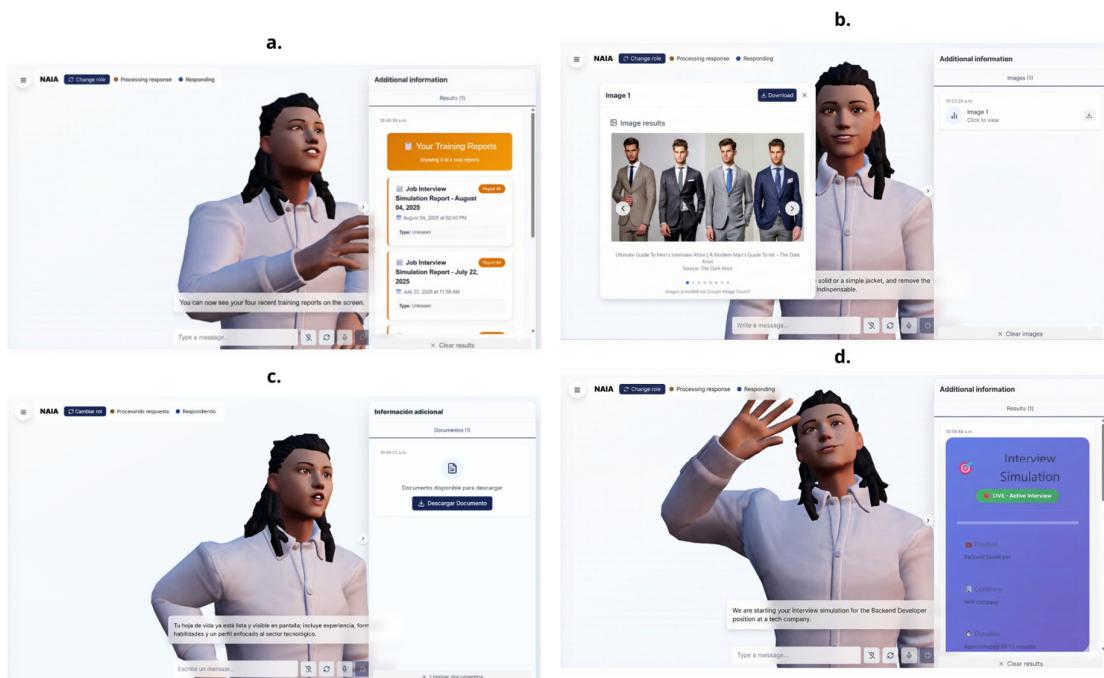
### 3.2.4. Personal Skills Trainer Role

The Personal Skills Trainer role focuses on professional development and skill enhancement through six specialized functions that combine AI-powered simulation with computer vision analysis and personalized feedback generation.

The job interview simulation function creates interactive practice sessions tailored to specific positions and company types, adapting conversation flows and question patterns to match desired career trajectories. Professional appearance analysis leverages computer vision capabilities to evaluate the user's visual presentation, providing specific feedback on professional attire and presentation elements.

The CV generation system creates fully customized resumes without format or style limitations, adapting content and presentation to user specifications and career objectives. The training report generation produces comprehensive HTML-formatted documents that include performance analysis, improvement recommendations, and progress documentation.

The training history function maintains accessible records of previous training sessions and performance reports, enabling progress tracking and skill development monitoring. Professional communication capability provides email functionality focused on career development and professional correspondence. Figure 13 highlights examples of the functions previously discussed.



**Figure 13.** Skills trainer role function examples ((a)). Training history, (b). Appearance analysis and recommendations, (c). CV generation, (d). Job interview simulation).

### 3.2.5. Receptionist Role

The Receptionist role serves as a comprehensive information and assistance hub, providing seven functions that focus on campus navigation, local information access, and visitor support services.

The campus facilities information system provides detailed information about university restaurants, sports facilities, academic buildings, retail locations, and digital zones, including operating hours and available services. The university contact search maintains the same functionality as other roles, ensuring consistent access to institutional directory information.

Local events search capabilities locate activities and events in specified cities through interactive calendar interfaces, while restaurant discovery provides gastronomic options beyond campus boundaries with integrated mapping functionality. The tourist attraction search offers comprehensive travel guidance for any location with interactive travel guides and recommendations.

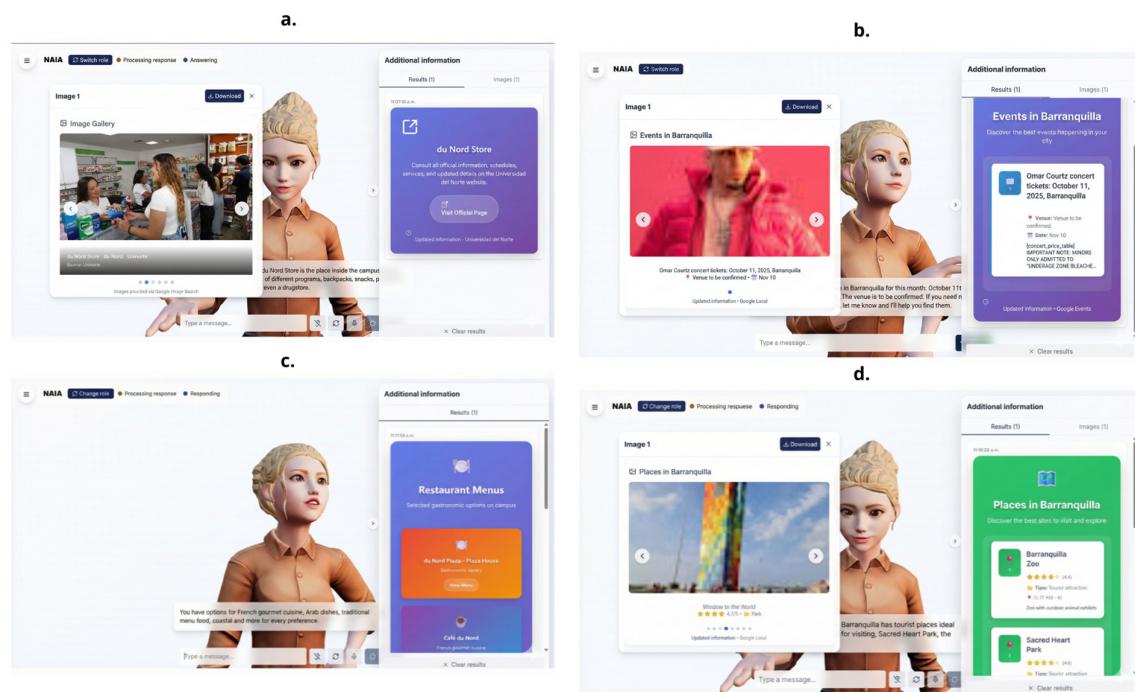
The role maintains standard email functionality through the project's official communication system and provides commercial campus information covering university-operated restaurants, bookstores, snack vendors, and other campus retail services. Figure 14 illustrates additional examples of the functions discussed above.

Table 2 provides a comprehensive overview of the functional design of each specialized agent.

### 3.3. Functional Design

All the previous technologies work together through a well-defined process flow that not only provides a clear structure to follow but also helps maintain control over the tools powering NAIA's capabilities. Figure 15 illustrates the workflow the app follows to operate.

This process ensures an orderly management of NAIA's functionality throughout the entire workflow. The first step, as shown in the figure, involves user input, which can be entered manually or provided through voice. In the latter case, Speech-to-Text (STT) is required to transform the user's voice into text.



**Figure 14.** Receptionist role function examples. (a) Campus facilities information. (b) Local events. (c) University restaurants. (d) Local restaurants discovery.

**Table 2.** Functional Design of NAIA's Specialized Roles.

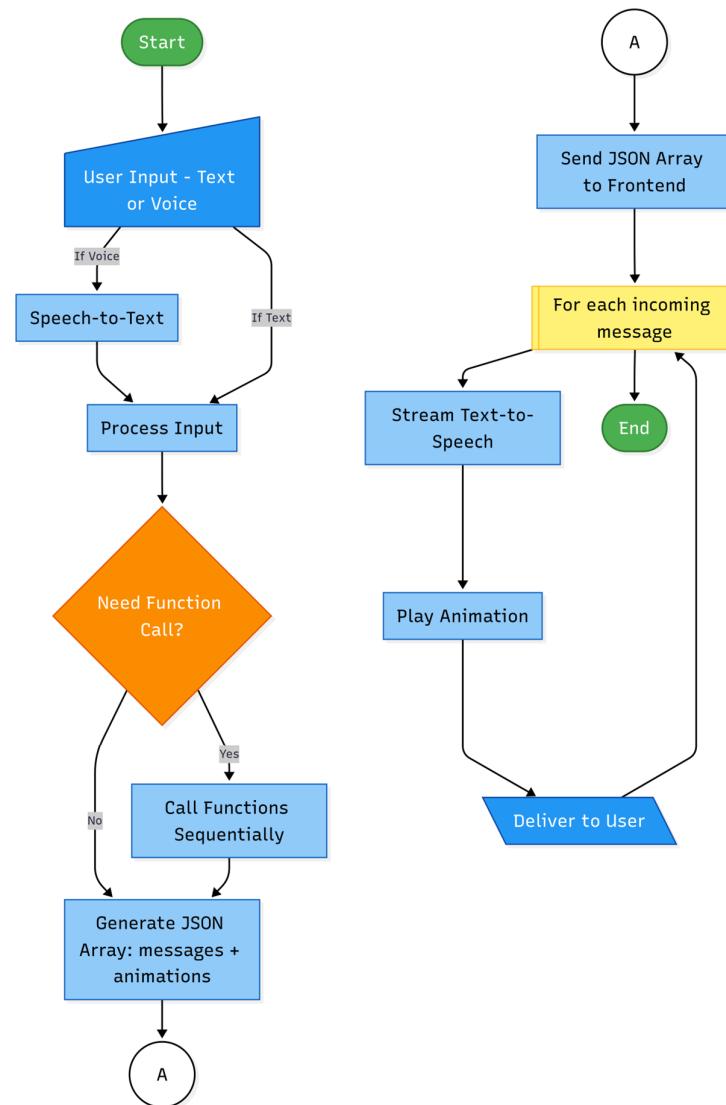
Role	Primary Objective	Key Functionalities
Researcher	To assist with academic research and writing tasks.	Bibliographic search, academic content with citations, PDF document formatting, and interactive querying of uploaded documents.
University Guide	To provide accurate and context-aware institutional information.	Answering queries based on official university documents, delivering information via institutional email, and providing campus navigation assistance.
Personal Skill Trainer	To provide personalized professional development and soft skills coaching.	Interview simulations, appearance analysis via computer vision, CV generation, HTML training reports, training history, and professional correspondence.
Personal Assistant	To integrate with and automate the user's digital productivity workflow.	Personal email management, inbox analysis, calendar reading/writing, contact search, news access, and weather information.
Receptionist	To manage campus resources, visitor assistance, and local information services.	Campus facilities information, contact directory, local event search, restaurant discovery, tourist guidance, email functionality, and commercial services info.

Next, the input text is processed by a Large Language Model (LLM), which determines whether the user's request requires the execution of a function or can be addressed with a standard response (pure text with no external capabilities). If a function is needed, the application handles it according to the specifications provided by the model. When multiple tools are required, they are executed sequentially. Once completed, the results are passed back to the LLM, which integrates them into a final, context-aware response. If no function is needed, the model simply generates a direct reply. (Further details on how this decision process is implemented are discussed in Section 3.3.1).

Regardless of whether functions are executed, the final output is a JSON array containing messages. Each message object consists of three components that together form the system's response: text, a facial expression identifier, and an animation identifier.

Then, each message is processed: the text is transformed into speech through the TTS system, while the digital avatar simultaneously displays the corresponding animation and facial expressions to the user.

This workflow is executed every time the user provides an input. From the initial capture of the request to the final delivery of text, voice, and avatar expressions, the process ensures consistency and control in the way NAIA responds. This cycle represents the core operational loop of the system and is repeated continuously with each user interaction.



**Figure 15.** NAIA's flow diagram.

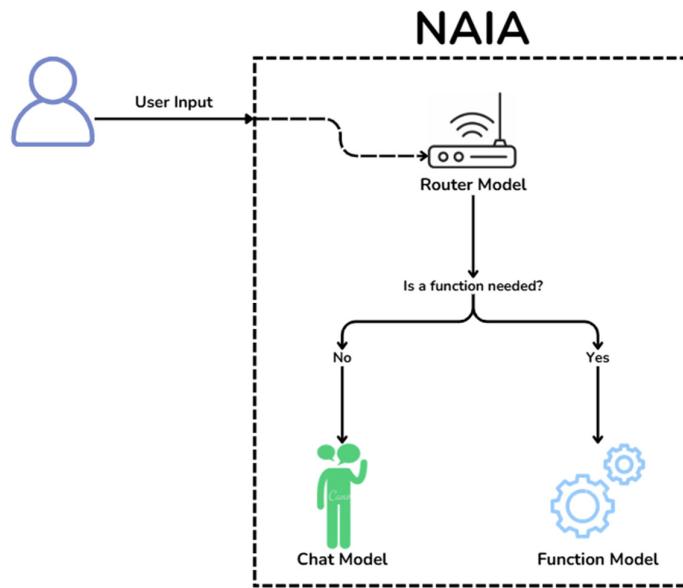
### 3.3.1. Decision-Making Process

NAIA's integration with Large Language Models (LLMs) does not rely on a single all-purpose model. Instead, a three-model architecture system is implemented to overcome the limitations of using just one LLM. The main issues encountered with a single-model approach are directly related to user experience. While some models excelled in response speed, this often comes at the cost of accurately understanding user requests and calling the appropriate functions.

Another recurrent problem is hallucination during long conversations that mix regular responses with function calls. In these cases, the model tends to stop executing functions and instead produces outputs such as "I will do it," while no function is generated. Al-

though this unreliability in function calling posed a challenge for NAIA, the short response time of such models remains highly desirable for pure conversational tasks.

The three-model architecture, shown in Figure 16, combines the strengths of the models available through the OpenAI API.



**Figure 16.** Three-model architecture diagram.

The architecture manages the user's request by using three sophisticated LLMs:

## 1. Router model

The router model serves as the initial decision-making component, determining the processing pathway for each user request. It is implemented as a lightweight model specialized in simple classification tasks. Its main role within the architecture is a binary classification that decides whether the user's input requires a function call or can be handled directly as text.

The router employs a specialized prompt that includes detailed descriptions of the functions available under the active NAIA role, enabling it to make context-aware decisions during the conversation. For this purpose, models such as GPT-4.1-nano and GPT-4o-mini are used, corresponding to those available at the time of development.

## 2. Function model

If the router determines that a function call is required, the user request is redirected to the function model. This component is implemented as a larger model specialized in instruction-following, ensuring that the correct functions are invoked with the appropriate parameters. Importantly, the function model operates strictly within the set of functions available under the active NAIA role, preventing calls to functions outside the current role's scope.

Beyond executing functions, this model is also responsible for generating the conversational response delivered to the user. By leveraging a specialized prompt that provides detailed context about each function, it not only announces the results of the executed functions but also explains them in a clear and natural way. This dual capability makes the function model central to NAIA's ability to handle complex interactions while maintaining an effective user experience. For this purpose, GPT-4.1 is used, as it is the model available at the time of development.

### 3. Chat model

When the router determines that no external functions are required, the request is redirected to the chat model. This component is slightly more complex than the router model but significantly lighter than the function model, striking a balance between speed and basic instruction-following capability.

The chat model is exclusively responsible for generating conversational responses, with no access to external tools or functions. Its role does not require the depth of understanding or reasoning of a large model; instead, it prioritizes responsiveness and fluency. Importantly, it is still capable of following simple instructions, such as delivering answers in the required JSON format previously described. By focusing on efficiency and clarity, the chat model ensures a smooth and natural user experience for interactions that remain purely conversational. For this purpose, GPT-4.1-mini is used, corresponding to the model available during development.

The three-model architecture provides NAIA with a balanced combination of efficiency, scalability, and user experience. By delegating the initial decision-making to a lightweight router model, the system optimizes computational costs and minimizes latency for simple requests. The chat model ensures natural and fluid conversations in cases where no external tools are needed, prioritizing responsiveness while maintaining contextual coherence. Meanwhile, the function model handles the most complex scenarios, orchestrating tool calls and delivering context-aware explanations of their results. Together, these three components create a specialized pipeline where each model is optimized for its specific role, classification, conversation, or tool integration, achieving both performance gains and a more reliable interaction flow.

#### 3.3.2. History Conversation Management

All responses, whether generated by the chat model or resulting from function executions, need to be stored to preserve the context of the interaction. In the first prototype [7], this is handled through two types of records: complete conversations and historical summaries. Complete conversations contained the full dialogue with the user, including both conversational responses and the results of executed functions, until a predefined token limit is reached. Once this happens, the conversation is achieved, and a summary is created to capture the essential information. This summary then served as the starting point for a new complete conversation, repeating the cycle whenever the limit is reached [7]. While this mechanism provided a basic way to manage context, in practice, it introduced additional complexity and did not add substantial value.

In the current version, NAIA simplifies this process by storing only one type of record: the conversation itself. For each user, only one conversation per day is kept, containing the entire interaction, including both dialogue and function outputs. When this conversation approaches the token limit, it is replaced by a summary that captures the most relevant information, and a new complete conversation begins from that summary. The cycle continues if the user keeps interacting, but at any given time, only a single conversation is stored per day. This design reflects the fact that retaining the full history of every interaction does not provide meaningful benefits, since users value NAIA's agentic capabilities, its ability to call functions and sustain natural dialogue, more than the retention of every past detail. A well-crafted summary is sufficient for the language model to preserve the relevant context for future interactions. To achieve this, the summarization process is guided by a specialized agent prompt that not only condenses the dialogue but also identifies and retains key details such as names, dates, numerical data, and specific tasks that may be important for the user to recall later.

### 3.3.3. Software Architecture

The comprehensive functionality and multimodal integration described in the previous sections require a robust software architecture that can efficiently manage complex interactions while maintaining code maintainability and scalability. Building upon the foundation established in NAIA's initial implementation, the current system represents an architectural evolution that enhances organization, modularity, and scalability to better support the platform's expanded capabilities.

The initial NAIA version successfully demonstrated the viability of the multimodal virtual assistant concept through a unified backend-frontend architecture that streamlined development and deployment. As the project matured and requirements expanded, the opportunity emerged to implement a more structured architectural approach. This evolution led to the adoption of modular monolithic architecture that facilitates the organization of specialized capabilities into distinct, manageable components while promoting code reusability and system extensibility within a unified deployment unit.

Figure 17 illustrates the comprehensive software architecture that underlies NAIA's modular design. The diagram demonstrates how domain-specific modules integrate with centralized services to create a cohesive system architecture that supports the platform's diverse functionalities.

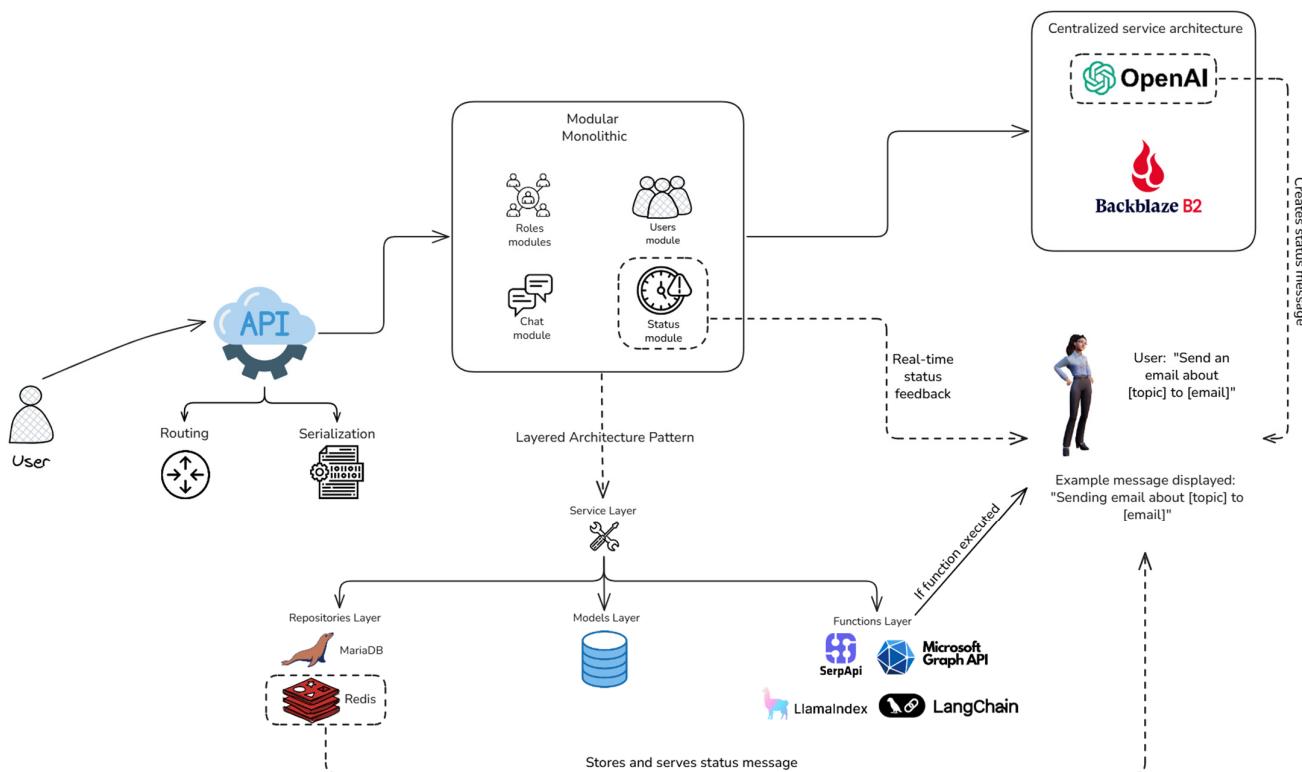
The software architecture follows a modular monolithic approach that organizes functionality into two types of modules: core system modules and specialized role modules. The core system modules include chat for managing all conversational flow and interactions across any role, users for handling user registration and user-related operations, and status for managing operational notifications during function execution. The specialized role modules correspond to NAIA's five distinct roles: researcher, university guide, skills trainer, personal assistant, and receptionist. Each module encapsulates domain-specific logic and functionality, enabling independent development and maintenance while ensuring clear separation of concerns across the platform's diverse capabilities.

Within each domain module, the system implements a consistent layered architecture pattern that promotes code organization and maintainability. The models layer defines data structures and database relationships specific to each domain. The service layer contains core business logic and coordinates interactions between different system components. The functions layer implements specialized utility operations, direct integrations with external APIs like Microsoft Graph API and SerpAPI, and manages access to vector databases through frameworks like LangChain and LlamaIndex. The repositories layer provides structured data access operations for database interactions.

The platform implements two primary centralized services that provide shared functionality across all domain modules through singleton patterns and dependency injection. OpenAI integration operates through a centralized service layer that abstracts API complexity from domain modules, implementing dynamic model routing, unified conversation state management, and consistent retry mechanisms across all role interactions. Backblaze B2 integration functions as a centralized storage abstraction that manages role-specific document repositories through standardized interfaces, enabling secure file operations and cross-module resource sharing while maintaining namespace isolation. Both services utilize dependency injections to eliminate direct API coupling within domain modules, enabling unified configuration management, consistent authentication handling, and simplified testing through service interface abstraction while preserving the modular independence that characterizes the overall architecture.

The RESTful API architecture organizes endpoints through a versioned structure with domain-specific routing that mirrors the modular application structure. Each specialized module handles its own HTTP request processing through dedicated view components

and manages data serialization within its domain boundaries. This distributed approach allows each module to maintain control over its specific API behavior and response formats while following consistent patterns across the system.



**Figure 17.** Software Architecture Overview: Modular Design and Centralized Service Integration.

The architecture implements sophisticated state management through a combination of persistent storage using MariaDB for relational data and Redis for high-performance caching operations. Conversation state is maintained through flexible data structures that store complex interaction histories, while user preferences and role configurations are managed through structured database relationships. Additionally, the system includes a lightweight status notification mechanism that provides user feedback during function execution, operating on a per-user, per-role basis and automatically clearing status indicators upon function completion.

### 3.4. Architectural Evolution from the Prior Prototype

The transition from the initial NAIA prototype to the present system involved a fundamental architectural reconfiguration informed by the insights obtained during the exploratory stage. The prototype validated the feasibility of combining speech processing, vision-based analysis, large language models, and avatar-mediated interaction within the same platform, while also highlighting the need for a more scalable and coherent processing pipeline capable of supporting expanded functionality. These observations motivated the introduction of a refined architecture centered on specialized agents, distributed function execution, and a structured multimodal workflow.

The new system consolidates three major architectural advancements. First, the role framework was expanded and redefined, allowing each agent to incorporate domain-specific capabilities, interaction patterns, and knowledge access pathways aligned with academic workflows. Second, the processing pipeline was reorganized through a modular LLM triad composed of a router model, a function-execution model, and a chat model, enabling more efficient request handling and reducing response delivery time without

compromising contextual reasoning. Third, the integration with institutional resources advanced from basic informational retrieval in the prototype to a fully interconnected ecosystem that leverages Microsoft Graph for email, calendar, and contact operations, as well as RAG-based access to official university documents stored in Backblaze B2.

Together, these changes establish a clear architectural distinction between the prototype and the current system. While the earlier version confirmed user interest in multimodal, role-based interaction, the present architecture transforms these foundations into a cohesive institutional framework with expanded capabilities, optimized response flow, and a scalable design suitable for real academic environments.

Unlike the initial prototype, which primarily served as a feasibility exploration, the current system establishes a coherent multimodal architecture, strengthened role specialization, and expanded institutional connectivity, positioning NAIA as a mature framework suited for practical deployment in academic environments.

### 3.5. Methodology for User Experience Evaluation

This study is primarily concerned with evaluating the user experience and perceived effectiveness of NAIA within the institutional context in which it has been conceived. NAIA has been designed and deployed under the institutional Uninorte IA initiative, which promotes the development of and evaluation of artificial intelligence solutions with students and staff from Universidad del Norte. Within this framework, the evaluation focuses on a concrete deployment of NAIA with a cohort of thirty (30) participants from the university community, reflecting the academic ecosystem targeted by the initiative.

The evaluation of NAIA's effectiveness within academic environments employs a comprehensive mixed-methods approach designed to assess both technical performance and user acceptance across diverse stakeholder groups. This methodology centers on the Technology Acceptance Model 2 (TAM2) framework, extended with TAM3 constructs for AI virtual assistant evaluation in educational contexts.

The Technology Acceptance Model (TAM) is one of the most influential frameworks utilized to evaluate the impact and user acceptance of new technologies [43]. TAM originally included five constructs: perceived usefulness (PU), perceived ease of use (PEOU), attitude toward using (ATU), behavioral intention to use (BI), and actual use. The TAM literature considers the removal of the construct attitude evolving to TAM2 and TAM3 takes the variables from TAM2, proposing other variables (e.g., computer self-efficacy) [44]. In this approach, the core TAM2/TAM3 constructs of PU, PEOU, BI, and Computer Self-Efficacy are implemented as the main lens for user experience and adoption analysis.

Although TAM2/TAM3 and the System Usability Scale (SUS) both address aspects of user evaluation, they capture complementary constructs. TAM2/TAM3 focuses on cognitive determinants of adoption, perceived usefulness, perceived ease of use, behavioral intention to use, and computer self-efficacy. Providing insight into the psychological factors shaping acceptance. In contrast, SUS yields a standardized global usability score that integrates ease of learning, efficiency, and satisfaction into a single metric widely used for benchmarking interactive systems. Combining both instruments enables an integrated assessment in which TAM2/TAM3 explains the underlying drivers of adoption, while SUS offers an interpretable, industry-recognized indicator of overall usability. This complementary structure makes the joint use of TAM2/TAM3 and SUS particularly suitable for evaluating generative AI assistants deployed in academic settings, where both cognitive acceptance factors and practical usability influence sustained adoption.

### 3.5.1. Participants and Sampling

The participant recruitment strategy employs purposive sampling within the Uninorte IA initiative to ensure representative diversity across multiple demographic and academic dimensions of Universidad del Norte's community. A cohort of thirty (30) participants is selected, including undergraduate students from Electrical and Electronic Engineering and Marketing programs, graduate students, and administrative staff. This sampling approach ensures adequate representation across gender distribution, age, technological proficiency levels, and academic disciplines to capture diverse user perspectives and needs within the academic environment. Participants are between 18 and 35 years old, and their distribution across age groups and institutional roles is summarized in Table 3.

**Table 3.** Cross-tabulation of Age Groups and User Type.

Age Group	Undergraduate	Graduate	Administrative Staff	Total
18–22	12 (80%)	2 (13%)	1 (7%)	15
23–27	5 (50%)	3 (30%)	2 (20%)	10
28–35	2 (40%)	2 (40%)	1 (20%)	5
Total	19 (63%)	7 (23%)	4 (14%)	30

The evaluation protocol involves a two-week interaction period with NAIA, during which participants engage with the system's multimodal capabilities and specialized agent functionalities. This extended interaction timeframe allows users to develop familiarity with NAIA's features while providing sufficient exposure to assess both immediate usability and sustained engagement patterns. The two-week period enables participants to integrate NAIA into their academic workflows naturally, providing authentic feedback on system utility and performance across different academic contexts and usage scenarios.

Participants interact with NAIA under naturalistic usage conditions, using their personal laptops or mobile devices through the university's official web-based interface. Access occurs through the institutional Microsoft environment, ensuring that interactions take place within the same operational context in which the assistant functions. Throughout the two-week period, participants engage with NAIA during their regular academic or administrative activities, exploring the five specialized roles and utilizing both voice and text modalities. Participants formulate queries relevant to their coursework, information needs, or job functions, enabling the evaluation to capture authentic perceptions of usability, usefulness, and conversational effectiveness within realistic academic workflows.

### 3.5.2. Data Collection Instrument

The comprehensive survey instrument integrates multiple validated measurement frameworks beyond the core TAM2/TAM3 construct. The System Usability Scale (SUS) complements TAM measurements with standardized usability assessment questions rated on five-point scales, providing an established benchmark for system usability evaluation in educational technology contexts [45]. Additionally, the instrument incorporates demographic data collection to ensure representative sampling across stakeholder categories, academic disciplines, and technological proficiency levels.

The evaluation instrument is deployed through Universidad del Norte's institutional Microsoft Forms infrastructure, leveraging the university's Microsoft tenant to ensure secure data collection and institutional compliance. This deployment approach maintains data governance standards while providing participants with familiar, accessible survey interfaces through their existing institutional credentials.

Technical component evaluation sections assess NAIA-specific functionalities, including voice quality assessment, response appropriateness evaluation, conversational

naturalness metrics, multimodal interaction capabilities, and role-specific performance across the five specialized agents. The instrument employs multiple rating scales to capture nuanced user perceptions: seven-point Likert scales for TAM3 constructs and technical component evaluation, five-point scales for SUS and efficiency measures, and Net Promoter Score methodology for recommendation likelihood assessment.

The survey structure encompasses ten comprehensive sections: demographic information collection, previous technology experience assessment, NAIA usage patterns documentation, TAM3 construct measurement, SUS evaluation, role-specific performance assessment, technical component evaluation, efficiency and productivity impact measurement, satisfaction and recommendation metrics, and open-ended feedback collection. This multi-dimensional approach enables both quantitative statistical analysis and qualitative insight generation to provide comprehensive system evaluation across technical performance and user acceptance dimensions.

## 4. Results

As mentioned before, the development of NAIA had a previous evaluation phase that established baseline performance metrics. The initial assessment involved 30 participants from Universidad del Norte, each interacting with one of NAIA's specialized roles for 30 min. The quantitative results showed promising user acceptance: an overall satisfaction score of 4.27 out of 5, with 90% of participants reporting that NAIA-assisted tasks have superior quality compared to traditional methods [7].

However, the qualitative analysis revealed important areas for improvement, including inconsistent platform performance, voice recognition instability, and accessibility limitations, particularly on mobile devices.

The current evaluation addresses these identified limitations through enhanced system architecture and expanded assessment protocols. The following analysis presents comprehensive quantitative results from the improved NAIA implementation, examining user experience across demographic profiles, technology acceptance dimensions, and system usability metrics. Additionally, a detailed comparison with baseline performance indicators demonstrates the measurable improvements achieved through systematic technological enhancements and refined user interaction capabilities.

### 4.1. Quantitative Analysis of User Experience

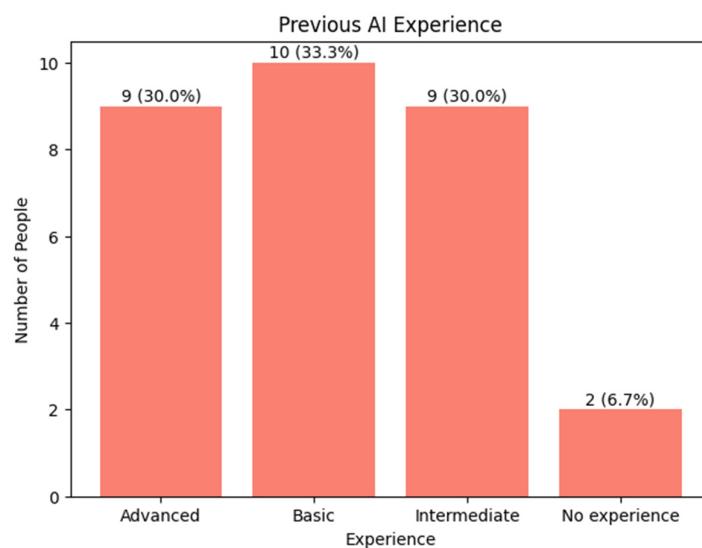
This section breaks down the quantitative findings into three main parts. First, a demographic analysis of the participants is presented to provide context on the user base that evaluated the platform. Following this, the analysis explores user perceptions through two established frameworks: the Technology Acceptance Model (TAM) and the System Usability Scale (SUS), which measure the system's perceived usefulness and ease of use. Finally, these results are compared against baseline metrics from the initial prototype to quantitatively demonstrate the system's evolution and improvements.

#### 4.1.1. Demographic Analysis

The demographic distribution highlights that most of NAIA's participants are undergraduate students, with notable representation across different age groups. Table 3 presents the cross-tabulation between age ranges and user types, showing that younger participants (18–22) are predominantly undergraduates, while older participants are more evenly distributed between graduate students and administrative staff. This distribution is particularly relevant for NAIA, as it demonstrates that the platform is evaluated by users with diverse academic roles and stages of technical adoption. By engaging both younger and older groups, the study ensures that NAIA's usability and acceptance are not

limited to a single demographic profile but extend across different types of users within the university ecosystem.

The analysis of participants' prior experience with artificial intelligence tools reveals a diverse background across the sample of 30 users. As shown in Figure 18, approximately one-third of respondents reported having basic experience (33.3%), while a similar proportion indicated an advanced (30.0%) or intermediate (30.0%) level of experience. Only a small fraction of participants reported having no prior experience with AI tools (6.7%). This distribution indicates that NAIA is evaluated by users with varying levels of technological expertise, ensuring that feedback is not limited to expert users but also incorporates the perspectives of those with limited or no prior exposure. Such diversity reinforces the robustness of the evaluation, as it demonstrates NAIA's capacity to be perceived as useful and accessible across heterogeneous user profiles.



**Figure 18.** Previous AI Experience.

The cross-tabulation of age groups and time dedicated to NAIA reveals important differences in engagement patterns. As shown in Table 4, the youngest participants (18–22 years) predominantly reported short interaction times, with 80% indicating 30–60 min, and only a minority extending beyond 2 h. In contrast, participants aged 23–27 years reported longer sessions, with 70% dedicating more than one hour and nearly one-third engaging for more than five hours. The oldest group (28–35 years) displayed a balanced distribution, splitting between 30–60 min and 1–2 h evenly, with some extending up to 2–5 h. These findings reinforce that while NAIA is frequently adopted in shorter sessions by younger undergraduates, older participants engaged in longer and more sustained interactions, suggesting differentiated usage behaviors depending on academic maturity and user profile.

**Table 4.** Cross-tabulation of Age Group and Time Dedicated to NAIA.

Age Group	30–60 min	1–2 h	2–5 h	5+ h	Total
18–22	12 (80%)	2 (13%)	1 (7%)	0 (0%)	15
23–27	0 (0%)	4 (40%)	3 (30%)	3 (30%)	10
28–35	2 (40%)	2 (40%)	1 (20.0%)	0 (0%)	5
Total	14	8	5	8	30

#### 4.1.2. Technology Acceptance Model (TAM) Analysis

As previously described in the methodological framework, the Technology Acceptance Model (TAM) is incorporated into the evaluation instrument to assess the users' perceptions of NAIA. In this section, the results of the TAM constructs as applied to this study are presented. The following tables and analysis illustrate how participants evaluated NAIA in terms of perceived usefulness, ease of use, behavioral intention, and technological self-efficacy, thereby providing a structured perspective on the system's acceptance across its core dimensions.

The results for perceived usefulness demonstrate that NAIA is strongly regarded as beneficial for academic and work-related activities. As shown in Table 5, 83.3% of participants agree or strongly agree that NAIA improves their academic/work performance, 86.7% report that it increases their productivity, and 93.3% indicate that it enhances the effectiveness of their tasks. Most notably, 93.3% consider NAIA overall useful for their activities, underscoring its strong perceived value. Strong agreement is especially evident in task effectiveness, where 20% of respondents select "Strongly Agree." Neutral responses remain minimal (up to 6.7%), and negative evaluations are virtually absent. These results highlight NAIA as a reliable academic support assistant that drives measurable improvements in productivity, task effectiveness, and overall usefulness.

**Table 5.** Distribution of Responses—Perceived Usefulness.

	Strongly Disagree	Disagree	Somewhat Disagree	Neutral	Somewhat Agree	Agree	Strongly Agree
Using NAIA improves my academic/ work performance	0 (0%)	0 (0%)	0 (0%)	0 (0%)	5 (16.7%)	22 (73.3%)	3 (10.0%)
Using NAIA increases my productivity	0 (0%)	0 (0%)	0 (0%)	0 (0%)	4 (13.3%)	24 (80%)	2 (6.7%)
Using NAIA improves the effectiveness of my tasks	0 (0%)	0 (0%)	1 (3.3%)	1 (3.3%)	0 (0%)	22 (73.3%)	6 (20%)
Overall, I find NAIA useful for my activities	0 (0%)	0 (0%)	0 (0.0%)	2 (6.7%)	0 (0.0%)	25 (83.3%)	3 (10%)

The results for Perceived Ease of Use confirm NAIA's intuitive design and accessibility. As shown in Table 6, 93.3% of participants agree or strongly agree that NAIA is easy to learn to use, and the same percentage indicates that it is easy to use overall. Similarly, 86.6% perceive the interaction with NAIA as clear and understandable, while 60% affirm that it is easy to become skilled in using the system. Neutral responses remain relatively low (between 3% and 10%), and disagreement is almost negligible (<3.3%). These findings emphasize that NAIA eliminates barriers to adoption and stands out as a user-friendly educational technology that facilitates learning and sustained usage.

**Table 6.** Distribution of Responses—Perceived Ease of Use.

	Strongly Disagree	Disagree	Somewhat Disagree	Neutral	Somewhat Agree	Agree	Strongly Agree
NAIA is easy to learn to use	0 (0%)	0 (0%)	0 (0%)	1 (3.3%)	1 (3.3%)	8 (26.6%)	20 (66.6%)
Interaction with NAIA is clear and understandable	0 (0%)	0 (0%)	1 (3.3%)	0 (0%)	3 (10%)	17 (56.6%)	9 (30.0%)
NAIA is easy to use	0 (0%)	0 (0.0%)	0 (0%)	1 (3.3%)	1 (3.3%)	5 (16.6%)	23 (76.6%)
It is easy to become skilled in using NAIA	0 (0%)	0 (0%)	1 (3.3%)	4 (10%)	10 (16.7%)	14 (56.7%)	1 (3.3%)

The results for Behavioral Intention strongly emphasize NAIA's potential for sustained adoption and dissemination. As presented in Table 7, 100% of participants intend to continue using NAIA in the future, and the same unanimous support is observed in the willingness to recommend NAIA to others. In addition, 93.3% report that they plan to use NAIA regularly, further reinforcing its role in academic and work-related contexts. Neutral or negative responses are completely absent across all items, which underscores a consistently positive perception. These findings clearly demonstrate NAIA's capacity to generate long-term engagement, positive adoption intentions, and strong peer recommendations.

**Table 7.** Distribution of Responses—Behavioral Intention.

	Strongly Disagree	Disagree	Somewhat Disagree	Neutral	Somewhat Agree	Agree	Strongly Agree
I intend to continue using NAIA in the future	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	25 (83.3%)	5 (16.7%)
I plan to use NAIA regularly	0 (0%)	0 (0%)	0 (0%)	0 (0%)	2 (6.7%)	18 (60%)	10 (33%)
I would recommend NAIA to others	0 (0%)	0 (0%)	0 (%)	0 (0%)	0 (0%)	23 (76.7%)	7 (23.3%)

The results for Technological Self-Efficacy indicate that users feel confident in their ability to operate NAIA effectively and independently. As shown in Table 8, 86.6% of participants agree or strongly agree that they feel confident using technologies like NAIA, while 90% report that they can use NAIA without assistance. Similarly, 90% state that they possess the necessary skills to use the system. Neutral responses remain minimal (3.3–6.7%), and disagreement is nonexistent across all items. These findings highlight that NAIA does not demand excessive technological proficiency, supporting its accessibility and ensuring successful adoption across diverse user profiles.

**Table 8.** Distribution of Responses—Technological Self-Efficacy.

	Strongly Disagree	Disagree	Somewhat Disagree	Neutral	Somewhat Agree	Agree	Strongly Agree
I feel confident using technologies like NAIA	0 (0%)	0 (0%)	0 (0%)	1 (3.3%)	3 (10%)	16 (53.3%)	10 (33.3%)
I can use NAIA without assistance	0 (0%)	0 (0%)	0 (0%)	1 (3.3%)	2 (6.7%)	17 (56.7%)	10 (33.3%)
I have the necessary skills to use NAIA	0 (0%)	0 (0%)	0 (0%)	1 (3.3%)	2 (6.7%)	15 (50%)	12 (40%)

To estimate how the observed proportions from the 30-user sample would behave in a larger population, a binomial proportion analysis is performed following the method described by [46]. The 95% confidence interval (CI) for each proportion is computed using the following formula:

$$\hat{p} \pm z \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}} \quad (1)$$

where the following applies:

- $\hat{p}$  = observed sample proportion (e.g., 0.93 for 93%);
- $z = 1.96$ , corresponding to a 95% confidence level;
- $n = 30$ , representing the total number of participants.

This expression provides the range within which the true population proportion ( $p$ ) is expected to fall with 95% confidence. For instance, if 93.3% of participants (in the case

of the ‘Using NAIA improves the effectiveness of my tasks’ variable) agreed with that statement, the confidence interval would be calculated as follows:

$$0.933 \pm 1.96 \sqrt{\frac{0.933(1 - 0.933)}{30}} = 0.933 \pm 0.089$$

Therefore, the 95% confidence interval is [0.844, 1.000].

When extrapolated to a population of 500 users, this interval corresponds to an estimated range between 422 and 500 users who are likely to express agreement under similar conditions.

Table 9 summarizes the relationship between each Technology Acceptance Model (TAM) construct and its corresponding observed proportions, together with the 95% confidence intervals and their extrapolated estimates for a population of 500 users. This table provides a quantitative overview of how users’ perceptions—across dimensions such as Perceived Usefulness (PU), Perceived Ease of Use (PEOU), Behavioral Intention (BI), and Technological Self-Efficacy (TSE)—are expected to scale in larger populations. The extrapolated values demonstrate that the observed trends remain consistent, reinforcing the stability and generalizability of the results obtained from the 30-user sample.

**Table 9.** Extrapolated TAM Constructs and Confidence Intervals Supporting NAIA’s User Acceptance Validation.

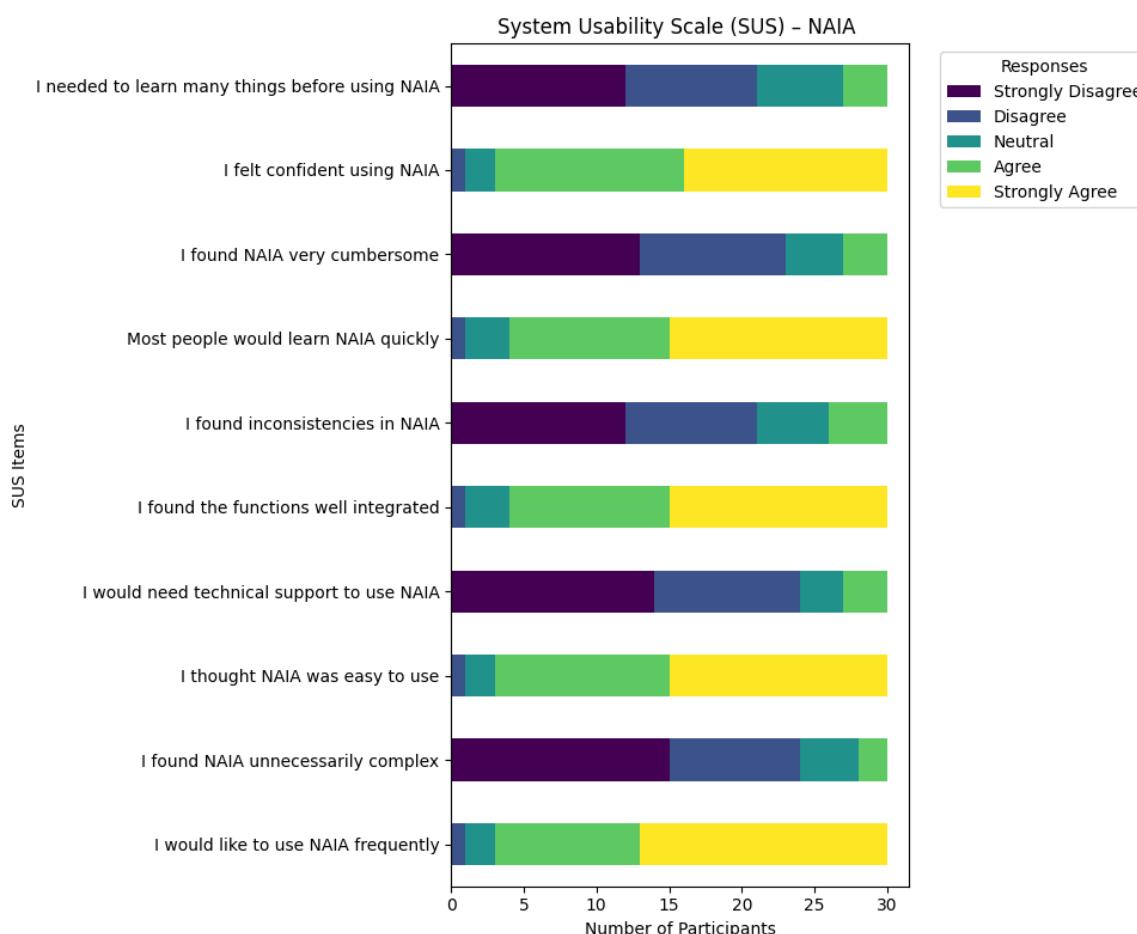
Construct	Item	$\hat{p}$ (Observed%)	95% CI (Proportion)	95% CI (Users out of 500)
Perceived Usefulness	Using NAIA improves the effectiveness of my tasks	83.3%	0.700, 0.966	350–483
Perceived Usefulness	Overall, I find NAIA useful for my activities	93.3%	0.844, 1.000	422–500
Perceived Ease of Use	NAIA is easy to learn to use	93.3%	0.844, 1.000	422–500
Perceived Ease of Use	Interaction with NAIA is clear and understandable	86.6%	0.744, 0.988	372–494
Behavioral Intention	I intend to continue using NAIA in the future	100%	1.000, 1.000	500–500
Behavioral Intention	I would recommend NAIA to others	100%	1.000, 1.000	500–500
Technological Self-Efficacy	I can use NAIA without assistance	90.0%	0.793, 1.000	396–500
Technological Self-Efficacy	I have the necessary skills to use NAIA	93.3%	0.844, 1.000	422–500

This analysis confirms that the main Technology Acceptance Model constructs, Perceived Usefulness (PU), Perceived Ease of Use (PEOU), and Behavioral Intention (BI) maintain stable proportional relationships even when projected to larger populations, supporting the generalizability of the results.

#### 4.1.3. System Usability Scale (SUS) Analysis

Building on the insights provided by the TAM constructs, where NAIA demonstrated strong perceptions of usefulness, ease of use, behavioral intention, and self-efficacy, we further examined the system’s usability through the System Usability Scale (SUS). While TAM focuses on user acceptance and intention to adopt the technology, SUS provides a standardized benchmark for evaluating overall usability and user experience. The following results illustrate how participants rated NAIA across the ten SUS items, offering complementary evidence that reinforces and extends the TAM findings.

The SUS analysis confirms NAIA's strong usability and reinforces the findings from the TAM constructs. As shown in Figure 19, responses to the positive SUS items (e.g., willingness to use NAIA frequently, ease of use, integration of functions, confidence, and quick learnability) cluster overwhelmingly around Agree and Strongly Agree, with agreement levels consistently above 85%. Conversely, the negative items (e.g., complexity, inconsistencies, cumbersomeness, and need for technical support) concentrate predominantly in Disagree and Strongly Disagree, with more than 75% of participants rejecting these potential barriers. Neutral responses remain low, and disagreement on positive items is virtually absent. Taken together, these results demonstrate that NAIA is not only perceived as useful and easy to use but also as a system that can be adopted smoothly without requiring significant technical assistance. Overall, the SUS outcomes place NAIA well within the "Excellent" usability range, validating its role as an accessible, user-friendly academic assistant.



**Figure 19.** System Usability Scale (SUS) response distribution for NAIA.

#### 4.1.4. Comparative Analysis with Previous NAIA Evaluation

To assess the evolutionary impact of NAIA's enhanced implementation, this section presents a quantitative comparison between the current evaluation results and the baseline metrics established in the initial NAIA assessment. The comparison focuses on key performance indicators across Technology Acceptance Model constructs, user satisfaction metrics, and system usability measures to demonstrate the measurable improvements achieved through the system's technological and architectural enhancements.

Technology Acceptance Model (TAM) metrics reveal enhanced measurement precision and greater participant diversity, providing more reliable adoption indicators compared to the baseline evaluation. While the initial study reported 100% recommendation rates

within a homogeneous group of users, the current assessment incorporates a broader demographic representation, including undergraduate students, graduate students, and administrative staff. This diversification enables more granular TAM measurements across stakeholder groups. Current results show that 83.3% of participants agree or strongly agree that NAIA improves academic/work performance, 86.7% confirm productivity gains, and 93.3% report that NAIA improves task effectiveness. Most notably, 93.3% consider NAIA overall useful for their activities. Regarding ease of use, 93.3% find NAIA easy to learn and to use, 86.6% perceive the interaction as clear and understandable, and 60% indicate that it is easy to become skilled in using the system. These findings reflect authentic adoption patterns across varied academic roles, moving beyond the uniformly positive but less diverse feedback captured in the baseline.

User satisfaction metrics reveal a more nuanced performance profile when compared to baseline measurements. The previous evaluation reported an overall satisfaction score of 4.27 out of 5, with 100% of participants indicating complete needs fulfillment and universal recommendation intentions. In contrast, the current study shows 100% of participants intend to continue using NAIA in the future, with 100% willing to recommend NAIA to others, and 93.3% plan to use it regularly. While the baseline study achieved 90% agreement on task quality improvement, current results demonstrate 86.7% agreement on productivity gains and 83.3% on performance improvement, reflecting the implementation of more stringent evaluation criteria and extended participant engagement periods.

Technological self-efficacy outcomes also highlight significant improvements. A large majority of participants report confidence in using technologies like NAIA (86.6%), the ability to use NAIA without assistance (90%), and possession of the necessary skills to operate the system effectively (93.3%). Neutral responses remain minimal (3.3–6.7%), and no disagreement was recorded. These results confirm that NAIA supports diverse user groups without requiring excessive technological proficiency, ensuring broad accessibility and adoption.

System Usability Scale (SUS) analysis complements the TAM results by providing a standardized usability benchmark. Responses to positive items (e.g., frequent use, ease of use, function integration, confidence, quick learnability) concentrate overwhelmingly in Agree and Strongly Agree, consistently above 85%. Negative items (e.g., complexity, inconsistency, cumbrousness, need for technical support) are concentrated in Disagree and Strongly Disagree, with more than 75% of participants rejecting these potential barriers. Neutral responses remain low, and disagreement on positive items is virtually absent. These results position NAIA within the “Excellent” usability range, reinforcing its role as an accessible, user-friendly academic assistant.

Efficiency and productivity assessments also demonstrate methodological maturation. While the baseline study produced notably high efficiency metrics under short, controlled sessions (e.g., 96% agreement on NAIA assisting with tasks, 90% agreement on higher task quality, 83.3% reporting performance improvements), the current evaluation offers more sustainable indicators gathered across extended two-week integration periods with a more diverse participant base. This methodological shift provides a more realistic view of NAIA’s effectiveness within authentic workflows, emphasizing sustained productivity rather than immediate impressions.

Overall, the comparative analysis demonstrates that NAIA has evolved from a promising prototype into a robust, methodologically validated academic assistant. While baseline results reflect universal enthusiasm under controlled conditions, the current evaluation highlights strong, sustainable performance across real-world academic environments. By resolving prior technical limitations—such as platform consistency, voice recognition stability, and accessibility—NAIA now delivers improved usability, enhanced confidence, and

reliable adoption indicators. Most importantly, the transition to extended integration periods enables more authentic evaluation of NAIA's practical utility, supporting its sustained adoption across diverse academic roles while maintaining strong user acceptance.

#### 4.2. Qualitative Analysis of User Experience

To complement the quantitative data, this section presents a thematic analysis of qualitative feedback obtained through an in-depth focus group approach. A selected subset of participants, specifically Marketing students from the university, participated in an extended qualitative evaluation process. This group attended an initial briefing session where the research team provided a comprehensive introduction to NAIA's platform and capabilities. Following the two-week interaction period, these participants not only completed the standard survey instrument but also provided direct feedback to the development team through structured qualitative methods. The students employed various analytical strategies, including direct observation, individual experience logs, Customer Journey Maps, Empathy Maps, and facilitated group discussions, to capture comprehensive insights into their user experience.

The thematic analysis of this qualitative feedback revealed several key themes regarding user perceptions of the platform across the extended interaction period.

##### 4.2.1. Perceived Strengths and Value

Across the evaluation groups, NAIA is consistently perceived as an innovative, practical, and useful tool with high potential for the university environment. The thematic analysis revealed three primary dimensions of value that participants identified throughout their interaction with the platform.

##### Core Functionality and System Integration

The multi-role architecture of NAIA emerged as one of its most valued characteristics among participants. In the role of Personal Assistant, NAIA is recognized for its utility in organizing academic and personal activities, such as managing emails, creating reminders, and administering schedules. Students reported that these organizational capabilities helped them maintain better control over their academic responsibilities, reducing the cognitive burden of tracking multiple commitments across different platforms.

As a Researcher, NAIA is seen as a valuable academic support tool for finding information, drafting texts, and generating content with citations and references. Participants particularly appreciated the platform's ability to assist with initial research phases, providing structured starting points for academic assignments and helping them identify relevant sources. The integration of citation generation was highlighted as a practical feature that streamlined their academic writing workflow.

The University Guide and Receptionist roles are highlighted for providing complete and reliable information about university services, events, and campus locations. Students noted that having immediate access to institutional information eliminated the need to navigate multiple websites or contact different administrative offices, making NAIA a convenient first point of contact for campus-related queries.

Furthermore, a significant value proposition highlighted by participants is the platform's ability to centralize these diverse functions into a single, cohesive system. This integration is seen as a key factor in positioning NAIA as a comprehensive support tool for university life, capable of managing everything from academic research and personal scheduling to campus navigation. Users reported that this centralization provided concrete benefits by streamlining the organization of their academic life and alleviating their mental load. Rather than switching between multiple applications and platforms, students

appreciated having a unified interface that could address varied needs throughout their academic day.

#### User Interface and User Experience Design

The visual and interactive design elements of NAIA received consistently positive feedback from participants. The platform is noted for having a simple and user-friendly interface that facilitates intuitive navigation without requiring extensive learning curves. Participants appreciated that they could begin using the platform effectively with minimal instruction, suggesting that the interface design successfully prioritizes accessibility and ease of use.

The visual design of the 3D avatars is frequently described as modern, attractive, and aesthetically pleasing, which generated positive engagement and satisfaction. Students expressed that the avatar representation added a distinctive character to the platform, differentiating it from conventional chatbot interfaces. The visual presence of the avatars was noted to make interactions feel more engaging and personalized, contributing to a more immersive user experience. Several participants mentioned that the aesthetic quality of the avatars reflected positively on the platform's overall professionalism and attention to design detail.

The combination of functional clarity and visual appeal in the interface design contributed to high levels of user satisfaction, with participants indicating that the platform was pleasant to use and that the design choices supported rather than hindered their interaction goals.

#### Emotional Engagement and Institutional Connection

Beyond functional utility, the qualitative feedback revealed significant emotional and community-building dimensions to the NAIA experience. Participants reported forming a positive connection with the tool, expressing feelings of curiosity and enthusiasm throughout the evaluation period. This emotional engagement manifested not only in sustained platform usage but also in participants' eagerness to explore different features and roles within NAIA.

A particularly notable finding was the sense of pride and motivation derived from using an innovative project developed at their own university. Students expressed appreciation for being part of a cutting-edge initiative within their institutional context, which generated trust and a sense of community. This institutional connection appears to have strengthened participants' investment in the platform's success and their willingness to engage deeply with its capabilities. Multiple participants noted that knowing NAIA was developed locally made them feel they were contributing to the advancement of their university's technological ecosystem.

The emotional resonance of the platform extended to creating a sense of ownership and belonging, with students viewing NAIA not merely as a tool but as a representation of their university's innovative capacity. This institutional pride factor emerged as an unexpected but significant contributor to user satisfaction and may represent an important element in NAIA's adoption and long-term engagement within the university community.

## 5. Discussions

When comparing NAIA to existing virtual assistants presented in the literature, several distinctive evolutionary contributions emerge that address limitations identified in previous implementations while building upon their proven approaches.

### 5.1. Comparison with University-Focused Systems

Unlike systems such as UBOT [27], which improve institutional FAQ response times through purely text-based interfaces, NAIA integrates synchronous voice, image, and text processing to deliver five role-specialized agents within a single platform. While UBOT demonstrates effectiveness in providing contextually relevant responses using institution-specific information, its limitation to text-based interaction revealed the opportunity for enhanced user engagement through multimodal interfaces. NAIA addresses this limitation through its integration of voice, vision, and gesture recognition to create more immersive academic interactions.

Similarly, while Edith [36] demonstrates multi-interface deployment capability and efficient resolution of academic process queries, its single-role approach highlights the potential for role specialization. NAIA's five-role architecture addresses this by maintaining specialized capabilities for specific academic functions rather than attempting universal coverage.

### 5.2. Advancement over Specialized Domain Applications

The museum-oriented Alice Assistant [28,29] enables natural speech communication and demonstrates effective voice-based interactions but lacks adaptive conversational memory. NAIA addresses this limitation through an extended-context LLM combined with Redis caching, sustaining multi-session dialogue even under high token loads. Additionally, while Alice's success in creating intuitive visitor interactions validates voice-to-voice interaction flow, Alice's single-domain focus revealed the potential for multi-domain capability within a unified platform, a concept that became central to NAIA's multi-role architecture.

### 5.3. Enhanced Multimodal Integration

Fernández and Venezuela [37] provided early validation for multimodal approaches by demonstrating that computer vision enhances academic assistance beyond traditional text-based interactions. NAIA extends this concept by incorporating real-time visual context awareness across all roles, enabling natural comments about user appearance and surroundings to enhance interaction immersiveness.

### 5.4. Improvements over Multi-Domain Systems

YIO's context-aware agent [30] spans multiple domains but remains unimodal, whereas NAIA's computer-vision layer supplies visual context that dynamically influences responses across all roles. While YIO's versatility in handling multiple domains without specialized configurations provides validation for multi-role concepts, YIO's general-purpose approach revealed the opportunity for deeper specialization, leading to NAIA's design, where each role maintains dedicated knowledge bases and function sets optimized for specific academic tasks.

### 5.5. Knowledge Integration Advancements

NAIA's implementation differs from knowledge graph-based approaches like KBot [41] and AIDA-Bot [42] by emphasizing multimodal interaction (voice, vision, and text) rather than focusing solely on structured knowledge representation. While KBot demonstrates improved accuracy through structured knowledge integration and AIDA-Bot shows capability in academic research assistance, both systems' text-only interaction revealed the opportunity for enhanced user engagement through multimodal interfaces. NAIA addresses this by integrating visual context awareness alongside structured academic information.

### 5.6. Architectural Innovations

The system's technical architecture builds upon the strengths of existing systems while addressing their limitations through innovative integration. NAIA combines the institutional knowledge approach of UBOT [27] with multimodal capabilities demonstrated by Fernández and Venezuela [37], while implementing role specialization concepts validated by healthcare applications [33,34] and museum systems [28,29]. This synthesis creates a comprehensive academic assistant that maintains the depth of specialized systems while providing the versatility of multi-domain platforms.

The multimodal integration in NAIA represents an evolution beyond existing approaches by implementing simultaneous voice, vision, and text processing rather than the sequential processing typical of previous systems. This enhancement, combined with role-specific function calling capabilities, transforms NAIA from a purely conversational system into an actionable academic assistant capable of generating meaningful results across diverse academic workflows.

### 5.7. Tool Limitations, Challenges, and Future Directions for Generative AI Evaluation

Although NAIA demonstrates stable performance across modalities, its main limitation lies in its dependence on external AI services such as OpenAI, Microsoft Graph, and SerpAPI. These services enable advanced functionality but reduce institutional control over model updates, availability, and long-term governance. This dependency does not affect NAIA's current operation, but it highlights the importance of designing the platform to remain adaptable as generative AI ecosystems evolve.

A core technical challenge involves the coordination of multi-step tasks through chained function calls. NAIA successfully manages this through iterative execution loops, yet this sequential design may not be the most efficient approach for increasingly complex workflows. Future work should explore more direct mechanisms for inter-function communication or parallel execution strategies.

Looking ahead, NAIA's modular and decoupled architecture makes the system highly adaptable to the rapid evolution of generative AI technologies. Because each component operates independently, new capabilities such as WebRTC-based low-latency communication, audio-to-audio interaction, asynchronous function execution through MCP servers, and enhanced scalability monitoring can be integrated without requiring major structural modifications. The adaptability of the platform was demonstrated during the initial exploration of emerging real-time speech technologies, where the migration process required only targeted adjustments rather than a full architectural redesign. This flexibility naturally anticipates the opportunities discussed in Section 5.8, where real-time conversational models broaden the potential for improved responsiveness and multimodal interaction fluidity.

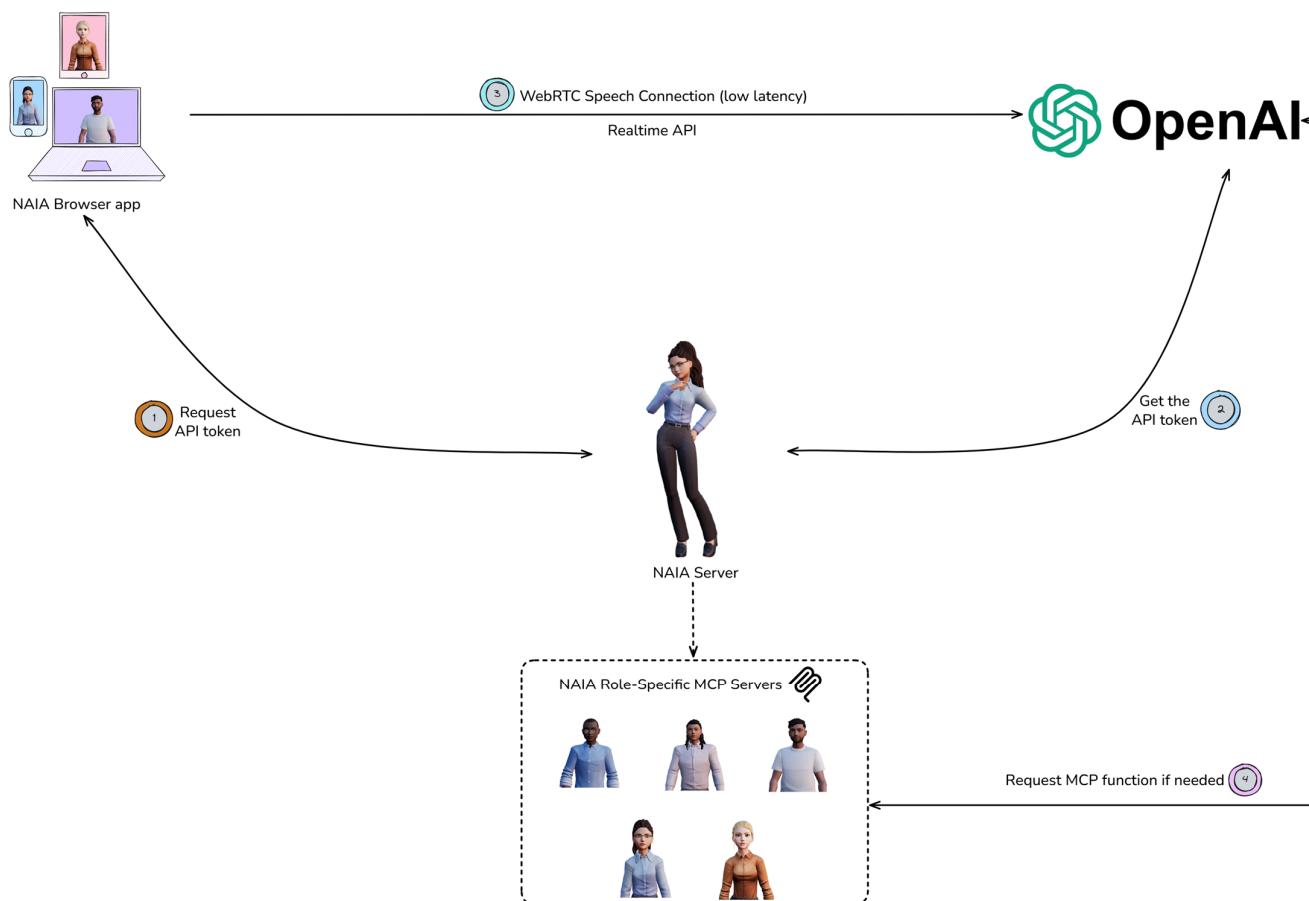
### 5.8. Emerging Real-Time Speech Technologies and Implications for Future Development

During the final stages of this research, OpenAI released significant updates to their "Realtime API", introducing capabilities specifically designed for low-latency conversational applications. This API enables direct speech-to-speech interactions without requiring separate STT or TTS components, as all audio processing is managed internally by OpenAI's infrastructure. The update introduced the "gpt-realtime" model, which demonstrates advanced capabilities including understanding of non-verbal audio cues, dynamic tone adaptation, asynchronous function calling, and support for Model Context Protocol (MCP) servers.

The current NAIA architecture takes between 3 and 5 s to deliver pure conversational responses. While this may appear reasonable, user feedback indicates a clear preference for faster interaction speeds, highlighting an area where the Realtime API could provide signif-

ificant improvements. The Realtime API offers two implementation approaches: WebSocket connections and WebRTC integration. Given NAIA's web-based architecture, WebRTC integration would be the optimal choice, enabling persistent, low-latency connections with OpenAI servers. This approach could deliver near-instantaneous responses while supporting voice-based interruption capabilities, allowing users to interact naturally during NAIA's speech. Additionally, the asynchronous function calling capability would prevent long-running operations from blocking conversational flow, enabling users to continue interacting while awaiting function results.

Moreover, the conversation state is fully controlled by OpenAI. Thus, it is not necessary to implement server-side modules mentioned before, such as the chat module, because tasks like maintaining the current context conversation are covered by the Realtime API. Since using WebRTC means that the connection to the API needs to be on the client side (frontend), the function calling capability must be restructured because all functions on NAIA are designed to run on the server side. Thanks to the update, this problem is easy to solve by developing MCP servers for each role; in this way, a safe platform is guaranteed by maintaining all function execution on the NAIA server. Figure 20 illustrates the proposed architectural integration of NAIA with OpenAI's Realtime API, demonstrating the WebRTC-based client-side connection, MCP server implementation for secure function execution, and the direct speech processing pipeline that achieves sub-second response times.



**Figure 20.** NAIA Realtime API Integration Architecture.

However, this architectural transition introduces significant context management challenges. The gpt-realtime model operates with a 32,000-token context window, which is considerably smaller compared to contemporary LLM standards. This limitation becomes more pronounced when incorporating NAIA's multimodal capabilities, as image processing

consumes additional token capacity. Furthermore, the Realtime API deviates from conventional message array structures used by traditional LLM APIs, requiring novel approaches to conversation history management. Long-term memory mechanisms need careful examination and optimization to ensure comprehensive context retention without exhausting the available token space required for active conversation flow. This constraint necessitates the development of intelligent context summarization strategies that preserve essential conversational elements while maintaining sufficient capacity for real-time interaction.

Another significant challenge emerges in avatar animation synchronization and contextual gesture mapping. NAIA's current architecture generates structured JSON responses containing text content alongside corresponding animation and facial expression metadata (as described in Section 3.3), enabling precise synchronization between speech, gestures, and facial expressions. However, Realtime API's direct speech output creates a fundamental incompatibility with this approach: requesting the model to generate JSON metadata would result in the model vocalizing the JSON structure itself, creating an unacceptable user experience. Although the generated speech text can be extracted for processing, the inability to request structured metadata output forces the implementation of generic, randomized talking animations and neutral facial expressions. This creates a paradoxical trade-off: despite significantly enhanced vocal expressiveness and emotional nuance in speech synthesis, there is a corresponding reduction in avatar immersiveness due to the loss of contextually appropriate animations. Complex gestures such as laughter, greeting, or emotion-specific expressions cannot be accurately synchronized without proper contextual matching mechanisms, necessitating the development of alternative strategies for maintaining avatar-speech coherence in real-time conversational systems.

## 6. Conclusions

The development of NAIA, a multi-role and multimodal academic assistant, is successfully accomplished by integrating a cohesive stack of web, data, and AI services into a single institution-aware solution. This integration unifies language, vision, and voice across specialized roles (researcher, university guide, skills trainer, personal assistant, and receptionist), enabling NAIA to move beyond question-answering toward real academic workflows such as email, calendar, and document support. Unlike traditional single-purpose or unimodal assistants, NAIA's modular architecture (frontend avatars, Django REST services, MariaDB, Redis for memory and caching, secure SSO, Graph integration, RAG over institutional content, and cloud storage) delivers personalized, context-aware guidance while preserving depth within each role.

The effectiveness of the system is underscored by user evaluation metrics and system performance. In this study, most participants reported favorable usability and acceptance (e.g., high ease-of-use and short learning curves). These outcomes indicate growing trust in NAIA's capabilities to support decision-making and routine academic tasks, while the role-specialized design and multimodal interfaces contributed to a more guided and engaging user experience. Qualitative feedback aligned with these results, highlighting value in institution-specific answers and suggesting improvements such as richer visual summaries, tighter academic reminders, and clearer progress cues.

Future development could explore real-time streaming to further reduce latency and enable natural turn-taking, the adoption of role-scoped agent runtimes to strengthen secure tool execution, and enhancements to long-term memory and recommendation techniques to deepen personalization. Additional validation should assess decision-making efficiency and time savings in authentic campus scenarios, providing a more comprehensive measure of practical impact. Together, these directions position NAIA as a robust foundation for scalable, high-impact human–AI interaction in higher education.

**Author Contributions:** Conceptualization, A.F.P.M., K.J.B.Q., S.D.S.C. and C.G.Q.M.; methodology, A.F.P.M., K.J.B.Q., S.D.S.C. and C.G.Q.M.; software, A.F.P.M., K.J.B.Q. and S.D.S.C.; validation, A.F.P.M., K.J.B.Q., S.D.S.C. and C.G.Q.M.; investigation, A.F.P.M., K.J.B.Q., S.D.S.C. and C.G.Q.M.; data curation A.F.P.M., K.J.B.Q. and S.D.S.C.; writing—original draft preparation, A.F.P.M. and K.J.B.Q.; writing—review and editing, A.F.P.M., K.J.B.Q. and C.G.Q.M.; supervision, C.G.Q.M.; project administration, C.G.Q.M.; funding acquisition, C.G.Q.M. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the Project NAIA: Virtual Assistant to Support the University Community Powered by Artificial Intelligence—Uninorte. IA Call under Grant FOFICO 32101 PE0031.

**Data Availability Statement:** No new datasets were generated or analyzed in this study. All data used for evaluation are internal to Universidad del Norte and cannot be publicly shared due to institutional privacy restrictions.

**Acknowledgments:** This work was supported by Universidad del Norte, Barranquilla, Colombia.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. Lee, G.-G.; Shi, L.; Latif, E.; Gao, Y.; Bewersdorff, A.; Nyaaba, M.; Guo, S.; Liu, Z.; Mai, G.; Liu, T.; et al. Multimodality of AI for Education: Towards Artificial General Intelligence. *arXiv* **2023**, arXiv:2312.06037. [[CrossRef](#)]
2. Zawacki-Richter, O.; Marín, V.I.; Bond, M.; Gouverneur, F. Systematic review of research on artificial intelligence applications in higher education—Where are the educators? *Int. J. Educ. Technol. High. Educ.* **2019**, *16*, 39. [[CrossRef](#)]
3. Holmes, W.; Bialik, M.; Fadel, C. *Artificial Intelligence in Education Promises and Implications for Teaching and Learning*; Center for Curriculum Redesign: Boston, MA, USA, 2019. Available online: [https://www.researchgate.net/publication/332180327\\_Artificial\\_Intelligence\\_in\\_Education\\_Promise\\_and\\_Implications\\_for\\_Teaching\\_and\\_Learning](https://www.researchgate.net/publication/332180327_Artificial_Intelligence_in_Education_Promise_and_Implications_for_Teaching_and_Learning) (accessed on 16 June 2025).
4. Challenges and Opportunities for Sustainable Development Education Sector United Nations Educational, Scientific and Cultural Organization. 2019. Available online: <https://en.unesco.org/themes/education-policy-> (accessed on 16 June 2025).
5. Rapanta, C.; Botturi, L.; Goodyear, P.; Guàrdia, L.; Koole, M. Online University Teaching During and After the COVID-19 Crisis: Refocusing Teacher Presence and Learning Activity. *Postdigital Sci. Educ.* **2020**, *2*, 923–945. [[CrossRef](#)] [[PubMed](#)]
6. Steenbergen-Hu, S.; Cooper, H. A meta-analysis of the effectiveness of intelligent tutoring systems on college students' academic learning. *J. Educ. Psychol.* **2014**, *106*, 331–347. [[CrossRef](#)]
7. Mendoza, A.P.; Quiroga, K.B.; Celis, S.D.S.; Quintero, C.G.M. NAIA: A Multi-Technology Virtual Assistant for Boosting Academic Environments—A Case Study. *IEEE Access* **2025**, *13*, 141461–141483. [[CrossRef](#)]
8. Ganesh, P.; Chen, Y.; Lou, X.; Khan, M.A.; Yang, Y.; Sajjad, H.; Nakov, P.; Chen, D.; Winslett, M. Compressing large-scale transformer-based models: A case study on bert. *Trans. Assoc. Comput. Linguist.* **2021**, *9*, 1061–1080. [[CrossRef](#)]
9. Etaiwi, W.; Alhijawi, B. Comparative evaluation of ChatGPT and DeepSeek across key NLP tasks: Strengths, weaknesses, and domain-specific performance. *Array* **2025**, *27*, 100478. [[CrossRef](#)]
10. Xiao, H.; Zhou, F.; Liu, X.; Liu, T.; Li, Z.; Liu, X.; Huang, X. A comprehensive survey of large language models and multimodal large language models in medicine. *Inf. Fusion* **2025**, *117*, 102888. [[CrossRef](#)]
11. Gao, Y.; Xiong, Y.; Gao, X.; Jia, K.; Pan, J.; Bi, Y.; Dai, Y.; Sun, J.; Wang, M.; Wang, H. Retrieval-Augmented Generation for Large Language Models: A Survey. *arXiv* **2023**, arXiv:2312.10997. [[CrossRef](#)]
12. Wang, Y.; Lipka, N.; Zhang, R.; Siu, A.; Zhao, Y.; Ni, B.; Wang, X.; Rossi, R.; Derr, T. Augmenting Textual Generation via Topology Aware Retrieval. *arXiv* **2024**, arXiv:2405.17602. [[CrossRef](#)]
13. Grigoryan, A.; Madoyan, H. Building a Retrieval-Augmented Generation (RAG) System for Academic Papers. 2024. Available online: <https://cse.aua.am/files/2024/05/Building-a-Retrieval-Augmented-Generation-RAG-System-for-Academic-Papers.pdf> (accessed on 17 June 2025).
14. Pereira, R.; Lima, C.; Pinto, T.; Reis, A. Virtual Assistants in Industry 4.0: A Systematic Literature Review. *Electronics* **2023**, *12*, 4096. [[CrossRef](#)]
15. Imangi, I.; Lakshan, S.; De Silva, S.M.; Arachchi, H.A.D.M.; Samarasinghe, G.D. 3D—AI Avatar Attributes Impacting on Bank Customers' Perceived Experience. In Proceedings of the 2025 5th International Conference on Advanced Research in Computing: Converging Horizons: Uniting Disciplines in Computing Research through AI Innovation, ICARC 2025, Belihuloya, Sri Lanka, 19–20 February 2025. [[CrossRef](#)]

16. Xie, T.; Rong, Y.; Zhang, P.; Wang, W.; Liu, L. Towards Controllable Speech Synthesis in the Era of Large Language Models: A Systematic Survey. In Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing, Suzhou, China, 4–9 November 2025. [CrossRef]
17. Sethiya, N.; Maurya, C.K. End-to-End Speech-to-Text Translation: A Survey. *Comput. Speech Lang.* **2025**, *90*, 101751. [CrossRef]
18. Sajja, R.; Sermet, Y.; Cikmaz, M.; Cwiertny, D.; Demir, I. Artificial Intelligence-Enabled Intelligent Assistant for Personalized and Adaptive Learning in Higher Education. *Information* **2024**, *15*, 596. [CrossRef]
19. Gao, H.; Huai, H.; Yildiz-Degirmenci, S.; Bannert, M.; Kasneci, E. DataLIVR: Transformation of Data Literacy Education through Virtual Reality with ChatGPT-Powered Enhancements. In Proceedings of the 2024 IEEE International Symposium on Mixed and Augmented Reality, ISMAR 2024, Bellevue, WA, USA, 21–25 October 2024; pp. 120–129. [CrossRef]
20. Pardo, B.C.E.; Iglesias, R.O.I.; León, A.M.D.; Quintero, M.C.G. EverydAI: Virtual Assistant for Decision-Making in Daily Contexts, Powered by Artificial Intelligence. *Systems* **2025**, *13*, 753. [CrossRef]
21. Sajiukumar, A.; Ranjan, A.; Parvathi, P.K.; Satheesh, A.; Udayan, J.D.; Subramaniam, U. Generative AI-Enabled Virtual Twin for Meeting Assistants. In Proceedings of the 2025 8th International Women in Data Science Conference at Prince Sultan University, WiDS-PSU 2025, Riyadh, Saudi Arabia, 13–14 April 2025; pp. 60–65. [CrossRef]
22. Isaev, R.; Gumerov, R.; Esenalieva, G.; Mekuria, R.R.; Doszhanov, E. HIVA: Holographic Intellectual Voice Assistant. *arXiv* **2023**, arXiv:2307.05501. [CrossRef]
23. John, K.S.; Roy, G.A.; Bindhya, P.S. LLM Based 3D Avatar Assistant. In Proceedings of the 2024 1st International Conference on Trends in Engineering Systems and Technologies, ICTEST 2024, Kochi, India, 11–13 April 2024. [CrossRef]
24. Zhao, Z.; Yin, Z.; Sun, J.; Hui, P. Embodied AI-Guided Interactive Digital Teachers for Education. In Proceedings of the SIGGRAPH Asia 2024 Educator’s Forum, SA 2024, Tokyo, Japan, 3–6 December 2024. [CrossRef]
25. Pan, M.; Kitson, A.; Wan, H.; Prpa, M. ELLMA-T: An Embodied LLM-agent for Supporting English Language Learning in Social VR. In Proceedings of the DIS ’25: Proceedings of the 2025 ACM Designing Interactive Systems Conference, Madeira, Portugal, 5–9 July 2025. [CrossRef]
26. Lai, M.S.; Ooi, E.G.; Goh, I.X.; Teoh, K.L.; Pragasam, T.T.N.; Lim, S.W.; Teh, J.S.R.; Tang, L.J.; Tan, S.C. Real-Time Avatar-Based Speech-to-Speech Conversational AI Tutor on AI PC. In Proceedings of the 2025 IEEE 15th Symposium on Computer Applications & Industrial Electronics (ISCAIE), Penang, Malaysia, 24–25 May 2025; pp. 108–113. [CrossRef]
27. Rubio, J.M.; Neira-Peña, T.; Molina, D.; Vidal-Silva, C. Proyecto UBOT: Asistente virtual para entornos virtuales de aprendizaje. *Inf. Tecnológica* **2022**, *33*, 85–92. [CrossRef]
28. Deusens Hyperxperiece Sl. Alice Assistant | El Avatar Virtual que se Comunica de Forma Natural. Available online: <https://deusens.com/es/interactive-products/alice-assistant> (accessed on 27 September 2024).
29. Duguleană, M.; Briciu, V.A.; Duduman, I.A.; Machidon, O.M. A Virtual Assistant for Natural Interactions in Museums. *Sustainability* **2020**, *12*, 6958. [CrossRef]
30. Technologies, Y.I. Asistente Virtual Inteligente Web | YIO Technologies. Available online: <https://yio.ai/asistente-ai-inteligente-web-yio/> (accessed on 27 September 2024).
31. Shawar, B.A.; Atwell, E. ALICE Chatbot: Trials and Outputs. *Comput. Sist.* **2015**, *19*, 625–632. [CrossRef]
32. Ornelas, F.A.G. Diseño e Implementación de un Asistente Virtual (Chatbot) para Ofrecer Atención a los Clientes de una Aerolinea Mexicana por Medio de sus Canales Conversacionales. Ciudad de Mexico. 2020. Available online: <https://infotec.repositorioinstitucional.mx/jspui/handle/1027/402> (accessed on 4 October 2024).
33. Flores, M.F. Diseño de Avatares para una Aplicación de Metaverso en Salud. 2023. Available online: <https://upcommons.upc.edu/handle/2117/393983> (accessed on 4 October 2024).
34. Flores, C.; Harold, S.; Barrionuevo, L.; Harnold, K. Desarrollo de Asistente Virtual con Inteligencia Artificial para la Atención del Paciente del Instituto Nacional de Salud Niño. Repositorio Institucional—UCV. 2023. Available online: <https://repositorio.ucv.edu.pe/handle/20.500.12692/123752> (accessed on 4 October 2024).
35. Ramirez, H.; Carlos, J. Diseño e Implementación de un Asistente Virtual de Información Turística para la Ciudad de Gandia. 2023. Available online: <https://riunet.upv.es/handle/10251/197609> (accessed on 4 October 2024).
36. Arias, P.M.G. Diseño, Desarrollo e Implementación de una Asistente Virtual para la Resolución de Dudas Sobre los Procesos Académicos de la Universidad Politécnica Salesiana - Sede Cuenca Utilizando Inteligencia Artificial y Procesamiento de Lenguaje Natural. Cuenca. 2022. Available online: <http://dspace.ups.edu.ec/handle/123456789/22027> (accessed on 18 February 2025).
37. Fernández, M.; Venezuela, G. Tecnologías emergentes: Diseño de asistente virtual universitario basado en inteligencia artificial. *Obs. Del Conoc.* **2023**, *8*, 15–35. Available online: <https://revistaoac-cal.oncti.gob.ve/index.php/ODC/article/view/97> (accessed on 4 October 2024).
38. Herramienta, A.D.E.C.; De la Cruz Romero Asesor, D.M.L.; Paulino, M.D.C.O.; Perú, L. Asistente Virtual Basado en Inteligencia Artificial como Herramienta de Tesis Para Estudiantes Universitarios de la Carrera de Ingeniería. Bachelor’s Thesis, Universidad Privada del Norte, Trujillo, Peru, 2023. Available online: <https://repositorioslatinoamericanos.uchile.cl/handle/2250/6381910> (accessed on 4 October 2024).

39. Canosa, A.C. FutureLab: Diseño y Desarrollo de un Prototipo de alta Resolución de una App Móvil con Inteligencia Artificial, para Ayudar a los Estudiantes en la Elección de Formación Universitaria. Master's Thesis, Universitat Oberta de Catalunya (UOC), Barcelona, Spain, 2019. Available online: <https://openaccess.uoc.edu/handle/10609/97026> (accessed on 4 October 2024).
40. Takita, H.; Hashiura, K.; Hatada, Y.; Kodama, D.; Narumi, T.; Tanikawa, T.; Hirose, M. Do We Still Need Human Instructors? Investigating Automated Methods for Motor Skill Learning in Virtual Co-Embodiment. *IEEE Trans. Vis. Comput. Graph.* **2025**, *31*, 2455–2463. [[CrossRef](#)]
41. Ait-Mlouk, A.; Jiang, L. KBot: A Knowledge Graph Based ChatBot for Natural Language Understanding over Linked Data. *IEEE Access* **2020**, *8*, 149220–149230. [[CrossRef](#)]
42. Meloni, A.; Angioni, S.; Salatino, A.; Osborne, F.; Recupero, D.R.; Motta, E. Integrating Conversational Agents and Knowledge Graphs Within the Scholarly Domain. *IEEE Access* **2023**, *11*, 22468–22489. [[CrossRef](#)]
43. Kleine, A.-K.; Schaffernak, I.; Lermer, E. Exploring predictors of AI chatbot usage intensity among students: Within- and between-person relationships based on the technology acceptance model. *Comput. Hum. Behav. Artif. Hum.* **2025**, *3*, 100113. [[CrossRef](#)]
44. Racero, F.J.; Bueno, S.; Gallego, M.D. Predicting Students' Behavioral Intention to Use Open Source Software: A Combined View of the Technology Acceptance Model and Self-Determination Theory. *Appl. Sci.* **2020**, *10*, 2711. [[CrossRef](#)]
45. Vlachogianni, P.; Tseliots, N. Perceived usability evaluation of educational technology using the System Usability Scale (SUS): A systematic review. *J. Res. Technol. Educ.* **2022**, *54*, 392–409. [[CrossRef](#)]
46. Brown, L.D.; Cai, T.T.; Gupta, A.D. Interval Estimation for a Binomial Proportion. *Stat. Sci.* **2001**, *16*, 101–133. [[CrossRef](#)]

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.