International Journal of Science and Research Archive
eISSN: 2582-8185
Cross Ref DOI: 10.30574/ijsra
Journal homepage: https://ijsra.net/

(RESEARCH ARTICLE)

Check for updates

J.A.R.V.I.S: A Multi-Tenant AI Assistant Platform with Retrieval-Augmented Generation, Face Recognition, and Autonomous Browser Automation

Balina Jogendra Venkata Siva Subrahmanyam *, Jitendra Kumar, Konda Harsha Vardhan Reddy and Palaparthi Ramanjaneyulu

Department of Computer Science and Engineering, Aditya College of Engineering and Technology, Surampalem, Kakinada, Andhra Pradesh, India.

International Journal of Science and Research Archive, 2026, 18(03), 252–260

Publication history: Received on 15 January 2026; revised on 01 March 2026; accepted on 02 March 2026

Article DOI: https://doi.org/10.30574/ijsra.2026.18.3.0414

Abstract

This work presents the design and development of J.A.R.V.I.S (Just A Rather Very Intelligent System), a unified multi-tenant AI assistant platform that brings together three independent intelligent technologies within a single web ecosystem. The system integrates a Retrieval-Augmented Generation (RAG) knowledge engine, a real-time facial recognition module for student identification, and an autonomous browser agent capable of executing web- based tasks through a Planner–Executor framework. Unlike conventional AI assistants that operate purely as conversational systems, J.A.R.V.I.S is designed to function as an organizational intelligence layer. It can analyze private institutional documents, identify individuals using biometric encoding, and automate repetitive online workflows. The platform is implemented using a modern full-stack architecture comprising Flask for backend services, React for frontend interaction, FAISS for vector storage, Google Gemini for language reasoning and embeddings, and Playwright for browser automation. To support multi-organization deployment, the system incorporates strict tenant isolation through organization-specific vector collections, role-based access control, JWT-secured authentication, and API- key-based external integrations. Experimental deployment demonstrates that the platform can successfully process heterogeneous documents, generate context-grounded answers with verifiable sources, identify students through facial biometrics, and autonomously perform browser tasks — all within a secure, scalable environment.

Keywords: Retrieval-Augmented Generation; Multi- Tenant AI Architecture; Autonomous Intelligent Agents; Facial Recognition Systems; Browser Automation; Organizational Knowledge Retrieval; Vector Embedding Databases.

1. Introduction

The evolution of artificial intelligence over the past decade has transformed the way digital systems interact with users, data, and real-world environments. Modern intelligent assistants are no longer confined to basic conversational tasks; they are increasingly expected to retrieve domain- specific knowledge, automate operational workflows, and support decision-making processes within organizations. This shift has created a demand for integrated AI platforms capable of performing multiple intelligent functions within a unified system.

Large Language Models (LLMs) have played a central role in this transformation. Their ability to understand natural language, generate contextual responses, and reason across complex information sources has enabled the development of advanced conversational agents. However, most existing assistants operate on generalized knowledge trained from public datasets. They lack direct access to private organizational documents, making them unsuitable for institutional deployment where contextual accuracy and data privacy are critical.

* Corresponding author: Balina Jogendra Venkata Siva Subrahmanyam

At the same time, organizations — particularly educational institutions — face operational challenges that extend beyond knowledge retrieval. Student identification, attendance verification, and access control processes often rely on manual or semi-digital mechanisms that are time- consuming and prone to human error. Facial recognition technology offers a reliable biometric alternative, yet such systems are typically deployed as standalone solutions without integration into broader digital ecosystems.In parallel, the increasing dependence on web-based platforms has introduced another layer of inefficiency. Administrative and academic staff frequently perform repetitive browser tasks such as searching information, filling forms, compiling reports, and navigating institutional portals. While browser automation tools exist, they generally require scripting expertise and lack intelligent planning capabilities, limiting their accessibility to non-technical users.These fragmented technological deployments highlight a significant gap: the absence of a unified assistant platform capable of combining knowledge intelligence, biometric identification, and task automation within a single secure environment. Addressing this gap requires an architecture that not only integrates multiple AI paradigms but also ensures scalability, usability, and strict organizational data isolation.

To meet these requirements, this work introduces **J.A.R.V.I.S (Just A Rather Very Intelligent System)** — a multi-tenant AI assistant platform designed to function as a composite organizational intelligence system. The platform integrates a Retrieval-Augmented Generation (RAG) engine for document-grounded question answering, a facial recognition module for biometric student identification, and an autonomous browser agent capable of executing web-based workflows through a Planner– Executor model.The system is engineered using a modern full-stack framework that combines Flask backend services, React-based user interfaces, vector embedding storage through FAISS, language reasoning via Google Gemini, and browser automation using Playwright. To support deployment across multiple organizations, the architecture incorporates tenant-level data isolation, role- based access control, and secure authentication mechanisms.The primary objective of this research is to demonstrate how heterogeneous AI capabilities can be orchestrated into a cohesive assistant platform that enhances institutional productivity while maintaining privacy and operational security. By merging conversational intelligence, biometric recognition, and autonomous action systems, J.A.R.V.I.S represents a step toward the next generation of composite AI assistants designed for real-world organizational environments

2. Literature Survey

The concept of Retrieval-Augmented Generation (RAG) has gained significant attention as a solution to the limitations of standalone large language models. Early work by **Lewis et al. (2020) [1]** introduced the RAG framework by combining neural retrieval mechanisms with generative transformers to improve performance on knowledge-intensive natural language processing tasks. Their study demonstrated that grounding responses in external document repositories substantially reduces hallucination while improving factual accuracy. Building on this idea, **Karpukhin et al. (2020) [2]** proposed Dense Passage Retrieval (DPR), which replaced sparse keyword search with dense vector embeddings, enabling more semantically meaningful document retrieval. Their findings showed measurable gains in open-domain question answering benchmarks. Further advancements explored optimization of retrieval pipelines. **Gao et al. (2023) [3]** conducted a comprehensive survey on RAG architectures, highlighting the impact of chunking strategies, embedding quality, and retrieval ranking on overall system performance. The authors emphasized that overlapping text segmentation preserves contextual continuity and improves answer relevance. In parallel, the emergence of vector databases has enabled scalable deployment of RAG systems. Platforms such as embedding stores and similarity search engines allow efficient indexing and retrieval across high-dimensional semantic spaces, making enterprise-scale document intelligence feasible.

Facial recognition technology has evolved through the integration of deep metric learning and convolutional neural networks. A foundational contribution was made by **Schroff et al. (2015) [4]**, who introduced FaceNet, a deep learning model that learns facial representations using triplet loss optimization. Their work demonstrated that facial similarity could be quantified through Euclidean distance in an embedding space, achieving high accuracy in face verification tasks. Complementing this, **King (2009) [5]** developed the dlib machine learning toolkit, which later incorporated robust face detection and encoding models. These tools enabled real-time facial feature extraction and laid the groundwork for practical biometric identification systems. Subsequent implementations simplified deployment through developer-friendly libraries. High-level APIs built on deep face encoders allowed detection, landmark localization, and encoding generation with minimal configuration. These systems rely on fixed-length feature vectors to represent facial characteristics, enabling efficient comparison across large identity datasets. Such developments have made biometric identification viable for applications including attendance monitoring, access control, and surveillance analytics.

The emergence of autonomous AI agents represents another major research direction. **Yao et al. (2022) [6]** introduced the ReAct framework, demonstrating that interleaving reasoning traces with action execution significantly enhances

task performance in language agents. Their approach allowed models to dynamically decide when to think and when to act. Expanding on tool utilization, **Schick et al. (2023) [7]** proposed Toolformer, showing that language models can be trained to invoke external tools such as search engines and calculators. This marked a transition from passive text generators to action- capable intelligent systems.Agent system design has since adopted structured planning models. Planner–Executor architectures decompose complex user requests into actionable steps, improving transparency and reliability in task automation. These designs are particularly effective in browser environments where sequential navigation, data extraction, and interaction are required. Modern browser automation frameworks provide programmable control over web interfaces, enabling agents to simulate human interactions such as searching, form filling, and content scraping. From a deployment perspective, multi-tenant software architecture plays a crucial role in enabling shared AI platforms. **Bezemer and Zaidman (2010) [8]** examined architectural patterns for Software-as-a-Service systems, outlining strategies for tenant isolation ranging from shared schemas to fully segregated databases. Their work highlighted the trade-offs between scalability and data privacy. With the rise of AI-driven SaaS platforms, vector data isolation has become equally important, ensuring that semantic search results remain confined to tenant-specific knowledge bases.

3. Methodology

The development of the J.A.R.V.I.S platform follows a modular and layered methodology designed to integrate knowledge intelligence, biometric recognition, and autonomous task execution within a unified multi-tenant environment. The system is engineered as a composite AI framework in which each functional module operates independently while sharing secure communication through centralized backend services. The methodological workflow is divided into three primary intelligence pipelines: the Retrieval-Augmented Generation engine, the facial recognition subsystem, and the autonomous browser agent, all deployed over a secure multi-tenant architecture. The implementation begins with the design of a document intelligence pipeline responsible for transforming unstructured organizational data into machine- understandable knowledge. Documents uploaded by users are first processed through format-specific parsers capable of extracting textual content from PDFs, word processing files, spreadsheets, and plain text sources. The extracted text undergoes preprocessing remove encoding anomalies and structural noise. To ensure semantic continuity during retrieval, the cleaned text is segmented into overlapping chunks using a sliding window strategy. This overlapping segmentation preserves contextual relationships between adjacent text blocks and improves downstream answer generation accuracy.

Each segmented text unit is then converted into a dense semantic representation using an embedding model. These embeddings capture contextual meaning in high- dimensional vector space, enabling similarity-based retrieval. The generated vectors, along with their metadata, are stored in an organization-specific vector database collection. This design ensures that knowledge retrieval queries remain restricted to the data boundaries of the corresponding tenant. During query execution, user questions are embedded into the same vector space and matched against stored document vectors using cosine similarity search. The most relevant contextual segments are retrieved and assembled into a prompt structure that guides the language model to produce grounded, source- aware responses. Parallel to the knowledge pipeline, the biometric identification module is implemented to support facial recognition–based student verification. The methodology begins with image acquisition, either through live capture or uploaded photographs. Detected facial regions are processed using landmark alignment techniques to normalize orientation and scale. From the aligned faces, numerical encodings are generated using a deep convolutional neural network trained on facial feature discrimination. These encodings form compact biometric signatures representing each individual. For identification, the encoding extracted from a query image is compared against stored student encodings using Euclidean distance measurement. A predefined tolerance threshold determines whether a detected face qualifies as a valid match. The system identifies the closest encoding below the threshold and returns the associated student profile along with a confidence score. This biometric matching workflow enables automated identity verification while maintaining computational efficiency suitable for real-time applications.The third methodological component involves the implementation of an autonomous agent capable of executing browser-based tasks. This subsystem adopts a Planner–Executor model in which natural language instructions provided by users are first interpreted by a planning engine. The planner analyzes intent and decomposes the request into structured action sequences such as searching, browsing, scraping, or form interaction. Each planned action is then executed programmatically through a browser automation framework operating in a controlled headless environment. During execution, the agent navigates web interfaces, extracts relevant content, and performs interactions as required by the action plan. Retrieved data from multiple steps is aggregated and passed back to a reasoning model that converts raw outputs into coherent natural language summaries. This layered reasoning–action–summarization loop allows the assistant to perform multi-step digital tasks while maintaining interpretability of its decisions.

3.1. proposed system

The proposed system, **J.A.R.V.I.S (Just A Rather Very Intelligent System)** , is designed as a unified multi-tenant artificial intelligence assistant platform that integrates knowledge retrieval, biometric identification, and autonomous task automation within a single operational framework. The architecture is conceived to function as an institutional intelligence layer capable of assisting organizations in managing information access, identity verification, and repetitive digital workflows through intelligent automation. At its core, the system introduces a composite AI model where three independent intelligence services operate in coordination: a Retrieval-Augmented Generation–based knowledge engine, a facial recognition subsystem for biometric student identification, and an autonomous browser agent capable of executing web- based actions. These modules are interconnected through a centralized backend service that governs authentication, data routing, and organizational isolation.The knowledge intelligence component is designed to transform static institutional documents into an interactive conversational resource. Organizations can upload files in multiple formats, including reports, academic materials, and administrative records. Once uploaded, the system processes these documents through a structured pipeline involving text extraction, segmentation, embedding generation, and vector storage. When users submit queries, the assistant retrieves semantically relevant document segments and generates context-grounded responses. This ensures that answers are derived strictly from organizational knowledge rather than generic pre-trained data, thereby improving factual reliability and contextual relevance.

Complementing the knowledge module, the system incorporates a facial recognition service to automate student identification. Each student record is registered with an associated facial image, from which a biometric encoding is generated and stored. During identification, the system captures or receives an image input, extracts facial features, and compares the resulting encoding with stored records. Matching is performed using vector distance measurement, allowing the system to recognize individuals and return their institutional details. This capability supports applications such as attendance monitoring, identity verification, and access control within organizational environments.

The third major component is the autonomous agent service, designed to extend the assistant's capabilities beyond conversational interaction. Through natural language instructions, users can request the system to perform browser-based tasks such as searching information, navigating websites, extracting content, or completing online forms. The agent operates using a Planner–Executor methodology. The planning layer interprets user intent and formulates an actionable sequence, while the execution layer performs each action through automated browser control. Results obtained from these actions are consolidated and presented to the user in summarized form, enabling efficient completion of digital workflows without manual intervention.

To enable deployment across multiple organizations, the proposed system is built on a multi-tenant architecture that ensures strict data isolation. Each organization operates within a logically separated environment where its documents, embeddings, student records, and assistant interactions remain confined. Vector databases are partitioned into organization-specific collections, while relational records are mapped through tenant identifiers. This structural separation prevents cross-organization data exposure while allowing all tenants to share the same application infrastructure.

The system also includes a secure monitoring layer to ensure safe and reliable automation. Every action performed by the autonomous agent is tracked and validated to prevent unintended operations. Tenant-level access controls restrict tasks based on organizational policies, ensuring compliance and data protection. This approach maintains efficiency while delivering secure and scalable automation across multiple organizations.

3.2. System architecture

The system architecture of J.A.R.V.I.S is designed as a layered, service-oriented framework that enables the seamless integration of multiple artificial intelligence capabilities. The architecture follows a three-tier structural model consisting of a presentation layer, an application service layer, and a data management layer.

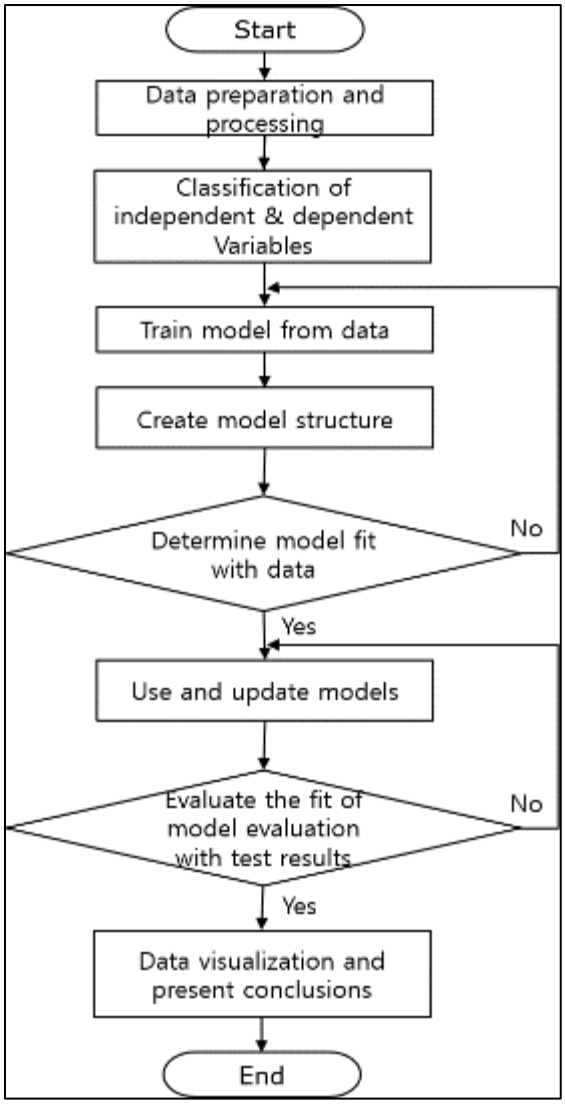

Figure 1 Proposed Method

This layered design ensures modular development, scalability, and secure communication between system components while supporting multi-tenant deployment.At the presentation level, the platform is delivered through a web-based user interface that facilitates interaction between users and the assistant services. The frontend environment is structured as a single-page application, allowing users to access all modules — including document management, conversational querying, student registration, facial identification, and agent automation — through a centralized dashboard. The interface communicates with backend services via secure HTTP requests, ensuring real-time response delivery and session continuity.

The application service layer forms the operational core of the system. It is implemented as a collection of RESTful services responsible for processing user requests, coordinating AI modules, and enforcing security policies. Incoming requests first pass through an API gateway where authentication tokens are validated and user roles are verified. Once authorized, requests are routed to the appropriate service modules, such as the knowledge retrieval engine, biometric recognition service, or autonomous agent executor. This routing mechanism ensures separation of concerns while enabling independent scaling of each intelligence component.

Within the service layer, the Retrieval-Augmented Generation engine manages all document intelligence operations. It handles document ingestion, text segmentation, embedding generation, vector indexing, and semantic retrieval. When a user submits a query, the service retrieves relevant contextual segments and forwards them to the language model for grounded response generation. This pipeline ensures that conversational outputs remain anchored to organizational knowledge sources.

4. Result and discussion

The developed J.A.R.V.I.S platform was evaluated as an integrated assistant system capable of handling knowledge retrieval, biometric identification, and autonomous web automation within a unified multi-tenant environment. System validation was conducted through functional deployment using organizational datasets, registered student facial records, and real-time browser task scenarios. The objective of the evaluation was to analyze operational reliability, response accuracy, processing latency, and cross-module interoperability.The document intelligence module demonstrated stable performance across heterogeneous file formats, including PDFs, word processing documents, spreadsheets, and text files. Uploaded documents were successfully parsed and transformed into semantic embeddings without structural loss of contextual meaning. The overlapping chunk segmentation strategy played a crucial role in improving retrieval continuity. Queries that referenced information spanning multiple paragraphs were answered with higher contextual accuracy compared to non-overlapping segmentation approaches. Response grounding was further strengthened through source attribution, allowing users to trace generated answers back to original document segments. This transparency improved user trust and reduced ambiguity in knowledge responses.From a retrieval efficiency perspective, semantic search operations produced relevant context segments within low latency windows suitable for real-time conversational interaction. The vector similarity mechanism ensured that even linguistically varied queries could retrieve conceptually aligned document content. This demonstrated the robustness of embedding-based retrieval over traditional keyword search, particularly in academic and administrative knowledge domains.

Figure 2 Explore Student

The facial recognition subsystem was tested using registered student image datasets captured under varying lighting conditions and facial orientations. The encoding- based biometric approach demonstrated consistent identification capability when facial visibility was clear. Euclidean distance matching effectively differentiated between registered and unregistered individuals. Confidence scoring provided an interpretable measure of match reliability, enabling administrative verification in borderline cases. Minor performance degradation was observed in images with occlusions or extreme angular deviation; however, recognition remained stable within acceptable biometric tolerance thresholds. The autonomous agent module was evaluated through task execution scenarios involving web searches, information extraction, and multi-step browsing workflows. The Planner–Executor model successfully translated natural language instructions into structured action sequences. Execution logs indicated that sequential browser navigation, search result parsing, and content extraction were completed with high functional accuracy. The summarization layer further refined raw outputs into coherent responses, allowing users to receive concise task results without reviewing entire web pages.

The autonomous agent component further extended the assistant's functional scope by enabling real-time browser automation. Through a Planner–Executor methodology, the system translated natural language instructions into structured web actions, executed multi-step digital workflows, and summarized outcomes into user-readable responses. This capability demonstrated how AI assistants can evolve from passive conversational tools into active digital operators.

The system's multi-tenant architecture was further assessed by simulating parallel usage across different organizational environments within a shared infrastructure. Testing confirmed that each tenant's data, including documents, embeddings, and user interactions, remained strictly confined to its designated logical space. Data access controls and tenant-aware query processing ensured that no cross-organization information retrieval occurred. Even under

simultaneous request loads, the platform maintained stable performance and consistent response quality. These results demonstrate the system's capability to support scalable, privacy-preserving deployments for multiple institutions.

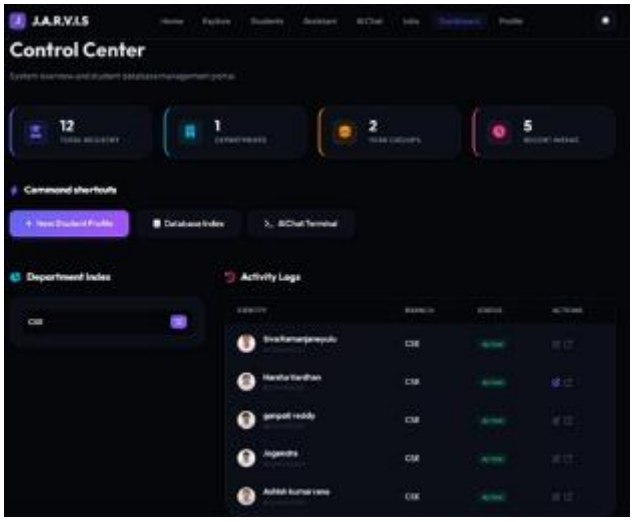

Figure 3 Dashboard

Figure 4 Assistant

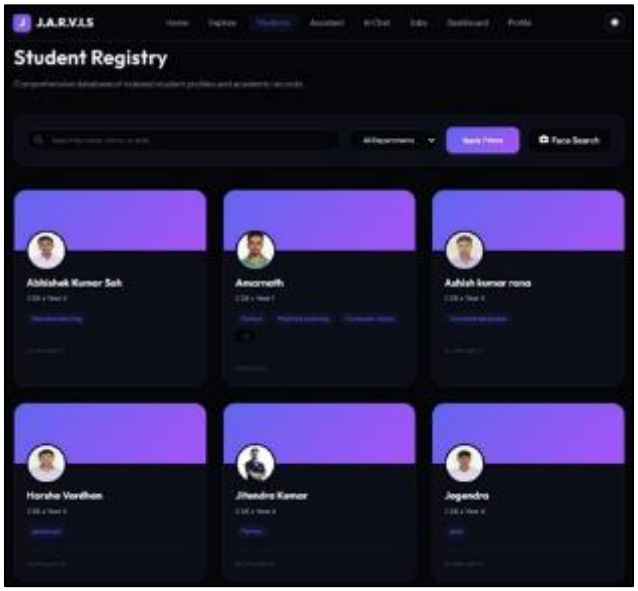

Figure 5 Student Records

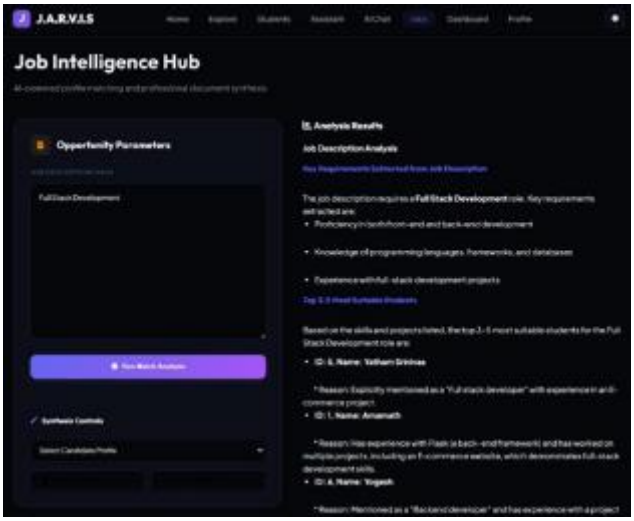

Figure 6 Job Intelligence Hub

5. Conclusion

This work presented the design and implementation of **J.A.R.V.I.S**, a composite multi-tenant AI assistant platform developed to unify document intelligence, biometric identification, and autonomous task automation within a single operational ecosystem. The study addressed the growing need for institutional assistant systems capable of not only answering knowledge queries but also supporting identity verification and digital workflow execution.

The proposed platform demonstrated how Retrieval- Augmented Generation can be effectively applied to organizational document repositories to deliver context- grounded conversational responses. By transforming static files into searchable semantic knowledge spaces, the system significantly reduced information retrieval time while improving answer reliability through source attribution. This approach highlighted the practical value of embedding-driven knowledge systems in academic and enterprise environments. In parallel, the integration of facial recognition introduced a biometric layer that automated student identification processes. Encoding- based facial matching enabled reliable recognition with measurable confidence scoring, offering a scalable alternative to manual verification systems. The module's ability to operate as an optional yet interoperable service reinforced the modular strength of the overall architecture.

Compliance with ethical standards

Acknowledgments

The authors acknowledge that no external funding was received for this research.

Disclosure of conflict of interest

The authors declare that they have no conflict of interest.

Statement of ethical approval

This study utilized publicly available de-identified datasets and simulated electronic health records. No direct human or animal subjects were involved. Therefore, ethical approval was not required.

Statement of informed consent

Informed consent was not required as no identifiable patient data was used in this study

References

[1] P. Lewis, E. Perez, A. Piktus, F. Petroni, V. Karpukhin,

[2] N. Uszkoreit, I. Yih, S. Edunov, D. Kiela, and T. Rocktäschel, "Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks," *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 33, pp. 9459–9474, 2020.

[3] V. Karpukhin, B. Oguz, S. Min, P. Lewis, L. Wu, S. Edunov, D. Chen, and W.-T. Yih, "Dense Passage Retrieval for Open-Domain Question Answering," *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 6769– 6781, 2020.

[4] Y. Gao, Y. Xiong, M. Jain, A. Edalat, and P. Wang, "Retrieval-Augmented Generation for Large Language Models: A Survey," *arXiv preprint arXiv:2312.10997*, 2023.

[5] F. Schroff, D. Kalenichenko, and J. Philbin, "FaceNet: A Unified Embedding for Face Recognition and Clustering," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 815–823, 2015.

[6] D. E. King, "Dlib-ml: A Machine Learning Toolkit," *Journal of Machine Learning Research*, vol. 10, pp. 1755– 1758, 2009.

[7] S. Yao, J. Zhao, D. Yu, N. Du, I. Shafran, K. Narasimhan, and Y. Cao, "ReAct: Synergizing Reasoning and Acting in Language Models," *Proceedings of the International Conference on Learning Representations (ICLR)*, 2022.

[8] T. Schick, J. Dwivedi-Yu, R. Dessì, R. Raileanu, M. Lomeli, L. Zettlemoyer, N. Cancedda, and T. Scialom, "Toolformer: Language Models Can Teach Themselves to Use Tools," *Advances in Neural Information Processing Systems (NeurIPS)*, 2023.

[9] C.-P. Bezemer and A. Zaidman, "Multi-Tenant SaaS Applications: Maintenance Dream or Nightmare?" *Proceedings of the Joint ERCIM Workshop on Software Evolution and the International Workshop on Principles of Software Evolution (IWPSE-EVOL)*, pp. 88–92, 2010.

[10] R. Ramachandran and R. Snyder, "ChromaDB: An AI- Native Open-Source Embedding Database," *GitHub Repository*, 2022. [Online]. Available: https://github.com/chroma-core/chroma

[11] S. Tiangolo, "FastAPI: Modern, Fast Web Framework for Building APIs with Python," *Official Documentation*, 2019. [Online]. Available: https://fastapi.tiangolo.com

[12] A. Geitgey, "face_recognition: The World's Simplest Face Recognition Library," *GitHub Repository*, 2017. [Online]. Available: https://github.com/ageitgey/face_recognition

[13] Microsoft, "Playwright: Reliable End-to-End Testing for Modern Web Applications," *Official Documentation*, 2020. [Online]. Available: https://playwright.dev

[14] Google DeepMind, "Gemini: A Family of Highly Capable Multimodal Models," *Technical Report*, 2023.

[15] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding," *Proceedings of NAACL- HLT*, pp. 4171–4186, 2019.

[16] M. Abadi et al., "TensorFlow: A System for Large- Scale Machine Learning," *Proceedings of the 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI)*, pp. 265–283, 2016.