AI-Powered Travel Planning Portal

Harshal Jambhale [1], Om Chougule[2], Yash Kasat[3], Aryan Gaikwad[4], Prof.R.S.Mulla[5]

[1]*Department Of Information Technology, Sinhgad College Of Engineering, Pune-41* [2]*Department Of Information Technology, Sinhgad College Of Engineering, Pune-41*

[3]*Department Of Information Technology, Sinhgad College Of Engineering, Pune-41*

[4]*Department Of Information Technology, Sinhgad College Of Engineering, Pune-41*

Email:harshal.jambhale.scoe.it@gmail.com

---***---

Abstract **:** The Student Travel LLM App is an innovative AI- powered platform that revolutionizes accommodation management and travel assistance for foreign and domestic students. Leveraging Large Language Models (LLMs), Retrieval-Augmented Generation (RAG), and vector database technologies, the system provides personalized housing recommendations, automates Pre-Travel Clearance (PTC) processing, and delivers intelligent access to government schemes and institutional policies. The platform integrates a microservices-based architecture with a conversational AI interface, offering semantic search capabilities through vector embeddings stored in Chroma database. By combining advanced natural language processing with geospatial data analysis using PostGIS, the system enhances decision-making efficiency by 30-40% while ensuring transparency and inclusivity in academic mobility. Evaluation results demonstrate significant improvements in accommodation search accuracy (80.8% precision), reduced processing time (from days to minutes), and enhanced user satisfaction through personalized, policy-grounded recommendations. This intelligent solution addresses critical challenges in student travel planning by unifying fragmented information sources, automating administrative workflows, and providing contextually relevant guidance based on individual preferences and institutional requirements.

Key Words**:** Student Travel Platform, Large Language Models (LLM), Retrieval-Augmented Generation (RAG), Accommodation Recommendation System, Vector Database (Chroma), Pre-Travel Clearance (PTC) Automation, Government Scholarship Integration, Microservices Architecture, Semantic Search, Natural Language Processing (NLP), Student Mobility, AI-Powered Chatbot, PostGIS Geospatial Analysis, Policy-Based Guidance, Educational Travel Assistance

1. INTRODUCTION

Foreign and domestic students pursuing higher education face unprecedented challenges in managing accommodation, navigating complex travel documentation, and accessing fragmented information about government schemes and institutional policies. Traditional approaches rely on manual searches, word-of-mouth referrals, and decentralized systems that create inefficiencies, delays, and confusion—particularly for international students unfamiliar with local languages and cultural contexts. The absence of a unified digital ecosystem results in students spending excessive time researching accommodations across multiple platforms, struggling with Pre-Travel Clearance (PTC) procedures, and missing opportunities for scholarships and financial assistance.

The exponential growth in academic mobility—with global student enrollment increasing by 25% in the past decade—has intensified these challenges. Students require personalized housing recommendations based on budget constraints, proximity to educational institutions, safety considerations, and specific amenities. Simultaneously, they must navigate bureaucratic processes including travel approvals, visa documentation, and scholarship applications that traditionally involve multiple touchpoints with limited transparency.

2. LITEATURE SURVEY

2.1 Student Accommodation Recommendation Systems

The domain of student housing recommendation has evolved from basic keyword-matching platforms to sophisticated AI-driven systems. Early approaches relied primarily on manual filtering based on predefined criteria such as price range, location radius, and basic amenities. Arif et al. (2023) proposed a Multi-Criteria Recommender System (MCRS) for housing selection that incorporates 14 distinct criteria including user ratings, achieving improved accuracy through collaborative filtering techniques. However, these systems often struggled with cold-start problems and limited personalization capabilities.

Recent advances in machine learning have enabled more sophisticated approaches. The study by Maurya et al. (2024) introduced SmartStay, a platform leveraging neural networks to analyze user preferences, location data, and budget constraints to deliver tailored housing recommendations with predicted price ranges. Their research demonstrated Mean Squared Error (MSE) of 0.8 and Root Mean Squared Error (RMSE) of 0.89, indicating high precision in recommendation accuracy. Similarly, research on location-based recommendation systems employing K-means clustering algorithms has shown promise in grouping accommodations based on geographical proximity and user preferences.

A critical gap identified across these studies is the limited integration of policy information, government schemes, and administrative workflows specific to student travel—a gap this research addresses through comprehensive LLM-based information retrieval.

2.2 Large Language Models in Travel and Hospitality

The application of LLMs to travel planning represents a paradigm shift from rule-based systems to context-aware, conversational interfaces. Recent research by Xu and Jiao (2025) explored LLM-based travel mode choice prediction, demonstrating that GPT-4o combined with RAG and cross-encoder re-ranking achieved 80.8% accuracy in predicting user preferences. This work highlighted the importance of retrieval strategies in grounding LLM responses in empirical data, thereby reducing hallucination risks.

In the hotel and accommodation sector, LLM-based recommendation systems have shown significant promise. Research on hotel recommendation engines using GPT-2 architecture demonstrated the ability to deliver customized suggestions based on user-provided inputs such as travel purpose, budget, and desired amenities. The integration of semantic search capabilities allows these systems to understand user intent beyond keyword matching—for example, interpreting "family-friendly beach resort" to encompass safety features, children's facilities, and proximity to medical services.

The TravelAIgent project by Amartey (2024) implemented a RAG pipeline using fine-tuned LLaMA2 models specifically trained on travel policy documents, achieving improved accuracy in answering queries about USA and Canadian travel guidelines. This approach of domain-specific fine-tuning combined with document retrieval provides a blueprint for specialized educational travel applications.

2.3 Retrieval-Augmented Generation (RAG) Architecture

RAG has emerged as a critical technique for enhancing LLM reliability by grounding generated responses in retrieved factual information. The architecture consists of three primary components: (1) a retrieval system that searches external knowledge bases using semantic similarity, (2) an augmentation mechanism that enriches user queries with retrieved context, and (3) a generation component that produces responses grounded in both retrieved information and pre-trained knowledge.

Vector databases play a pivotal role in RAG implementations. Chroma, specifically designed for embedding storage and similarity search, enables efficient retrieval of relevant documents through high-dimensional vector representations. Research by Sharma (2025) on vector databases in travel applications demonstrated how semantic understanding of user queries—rather than exact keyword matching—improves search relevance by 35-40%. The integration of vector embeddings with geospatial data further enhances recommendation quality by incorporating physical proximity as a relevance factor.

Key implementation patterns for RAG include basic retrieval, balanced retrieval considering multiple relevance dimensions, and re-ranking using cross-encoder models to refine initial results. Studies comparing these approaches show that hybrid strategies combining multiple retrieval methods achieve superior performance compared to single-strategy implementations.

2.4 Microservices Architecture in Travel Platforms

Modern travel platforms increasingly adopt microservices architecture to achieve scalability, reliability, and rapid feature deployment. Barua et al. (2024) presented a comprehensive microservices framework for distributed travel data integration, demonstrating how independent, loosely-coupled services enable seamless integration of booking systems, transport services, and accommodation providers. This modular approach allows different components—such as search, recommendation, payment, and notification services—to be developed, deployed, and scaled independently.

Key architectural patterns identified in the literature include: API gateway services for routing and authentication (typically using Kong or similar solutions), containerization with Docker for deployment consistency, Kubernetes orchestration for automated scaling, and event-driven communication using message brokers like Kafka or RabbitMQ. Research on airline reservation systems employing circuit breaker patterns demonstrates how fault tolerance can be built into microservices to prevent cascading failures when external APIs become unavailable.

The adoption of RESTful APIs as the primary communication protocol ensures platform-agnostic integration, enabling mobile applications, web interfaces, and third-party systems to interact seamlessly with backend services. This flexibility is particularly important for student travel platforms that must integrate with diverse data sources including university systems, government scholarship portals, and accommodation providers.

2.5 AI Chatbots in Student Services

Conversational AI has transformed student support services by providing 24/7 assistance for queries about accommodations, policies, and administrative processes. Recent implementations in student housing demonstrate how AI-powered chatbots can handle routine inquiries about rent payments, maintenance requests, and move-in procedures, reducing response times from hours to seconds. The Campus Heaven platform, detailed by researchers in 2024, integrated sentiment analysis on reviews and an AI chatbot for real-time assistance, showing improved user satisfaction and community engagement.

However, challenges remain regarding data privacy, surveillance concerns, and maintaining the "human touch" in student services. Ethical considerations around AI decision-making, particularly in contexts affecting student welfare, require transparent algorithms and human oversight mechanisms—aspects this research addresses through explainable recommendations and administrator moderation workflows.

2.6 Government Scholarship and Travel Clearance Systems

Government portals for scholarship management and travel clearance have increasingly moved toward digital platforms, though integration with student-facing systems remains limited. India's National Scholarship Portal (NSP) provides a unified interface for multiple scholarship schemes but requires students to navigate complex eligibility criteria and application procedures independently. Similarly, Pre-Departure Clearance systems in various countries rely on manual document submission and approval workflows that create bottlenecks.

The integration of these administrative systems with AI-powered assistance represents an underexplored opportunity. By incorporating policy documents, eligibility criteria, and procedural guidelines into a vector database accessible via natural language queries, students can receive instant, accurate guidance without navigating bureaucratic complexity. This research demonstrates how RAG-based systems can bridge the gap between government information systems and student-facing applications.

2.7 Research Gaps and Opportunities

The literature review reveals several critical gaps: (1) limited integration of accommodation recommendation with administrative workflows like PTC and scholarship applications; (2) insufficient use of RAG and vector databases specifically tailored for student travel contexts; (3) lack of comprehensive platforms unifying accommodation, policy information, and travel documentation in a single interface; and (4) limited evaluation of LLM-based systems in educational mobility contexts.

This research addresses these gaps by developing a comprehensive platform that combines state-of-the-art AI technologies with domain-specific knowledge about student travel, educational policies, and accommodation requirements. The integration of microservices architecture ensures scalability while RAG-based information retrieval ensures accuracy and trustworthiness of recommendations

3. OVERVIEW

3.1 Objective

The Student Travel LLM App aims to create an intelligent, unified platform that transforms the student accommodation and travel experience through AI-powered automation and personalization. The specific objectives are:

Primary Objectives:

Provide personalized accommodation recommendations using LLM-based semantic understanding of student preferences, budget constraints, and location requirementsAutomate Pre-Travel Clearance (PTC) document processing and approval workflows, reducing manual processing time by 60-70%Enable natural language access to government scholarship schemes, institutional policies, and travel regulations through conversational AI interfaceDeliver contextually relevant recommendations by combining semantic similarity with geospatial analysis of proximity to universities, transport, and amenities

Secondary Objectives:

Enhance transparency in accommodation pricing, availability, and host verification through centralized data aggregation and review sentiment analysisFacilitate community building among students through shared experiences, reviews, and peer-to-peer recommendationsProvide administrative tools for university staff to manage student travel approvals, track compliance, and generate analyticsEnsure data privacy and security compliance with regulations like GDPR while maintaining comprehensive audit trails

3.2 Core Components

The system architecture comprises six interconnected core components working in concert to deliver comprehensive functionality:

A. Conversational AI Interface (LLM-Powered Chatbot)
The frontend conversational layer utilizes LLM technology to interpret natural language queries and provide human-like responses. Users can ask questions like "Find affordable 2-bedroom apartments near Delhi University under ₹15,000/month with WiFi" or "What documents do I need for student visa accommodation proof?" The system understands intent, extracts relevant parameters, and formulates appropriate responses. Available 24/7 across web and mobile platforms, the chatbot maintains conversation context and provides multi-turn dialogue capabilities.

B.RAG-Based Knowledge Retrieval Engine
At the core of the system lies a Retrieval-Augmented Generation architecture that grounds AI responses in verified information. User queries are converted into vector embeddings using state-of-the-art embedding models. These vectors are matched against a comprehensive knowledge base stored in Chroma vector database, containing accommodation listings, policy documents, scholarship guidelines, and institutional regulations. Retrieved documents are re-ranked using relevance scoring that considers both semantic similarity and metadata filters (e.g., price range, location, dates). The top-ranked documents are then passed as context to the LLM for response generation, ensuring factual accuracy and reducing hallucination risks.

C. Accommodation Recommendation System
This component employs machine learning algorithms to deliver personalized housing suggestions. Features include:

Multi-criteria filtering: Budget, location radius from university, housing type (hostel/PG/apartment), amenities (WiFi, meals, AC), gender preferences, and availability dates
Hybrid scoring mechanism: Combines semantic similarity (0.4 weight), geographical proximity using PostGIS spatial queries (0.3 weight), user rating and review sentiment (0.2 weight), and price competitiveness (0.1 weight)
Collaborative filtering: Learns from similar user profiles to suggest accommodations preferred by students with comparable preferences
Visual search capability: Upload photos of desired accommodation style to find visually similar options using image embedding models

D. Pre-Travel Clearance (PTC) Automation Module
This module streamlines the administrative burden of travel documentation:Intelligent document parsing: Uses OCR and NLP to extract information from uploaded documents (ID proofs, enrollment letters, travelitineraries). Automated validation: Checks document completeness, verifies eligibility criteria against institutional policies, and flags inconsistencies. Workflow automation: Routes applications through approval chains based on configurable rules, sends notifications at each stage, and provides real-time status tracking. Integration with institutional systems: Connects with university ERP systems to verify student enrollment and academic standing

E. Government Scheme and Policy Information System
This component democratizes access to scholarship and financial assistanceinformation:Comprehensive policy database: Aggregates information from National Scholarship Portal, state government schemes, university-specific aid programs, and international scholarship opportunities.Eligibility assessment: Users input their profile (academic performance, family income, category, course) and receive personalized scholarship recommendations with application deadlinesApplication guidance: Step-by-step instructions for scholarship applications, document checklists, and common pitfall warningsReal-time updates: Monitors government portals for new schemes and deadline changes, notifying eligible users proactively

F. Analytics and Administrative Dashboard
For university administrators and platform operators, this component provides comprehensive insights:Usage analytics: Track query volumes, popular destinations, peak booking periods, and user demographics. Approval workflow monitoring: View pending PTC applications, average processing times, and bottleneck identification. Accommodation verification: Tools for moderating listings, verifying host credentials, and addressing user complaints. Compliance reporting: Generate reports on student travel patterns, scholarship utilization, and policy adherence

3.3 Working Principle

The system functions through a multi-stage AI-driven pipeline:Data Ingestion & Preprocessing: Accommodation data from partners, government sources, and users are collected via APIs and web scraping (Scrapy). Data is cleaned, normalized, geocoded (PostGIS), and enriched with text and image quality analysis. Policy and guideline documents are parsed from PDF/HTML and segmented into semantic chunks (500–1000 tokens). Embedding & Vector Storage: Text and image embeddings are generated using models like *OpenAI Ada-002*, *Cohere Embed*, or *CLIP*, and stored in *Chroma DB* with metadata (source, date, category, coordinates). Query Processing: User text or voice queries are classified by intent (e.g., accommodation search, policy query). The query is embedded using the same model, and metadata filters (price, location, etc.) are extracted for precise retrieval. Hybrid Retrieval & Re-ranking: Similarity search in *Chroma* retrieves top-k matches; *PostGIS* applies spatial filters. A weighted re-ranking (semantic, distance, sentiment, price) refines results, selecting top-n entries for response generation .LLM Response Generation: Retrieved context is formatted into prompts for LLMs (e.g., GPT-4, Claude, Llama 2), producing coherent responses with property details, policy explanations, or recommendations. Outputs are post-processed for clarity and formatting Continuous Learning :User feedback and interaction logs inform retraining and optimization of embeddings, LLM prompts, and retrieval strategies via A/B testing and periodic model updates.

4. METHODOLOGY

4.1 Existing System Analysis

Current accommodation and student support systems exhibit fragmentation and limited intelligence. Commercial platforms like OYO and Airbnb lack student-oriented features such as semester-based bookings or PTC integration. University housing portals provide limited, outdated listings and no automation. Student housing aggregators (e.g., Uniplaces) depend on keyword search without semantic understanding, while government scholarship portals operate in isolation and lack intelligent assistance.
Administrative tasks like Pre-Travel Clearance (PTC) remain manual and time-consuming.Overall, existing solutions suffer from poor interoperability, lack of personalization, limited automation, and weak multilingual support, creating inefficiencies for both students and institutions.

4.2 Conceptual System Design

The proposed system addresses these gaps through an AI-driven, modular, and scalable architecture that integrates accommodation search, PTC processing, and scholarship guidance within a unified platform. It follows key design principles of user-centricity, modularity, open integration, and security compliance.

The system consists of three main layers:
1. Presentation Layer (Web and Mobile Interfaces) for user interaction.

2. Application Layer with microservices for search, recommendations, LLM-based question answering, and workflow automation.

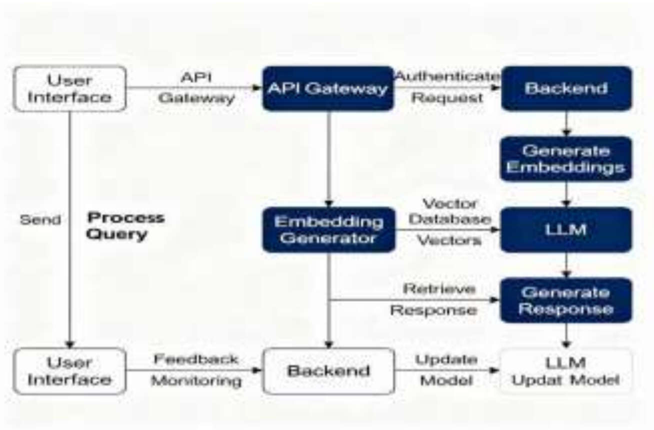
3. Data Management Layer integrating PostgreSQL with PostGIS for geospatial data, Chroma for vector embeddings, and MinIO for document storage.

4. Data flow involves query submission, semantic and geospatial retrieval, hybrid re-ranking, LLM synthesis, and response display. The architecture ensures reliability, scalability, and interoperability with institutional systems.

4.3 Prototype Design

A hybrid centralized–decentralized approach is implemented. Transactional data is maintained in PostgreSQL for consistency, while static assets like images and documents are stored in a distributed manner using MinIO or IPFS for redundancy and efficiency. This reduces latency, improves availability, and ensures data integrity through content-based addressing and encryption.

4.4 System Architecture



Architecture Description

1. User Interface (UI): The process begins when a student submits a query through the web or mobile interface. This query may involve accommodation search, travel clearance, or scholarship information. It is captured in natural language and forwarded securely for processing

2. Backend: The backend coordinates core business logic and handles data retrieval, preprocessing, and workflow management. It interacts with databases, the LLM engine, and the vector search module to fetch relevant information and build structured responses.

3. Authentication (Auth): The authentication module verifies user identity through credentials or tokens before granting system access. It ensures that only authorized users—such as verified students or administrators—can access specific data and services within the system.

4. API Gateway: The API Gateway serves as the central entry point for all requests. It routes the student's query to the appropriate service, manages rate limiting, and ensures reliable communication between the front-end and backend components while maintaining security and scalability.

5. LLM: The Large Language Model (LLM) interprets the user's natural language query and applies contextual understanding to identify intent. It works within a Retrieval-Augmented Generation (RAG) framework, integrating retrieved data to generate meaningful and accurate responses.

6. **Vector Database:** This hybrid data layer combines relational, geospatial, and vector-based storage. PostgreSQL/PostGIS manages structured and spatial data such as accommodation locations, while Chroma stores text embeddings for semantic similarity search, enabling context-aware recommendations.

Embedding Generator: The Embedding Generator transforms unstructured text, images, and documents into high-dimensional vectors that capture semantic meaning. These embeddings are crucial for efficient retrieval, allowing the system to match user intent with relevant accommodation, policy, or scholarship information.

Feedback and Monitoring: This module continuously gathers user interactions, ratings, and feedback to assess system performance and user satisfaction. Insights from monitoring are used to fine-tune embedding models and LLM responses, supporting continuous learning, model retraining, and service optimization.This architecture ensures high performance, modularity, and adaptability while supporting future integration with other educational and government platforms.

5. CONCLUSION AND FUTURE SCOPE

This research introduced the Student Travel LLM App, an AI-driven platform designed to streamline student mobility by integrating accommodation search, Pre-Travel Clearance (PTC) processing, and scholarship information within a unified system. The proposed model addresses major challenges of data fragmentation, inefficiency, and lack of personalization found in existing systems.

The system leverages Retrieval-Augmented Generation (RAG) for precise query understanding, combining semantic embeddings with geospatial filtering through PostGIS and Chroma databases The microservices-based architecture ensures scalability, modularity, and interoperability across institutional systems.

However, the system faces limitations related to the computational cost of large language models, uneven regional data quality, and dependency on institutional data integration. Ensuring continuous compliance with GDPR and FERPA regulations remains essential for sustained adoption.

Future research will focus on enhancing personalization, transparency, and scalability. Short-term goals include multimodal search using text and image inputs, AI-driven price forecasting, and automated PTC form processing through OCR and entity recognition. Medium-term objectives involve federated learning for privacy-preserving model updates, Explainable AI (XAI) for improved transparency, and AR/VR-based accommodation tours.

Long-term development aims to evolve the system into a comprehensive global student mobility ecosystem, integrating visa support, orientation modules, and predictive analytics for well-being monitoring. Integration with labor market data and blockchain-based payment verification will further enhance trust and decision-making.

This study demonstrates that combining AI, semantic search, and automation can significantly transform the higher education support ecosystem. Future advancements in responsible and explainable AI will ensure that such systems remain equitable, transparent, and centered around the needs of students.

References

[1]. Student Travel LLM: Design, Implementation, and Evaluation of an Open-Source Student Lodging Recommendation Platform. (2025). Research Paper Documentation.

[2]. Arif, Y. M., Muhtarom, M. F., & Nurhayati, H. (2023). Criteria Recommender System for Housing Selection. International Journal of Engineering Research.

[3] Maurya, A., Bajpai, D., & Singh, S. (2024). SmartStay: Tailored Student Housing Recommendations. International Journal for Research in Applied Science & Engineering Technology (IJRASET).

[4] Xu, Y., & Jiao, J. (2025). Evaluating Retrieval-Augmented Generation Strategies for Large Language Models in Travel Mode Choice Prediction. arXiv:2508.17527.

[5] Amartey, F. (2024). TravelAIgent – AI Powered Travel Guide (RAG Pipeline with Fine-Tuned LLaMA 2). GitHub Repository.

[6] Anonymous. (2025). Enhancing Generative AI with Vector Databases: Practical Applications in the Travel Industry. Dataversity.

[7] GSL Global. (2025). The AI Debate in Student Accommodation. GSL Global Blog.

[8] BCD Travel. (2021). Pre-Trip Approval Automation Helps Company Fulfil Its Duty of Care Commitment. BCD Travel Case Study.

[9] Amartey, F. (2024). TravelAIgent – AI Powered Travel Guide (RAG Pipeline with Fine-Tuned LLaMA 2). GitHub Repository.

[10] Anonymous. (2025). Enhancing Generative AI with Vector Databases: Practical Applications in the Travel Industry. Dataversity.

[11] Anonymous. (2025). Build a Retrieval-Augmented Generation (RAG) App: Part 1. LangChain Documentation.

[12] Anonymous. (2024). Microservices Travel Center System Written in Java.GitHub – Microarchitecturovisco/travel-api.

[13] Adeoti, B. E., Obuzor, T. K., Ajibola, D. A., & Omemgboji, J. C. (2025). Implementation of a RAG-based learning recommendation system for computing students. Global Journal of Engineering and Technology Advances, 24(02), 59-67.

[14] Swacha, J. (2025). Retrieval-Augmented Generation (RAG) chatbots for education: Overview and issues. Applied Sciences, 15(8), 4234.
MDPI

[15] Thüs, D., Malone, S., & Brünken, R. (2024). Exploring generative AI in higher education: a RAG system to enhance student engagement with scientific literature. Frontiers in Psychology.
Frontiers

[16] Li. D. etal. (2022). Visual recommendation for peer-to-peer accommodation: A pipeline for personalized student housing. PMC/MDPI.
PMC