ASPIRO AI: An Intelligent Mentoring and Career Guidance System Powered by Artificial Intelligence for Students and Professionals

Santhosh Kumar K
M.Sc. Computer Science, School of Computing Sciences
Vels Institute of Science, Technology and Advanced Studies (VISTAS), Pallavaram, Tamil Nadu, India

Under the Guidance of: Dr. K.S. Thirunavukkarasu, M.Sc., M.Phil., Ph.D.
Professor, Department of Computer Science and Information Technology, VISTAS, Chennai

Abstract
ASPIRO AI is an intelligent mentoring and career guidance system designed to support students and professionals through personalized, AI-driven insights and recommendations. The system integrates multiple advanced modules including an AI-powered chatbot, resume evaluation engine, skill recommendation system, and career roadmap generator to provide a comprehensive guidance ecosystem. The backend is developed using FastAPI, enabling scalable API-based communication between components, while a React frontend ensures an interactive user interface. The mentoring chatbot leverages the Google Gemini API to generate context-aware, supportive responses, simulating real-time mentorship interactions. Resume analysis functionality evaluates uploaded resumes against job descriptions using natural language processing techniques and token-based similarity scoring to provide objective feedback on candidate readiness. Additionally, the system recommends essential skills based on selected job roles and generates structured career roadmaps by analyzing user-defined skills and interests. A unique feature of ASPIRO AI is its integration of facial recognition-based attendance tracking using embedding techniques, enabling automated identification and monitoring of users. The system architecture ensures modularity, scalability, and efficient data handling through JSON-based storage and API-driven workflows. Overall, ASPIRO AI bridges the gap between education and employability by providing intelligent, data-driven, and personalized career guidance.

Keywords: *Artificial Intelligence, Career Guidance, Chatbot, Resume Analysis, Skill Recommendation, Facial Recognition, FastAPI, Google Gemini API, Natural Language Processing, Machine Learning.*

1. INTRODUCTION

In the modern digital era, career planning and professional development have become increasingly complex due to rapid technological advancements, evolving job market demands, and the vast availability of information. Students and professionals often face difficulties in identifying the right career path, understanding required skill sets, and aligning their competencies with industry expectations. Traditional career guidance systems rely heavily on manual counseling, static resources, and generalized advice, which lack personalization and scalability. As a result, there is a growing need for intelligent systems that can provide real-time, personalized, and data-driven mentoring solutions.

Artificial Intelligence (AI) has emerged as a transformative technology capable of addressing these challenges by analyzing user data, understanding preferences, and generating context-aware recommendations. Modern AI systems integrate natural language processing, machine learning, and predictive analytics to simulate human-like mentoring interactions. These advancements enable the development of intelligent platforms that not only guide users but also continuously adapt to their evolving needs and career goals.

ASPIRO AI is developed as a response to these challenges, aiming to create a unified platform that combines mentoring, skill development, and career planning into a single integrated system. By leveraging AI-based technologies, the system enhances decision-making processes and provides structured guidance tailored to individual users. The remainder of this paper is organized as follows: Section 2 reviews related literature, Section 3 analyzes limitations of existing systems, Section 4 describes the proposed system architecture, Section 5 details the implementation, Section 6 presents testing results, Section 7 discusses performance analysis, and Section 8 concludes with future directions.

2. LITERATURE REVIEW

The rapid advancement of AI and Machine Learning (ML) has significantly transformed the domain of career guidance and mentoring systems. Earlier career counseling depended on human experts who provided suggestions based on limited data, personal experience, and static evaluation techniques. These traditional approaches were often time-consuming, subjective, and lacked scalability, making them insufficient to meet the growing demands of modern education and employment systems.

2.1 Rule-Based and Early ML Approaches

Early research in career guidance focused on rule-based expert systems, which relied on predefined decision trees and if-then logic to provide recommendations [1]. Although effective for basic decision-making, these systems lacked flexibility and were unable to adapt to new or unexpected inputs. Subsequently, supervised learning algorithms such as Decision Trees, Naive Bayes, and Support Vector Machines (SVM) were employed to classify career options based on academic performance and skill data, yielding improved accuracy but struggling with unstructured inputs.

2.2 Natural Language Processing and Chatbot Systems

The emergence of NLP techniques enabled the analysis of user queries and generation of contextually relevant responses. Chatbot-based mentoring systems leveraging transformer architectures and deep neural networks gained attention due to their ability to simulate human-like interactions [2]. However, challenges such as maintaining long conversation context and avoiding biased outputs remain ongoing research problems.

2.3 Resume Analysis and Skill Recommendation

Resume analysis systems evolved from simple keyword matching to semantic analysis using TF-IDF, word embeddings, and deep learning-based scoring [3]. Skill recommendation systems employ collaborative filtering, content-based filtering, and hybrid approaches to suggest competencies aligned with user profiles and job market trends [4]. However, most systems operate in isolation and do not integrate these features into a unified pipeline.

2.4 Generative AI and Research Gap

Generative AI models based on transformer architectures have recently been applied to career planning, enabling structured path generation and mentoring conversation simulation [5]. Despite their capabilities, existing systems lack a cohesive, integrated platform that combines mentoring chatbots, resume analysis, skill recommendation, career roadmap generation, and attendance management into a single scalable framework. ASPIRO AI is designed to bridge this gap.

3. LIMITATIONS OF EXISTING SYSTEMS

A systematic review of existing career guidance platforms reveals four major architectural shortcomings that reduce their real-world effectiveness:

- Lack of integration: Most platforms operate as standalone tools. Users must navigate multiple applications separately for resume screening, skill development, and career planning, leading to fragmented experiences.
- Absence of personalization: Generalized recommendations that do not account for individual skills, interests, or career trajectories significantly reduce guidance quality.
- Reliance on static data: Recommendations based on fixed datasets quickly become outdated as industry requirements evolve. The inability to incorporate real-time data limits practical relevance.
- Limited AI utilization: Many platforms restrict AI to a single narrow function (e.g., keyword matching in resume screening), foregoing the richer contextual intelligence achievable through modern generative models.

These limitations collectively motivate the development of ASPIRO AI as a holistic, AI-native career guidance platform.

4. PROPOSED SYSTEM: ASPIRO AI

4.1 System Architecture

ASPIRO AI follows a three-tier layered architecture comprising a frontend presentation layer, a backend processing layer, and an AI inference layer. The React-based frontend provides an interactive user interface for querying the chatbot, uploading resumes, submitting job preferences, and triggering attendance. The FastAPI backend, deployed with asynchronous httpx, exposes RESTful API endpoints for each functional module. The AI layer integrates the Google Gemini API for generative language tasks, computer vision libraries for facial recognition, and NLP utilities for text processing.

Data persistence is implemented using a lightweight JSON file-based storage mechanism for embeddings and uploaded documents. A centralized configuration file (config.py) manages API keys and model parameters, ensuring maintainability.

Table 1. System Architecture Stack

Layer	Technology	Purpose
Frontend	React + Vite	Interactive user interface
Backend	FastAPI + async httpx	API routing, data processing
AI / NLP	Google Gemini API	Chatbot, roadmap, skill generation

Computer Vision	DeepFace + RetinaFace	Facial recognition & attendance
Storage	JSON File Store	Embeddings and upload metadata
Configuration	Python config.py	Centralized API key management

4.2 Module Descriptions

ASPIRO AI comprises five interdependent modules:

Chatbot Module — Provides real-time mentoring by processing user queries through the Google Gemini generative AI model. The chatbot simulates human-like conversation and maintains contextual relevance across interaction turns, returning supportive and educationally enriched responses via a dedicated FastAPI endpoint.

Resume Analysis Module — Accepts uploaded resume files (PDF/DOCX) and optional job description documents. Text is extracted and tokenized; a similarity score is computed using token-overlap analysis between the resume and job description. Users receive quantitative feedback indicating their profile alignment with the target role.

Skill Recommendation Module — Accepts a job title as input and queries the Gemini API with a structured prompt requesting a comma-separated list of role-critical competencies. The output is cleaned of Markdown artifacts and returned as a structured list.

Career Guidance Module — Accepts lists of skills and interests and generates a plaintext, step-by-step career development roadmap via the Gemini API. The module explicitly instructs the model to suppress Markdown formatting, ensuring consistent plain-text output suitable for rendering in the frontend.

Facial Recognition Module — Implements a two-stage pipeline: face enrollment (add_face endpoint) stores per-student facial embeddings extracted by DeepFace using the Facenet512 model following RetinaFace detection. During attendance processing (process_attendance endpoint), embeddings from the uploaded image are matched against the stored database using cosine similarity; identities falling below a distance threshold of 0.5 are recognized.

5. SYSTEM IMPLEMENTATION

5.1 Technologies Used

The backend is implemented in Python 3.10+ using FastAPI with CORS middleware enabled for cross-origin frontend communication. Uvicorn serves as the ASGI server for asynchronous request handling. The frontend is built with React 18 and Vite for fast hot-module-reload development. NLP preprocessing employs the NLTK library for tokenization and stopword removal. Computer vision tasks rely on the DeepFace library (Facenet512 backend) and RetinaFace for robust multi-face detection. All generative AI endpoints integrate the Google Gemini 1.5 Flash model via the google-generativeai SDK.

5.2 API Design

Five primary POST endpoints expose system functionality:

- /chatbot/ — Accepts a JSON body with a text field; returns an AI-generated mentoring response.
- /score_resume/ — Accepts multipart file uploads for resume and optional job description; returns a similarity score and file paths.
- /recommend_skills/ — Accepts a JSON body with a job_name field; returns a formatted skill list.

- /career_guidance/ — Accepts a JSON body with skills and interests arrays; returns a structured career roadmap.
- /add_face/ and /process_attendance/ — Handle facial embedding storage and recognition respectively, returning structured JSON responses.

5.3 Facial Recognition Implementation

The facial recognition pipeline converts uploaded image bytes to NumPy arrays via OpenCV, handles RGBA-to-RGB conversion, and invokes RetinaFace.detect_faces() to locate facial bounding boxes. Each detected face region is passed to DeepFace.represent() with enforce_detection=False to extract 512-dimensional embedding vectors. During recognition, cosine distance between the probe embedding and all stored embeddings for each enrolled student is computed; the identity with minimum distance below the threshold is returned.

6. SYSTEM TESTING

Comprehensive testing was conducted across four levels: unit testing (individual module endpoints), integration testing (frontend-backend API communication), system testing (end-to-end workflow validation), and user acceptance testing (usability evaluation with target users). The system was subjected to valid inputs, invalid inputs, and edge cases to ensure robustness.

Table 2. Sample Test Cases and Results

TC ID	Module	Input	Expected Output	Actual Output	Result
TC01	Chatbot	User query text	Relevant AI-generated response	Correct Response	Pass
TC02	Resume Analysis	Resume file upload	Resume similarity score generated	Score Displayed	Pass
TC03	Skill Recommendation	Job title input	List of relevant skills returned	Skills Listed	Pass
TC04	Career Guidance	Skills and interests	Structured career roadmap output	Roadmap Generated	Pass
TC05	Facial Recognition	Image with face	Recognized user name returned	Name Identified	Pass
TC06	Facial Recognition	Image without face	Error: No face detected	Error Handled	Pass
TC07	API Validation	Invalid/empty input	Proper HTTP error message	400 Error Returned	Pass

All test cases passed successfully. Error handling mechanisms correctly returned HTTP 400 and 404 responses for invalid inputs and unrecognized faces, respectively. The system maintained stability under concurrent multi-user simulation, demonstrating reliable asynchronous processing.

7. PERFORMANCE ANALYSIS

System performance was evaluated across five key metrics — response time, accuracy, reliability, throughput, and scalability — for each functional module. The chatbot module exhibited the lowest latency due to direct API pass-through to the Gemini cloud endpoint. The resume analysis and career guidance modules incurred slightly higher latency owing to file I/O and multi-token prompt construction, but remained within acceptable real-time thresholds. The facial recognition module's accuracy was influenced by input image quality and embedding database size, performing reliably under standard indoor lighting conditions.

Table 3. Module-wise Performance Summary

Module	Response Time	Accuracy	Scalability
Chatbot	Low	High	High
Resume Analysis	Medium	High	Medium
Skill Recommendation	Low	High	High
Career Guidance	Medium	High	High
Facial Recognition	Medium	High	Medium

The modular FastAPI architecture supports asynchronous request handling, enabling concurrent processing of multiple user sessions without performance degradation. Lightweight JSON-based storage minimizes I/O overhead for embedding retrieval. These design choices collectively ensure the system scales efficiently as user load increases.

8. RESULTS AND DISCUSSION

Implementation and testing confirm that ASPIRO AI successfully achieves its stated objectives. The chatbot module generates contextually coherent mentoring responses across diverse query types, including academic guidance, skill development advice, and professional development support. The resume analysis module produces meaningful similarity scores that accurately reflect alignment between candidate profiles and target job descriptions.

The skill recommendation module consistently identifies industry-relevant competencies across a broad range of job titles, with outputs structured for direct integration into learning plans. The career guidance module generates detailed, actionable roadmaps that users rated as practically useful in preliminary acceptance testing. The facial recognition module demonstrated reliable identification under controlled conditions, with graceful error handling for edge cases such as absent or multiple faces.

Compared to existing standalone systems, ASPIRO AI's integrated architecture eliminates the need for users to context-switch between multiple platforms. The unified interface — combining mentoring, career planning, and academic monitoring — represents a significant usability improvement. The use of generative AI for natural language tasks provides a level of response depth and personalization that rule-based and early ML approaches cannot match.

Limitations acknowledged in the current implementation include dependency on third-party API availability (Gemini), facial recognition accuracy sensitivity to image quality and lighting, and absence of real-time job market data integration. These are identified as priority targets for future development.

9. CONCLUSION

This paper presented ASPIRO AI, an intelligent mentoring and career guidance system that integrates an AI-powered chatbot, resume evaluation engine, skill recommendation system, career roadmap generator, and facial recognition-based attendance tracker into a unified full-stack platform. The system addresses well-documented limitations of existing career guidance tools — fragmentation, lack of personalization, reliance on static data, and limited AI utilization — through a modular, scalable architecture built on FastAPI and the Google Gemini generative AI API.

Empirical evaluation confirmed accurate, responsive performance across all five modules, with robust error handling under diverse input conditions. ASPIRO AI demonstrates that a holistic, AI-native approach to career guidance can meaningfully improve user decision-making, skill alignment, and professional development planning. The system contributes a replicable architectural blueprint for next-generation intelligent mentoring platforms.

9.1 Future Enhancements

Planned enhancements include: integration of large-scale pre-trained language models to further improve chatbot contextual depth; real-time job market data feeds from external APIs for dynamic skill and career recommendations; multilingual support to broaden accessibility; semantic deep learning-based resume scoring to replace token-overlap heuristics; cloud deployment for improved scalability; and user progress analytics dashboards to enable longitudinal career tracking.

REFERENCES

[1] J. Smith and A. Kumar, "Career Prediction Using Machine Learning Algorithms," International Journal of Computer Applications, vol. 178, no. 12, pp. 15-22, 2020.

[2] R. Sharma, P. Singh, and K. Verma, "AI-Based Chatbot for Student Guidance Using Natural Language Processing," IEEE Access, vol. 9, pp. 45678-45689, 2021.

[3] M. Lee and K. Park, "Automated Resume Screening Using NLP Techniques," Journal of Information Processing Systems, vol. 16, no. 3, pp. 567-580, 2020.

[4] S. Gupta and P. Verma, "Skill Recommendation System Using Collaborative Filtering," Procedia Computer Science, vol. 167, pp. 1120-1129, 2020.

[5] A. Brown, D. Wilson, and R. Taylor, "Generative AI for Career Path Recommendation," IEEE Transactions on Artificial Intelligence, vol. 3, no. 2, pp. 210-220, 2022.

[6] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient Estimation of Word Representations in Vector Space," Proc. ICLR, 2013.

[7] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," Proc. NAACL-HLT, pp. 4171-4186, 2019.

[8] A. Vaswani et al., "Attention Is All You Need," Advances in Neural Information Processing Systems, vol. 30, 2017.

[9] H. Schroff, D. Kalenichenko, and J. Philbin, "FaceNet: A Unified Embedding for Face Recognition and Clustering," Proc. CVPR, pp. 815-823, 2015.

[10] F. Chollet, Deep Learning with Python. Manning Publications, 2018.

[11] I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning. MIT Press, 2016.
[12] D. Jurafsky and J. H. Martin, Speech and Language Processing, 3rd ed. Pearson, 2021.
[13] S. Russell and P. Norvig, Artificial Intelligence: A Modern Approach, 4th ed. Pearson, 2020.
[14] Y. LeCun, Y. Bengio, and G. Hinton, "Deep Learning," Nature, vol. 521, pp. 436-444, 2015.
[15] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," Proc. CVPR, pp. 770-778, 2016.