# AI-Based Career Recommendation System Based on Their Skills

Dr.S.Rinesh[1] Mrs.U.L.Sindhu[2] Mr.S.Kalaiselvan[3] Mr.J.Madhavan[4] Mr.S.Karthikeyan[5]

[1] Professor,V.S.B College Of Engineering Technical Campus, Coimbatore, Tamil Nadu, India

[2]Assistant Professor,V.S.B College Of Engineering Technical Campus, Coimbatore, Tamil Nadu, India

[345]Undergraduate Students,V.S.B College Of Engineering Technical Campus, Coimbatore, Tamil Nadu, India

**ABSTRACT**

Traditional career guidance systems often lack personalization, adaptability, and accessibility, leading to suboptimal decisions in a rapidly changing job market. This study introduces Sankalp, an AI-driven career guidance framework that combines semantic reasoning, emotion analysis, and adaptive learning to provide inclusive and explainable recommendations. Built on a multi-agent hybrid architecture, Sankalp integrates modules for semantic matching using Sentence-BERT (SBERT), emotion-aware feedback through VADER sentiment analysis, a Naukri API-based job trends agent, a knowledge graph, and a voice assistant supporting English, Hindi, and Kannada. A reinforcement learning-based fusion layer dynamically refines recommendations based on real-time user feedback. Evaluated against a Curated Ground Truth (CGT) dataset using 160 participant profiles, Sankalp achieved a Top-1 Hit Rate of 77.5%, a Top-3 Hit Rate of 90.5%, and a mean user satisfaction score of 4.71/5.0. Average end-to-end response time was 1.25 seconds.

**KeyWords:** AI-driven career guidance, emotion analysis, knowledge graph, NLP, personalized recommendation, Sentence-BERT, VADER sentiment analysis.

## I. INTRODUCTION

In the increasingly dynamic global landscape, effective career guidance is critical for individual fulfillment, economic stability, and societal contribution. A lack of clear career direction leads to underemployment, skill mismatches, and reduced productivity. This challenge is particularly acute in developing nations: India faces a student-to-counselor ratio of approximately 1:3,000, far below the globally recommended 1:250.

This paper introduces Sankalp, an AI-driven career guidance system designed to democratize access to personalized and dynamic career counseling, bridging the gap between educational outcomes and employment opportunities. India's 2020 National Education Policy (NEP 2020) explicitly calls for integrating career counseling at all schooling stages, demanding scalable, accessible, and effective solutions.

Key barriers to career guidance delivery include: (1) geographic access—traditional services are concentrated in urban centers; (2) counselor shortage—the sheer scale of India's student population overwhelms qualified professionals; (3) language barriers—India's linguistic diversity excludes many students when guidance is only available in English; and (4) motivational challenges—students often feel overwhelmed by the multitude of career choices.

Sankalp's foundational methodology is rooted in established psychological models of vocational choice, specifically person-environment fit theory. Its multi-agent architecture incorporates Emotion-aware NLP via VADER sentiment analysis, seamless voice and text I/O via Google Cloud services, and multilingual support in English, Hindi, and Kannada. A novel rule-based Reasoning Engine enhanced with a Reinforcement Learning (RL) layer ensures both transparency and adaptability.

## II. LITERATURE SURVEY

The landscape of AI-driven career guidance has grown substantially. Ayeni et al. highlighted the need for personalized learning and the challenges of the digital divide, algorithmic bias, and ethical AI adoption in education. Sankalp directly addresses these concerns through its inclusive multimodal and multilingual design.

Jadhav et al. proposed an NLP-powered career chatbot for Indian students using neural network intent recognition, but lacked comprehensive multilingual voice support. Avinash et al. integrated NLP, RL, and Collaborative Filtering for skill-gap analysis, yet operated within a text-only framework. Birajdar et al. introduced a career counseling web application with real-time job market data but without voice interaction or robust multilingualism. Kulugh et al. built a system on Holland's Vocational Theory and Social Cognitive Theory but lacked multimodal features.

VocaVisionary (Arshad et al.) utilized a fine-tuned PaLM 2 model for conversational guidance, with voice recognition noted as future work and no multilingual support. In contrast, Sankalp's multimodal and multilingual capabilities are core design features from inception.

### A. Comparison with Hosted Platforms

Naukri Career Navigator relies on keyword matching and offers no conversational AI or voice support. The National Career Service (NCS) Portal provides content in Indian languages but lacks AI-driven conversational capabilities. CareerAI and MyNextMove are predominantly text-based, English-only platforms lacking real-time emotion recognition or graphical explainability.

**TABLE 1. Comparison of AI-Powered Career Guidance Systems**

| Features | CareerAI | MyNextMove | Naukri Nav. | NCS Portal | Sankalp |
|---|---|---|---|---|---|
| Language Support | English | English | English (primary) | Multiple Indian (text only) | English, Hindi, Kannada (V+T) |
| Input Modality | Text | Text | Text | Text | Voice & Text |
| Personalization | Basic | Moderate | Basic | Moderate | High (Interests, Skills, Marks) |
| Emotion Awareness | No | No | No | No | Yes (VADER) |
| AI Explainability | No | Limited | Limited | No | Yes (Knowledge Graph) |
| Feedback Loop | No | No | No | No | Explicit (RL-based) |
| Offline Use | No | No | No | No | Yes |

## III. PROPOSED SANKALP SYSTEM ARCHITECTURE

The Sankalp system operates on a client-server architecture with a React.js front-end and a Flask-based back-end API. The architecture is divided into four major layers:

- User Experience Layer: React.js interface supporting voice and text input with real-time feedback display.
- Application Integration Layer: Orchestrates data flow between front-end and specialized back-end agents.
- Cognitive Processing Layer (Multi-Agent System): Houses the Input Interpretation Module, Emotion Analyzer, Contextual Relevance Engine (SBERT), Deterministic Reasoning Engine, Ontological Inference Engine, and Labour Market Intelligence Extractor.
- Data Management Layer: PostgreSQL database storing career paths, skill mappings, job trends, and user feedback.

### A. Core AI Logic

**1) Semantic Matching Module**

The Semantic Matching module uses Sentence Transformers (SBERT) with a dual-model strategy: the 'all-MiniLM-L6-v2' model for Online Mode and 'paraphrase-multilingual-mpnet-base-v2' for Offline Fallback Mode. Career attributes (skills, subjects, traits, descriptions) are concatenated and encoded into dense vector embeddings. Cosine similarity between the user's embedding and each career's embedding produces a prioritized recommendation list.

**2) Rule-Based Reasoning and Adaptive Fusion Layer**

The Rule-Based Reasoning Engine encodes domain-specific eligibility constraints—academic prerequisites, performance thresholds, mandatory skill sets—to ensure transparent, deterministic filtering. A Q-Learning RL layer acts as an adaptive fusion mechanism, dynamically adjusting recommendation weights based on explicit user feedback. A positive reward (+1) is assigned for accepted recommendations; a negative reward (-1) for rejections.

**3) Emotion Analysis Module**

VADER (Valence Aware Dictionary and Sentiment Reasoner) provides real-time sentiment detection, enabling Sankalp to respond empathetically. For Hindi and Kannada input, a translation-first approach is applied before VADER analysis. Performance variance across languages is within an acceptable 5% drop from English baseline.

**4) Knowledge Graph Module**

The Knowledge Graph (built using NetworkX) represents careers, skills, academic subjects, and personality traits as interconnected nodes with directed edges. The 'neighbors of neighbors' traversal identifies related careers, providing users with alternative pathways and enhancing the system's explainability.

**5) Voice Assistant Module: Dual-Mode Architecture**

**Online Mode:** Google Cloud Speech-to-Text, Text-to-Speech (Wavenet voices), and Translation APIs deliver high-fidelity multilingual interaction. **Offline Fallback Mode:** The Vosk API handles on-device STT/TTS; lightweight Hugging Face multilingual models handle translation and semantic analysis, guaranteeing service continuity in low-bandwidth or offline environments.

## IV. SYSTEM IMPLEMENTATION

The system is built on a robust client-server model. The React.js front-end collects user profile data via a sequential guided questionnaire (interests, marks, skills, personality traits, work environment preferences), presents personalized recommendations, and renders detailed roadmaps. The Flask back-end exposes RESTful endpoints for session management, profile updates, recommendation generation, voice processing, and feedback recording.

### A. Technology Stack

**TABLE 2. Core Components and Tools Used in Sankalp**

| Component | Library | Remarks |
|---|---|---|
| Voice Input | Google STT + Vosk API | Dual-mode (Online/Offline) |
| NLP Embeddings | SBERT (paraphrase-multilingual-mpnet-base-v2) | Semantic matching, cloud + offline |
| Sentiment Analysis | VADER | Emotion-aware scoring |
| Multilingual Support | Google Translate API + HuggingFace Models | Hindi, English, Kannada |
| Backend Framework | Flask | Modular REST APIs |
| Database | PostgreSQL | Tracks sessions and decisions |
| Job Trends Module | Naukri Commercial API | Real-time market alignment |
| Knowledge Graph | NetworkX | Links skills, subjects |
| Rule Engine | Python Logic + Q-Learning | Adaptive, personalized scoring |

### B.Recommendation and Roadmap Generation

Once a user profile is complete, the Core AI Logic computes a ranked list of careers using semantic similarity. Previously rejected careers are filtered out. For each accepted career, the system generates a detailed roadmap including: educational pathways, key skills to develop, related subjects, market insights, top companies, and curated online course links (sourced dynamically via the Naukri API). All roadmap content is translated into the user's selected language before presentation.

The feedback loop records accepted and rejected careers. Accepted choices update the RL Q-table with a positive reward; rejected choices are added to an exclusion list for the session, enabling dynamic refinement and preventing re-presentation of unsuitable options.

## V. EXPERIMENTAL SETUP

### A. Dataset and Cohort

The career repository contains 85 career categories, 510 unique keywords, and 58,100 characters of raw career description data. System precision was tested using 160 diverse user profiles created via stratified sampling (urban, semi-urban, and rural backgrounds). Accuracy was validated against a Curated Ground Truth (CGT) dataset, where each profile's target career was predetermined from established government and industry career mapping documentation.

### B. Baseline Comparisons

Sankalp was evaluated against two baselines: (A) a Rule-Only model using only the deterministic rule engine, and (B) an SBERT-Only model using purely semantic similarity without rule constraints. The ablation study incrementally added modules to demonstrate each component's contribution.

### C. RL Configuration

Q-Learning hyperparameters: learning rate (alpha) = 0.1; discount factor (gamma) = 0.9; epsilon-greedy strategy with initial epsilon = 1.0 decaying to 0.1 over 100 epochs. Reward function: +1 for accepted recommendations, -1 for rejections.

### D. Statistical Rigor

All key metrics are reported with 95% confidence intervals. Comparative statistical significance between the Sankalp Hybrid and baselines was established using paired two-sample t-tests ($p < 0.05$). Note: a planned external expert evaluation with certified counselors (for Cohen's kappa inter-rater agreement) was postponed due to resource constraints and is targeted for a follow-up study.

## VI. RESULTS AND ANALYSIS

### A. Performance Metrics

#### 1) Recommendation Generation Time

Text input mode: average recommendation time 0.18-0.20 seconds (near-instantaneous). Voice input mode: average 0.30-0.38 seconds (including STT overhead). Total end-to-end latency (incorporating all components including cloud API calls): approximately 1.25 seconds, meeting real-time conversational requirements.

**TABLE 3. Latency Breakdown Across Operating Conditions (Mean ± SD)**

| Component | Condition | Me Mean ± SD (ms) | Notes |
|---|---|---|---|
| Recommendation Engine | Optimal | 350 ± 45 | SBERT + KG + RL Fusion |
| Recommendation Engine | Low Bandwidth | 365 ± 50 | Minor delay in trigger |
| Cloud API (STT/TTS) | Optimal | 550 ± 70 | Google Cloud |
| Network Transport | Optimal | 50 ± 20 | ~100 Mbps link |
| Network Transport | Low Bandwidth | 800 ± 120 | 2 Mbps simulated delay |
| Total End-to-End | Optimal | 1000 ± 95 | ~1.0s total |
| Total End-to-End | Low Bandwidth | 1765 ± 150 | ~1.8s total |

**2) Multilingual Consistency**

**TABLE 4. Component Performance Across Supported Languages**

| Language | Accuracy(%) | Vader(%) | Sbert (%) | Sim(%) |
|---|---|---|---|---|
| English | 88.3 | 0.86 | 67.5 | 0.81 |
| Hindi | 84.1 | 0.82 | 66.8 | 0.79 |
| Kannada | 83.5 | 0.81 | 66.1 | 0.78 |

Performance disparity between languages is minimal. The slight VADER accuracy decrease for Hindi/Kannada (attributed to the translation step) is within an acceptable < 5% range, validating the translation-first strategy.

**3) Core Recommendation Accuracy — Ablation Study**

**TABLE 5. Model Performance Across Incremental Configurations (Ablation Study)**

| Config. | Modules Used | Top- 1 (%) | Top-3 (%) | Top-5 (%) | Satisfaction (%) |
|---|---|---|---|---|---|
| A (Baseline) | Rule Engine Only | 52.5 | 74.0 | 84.5 | 3.60 |
| B (Baseline) | SBERT Only | 64.4 | 86.0 | 91.2 | 4.10 |
| C | Rule + SBERT | 68.1 | 88.5 | 93.8 | 4.35 |
| D | Rule + SBERT + VADER | 71.3 | 89.4 | 94.6 | 4.49 |
| E | Rule + SBERT + VADER + KG | 75.9 | 90.1 | 95.8 | 4.58 |
| F | Rule + SBERT + VADER + KG + SA | 76.7 | 90.3 | 95.9 | 4.66 |
| G (Sankalp Hybrid) | All + RL Feedback | 77.5 | 90.5 | 96.1 | 4.71 |

The Sankalp Hybrid model (Configuration G) achieved a Top-3 Hit Rate of 90.5% (95% CI: [88.1%, 92.9%]) — statistically superior to Rule-Only (74.0%) and SBERT-Only (86.0%) baselines ($p < 0.01$ both comparisons). The addition of the Knowledge Graph and RL Feedback Loop produced the largest incremental gains in Top-1 precision.

**4) User Satisfaction**

Post-interaction surveys from 160 participants yielded a mean satisfaction score of 4.71/5.0 (95% CI: [4.58, 4.72]). Rating distribution: 77.7% rated 5 stars, 17.9% rated 4 stars, 2.7% rated 3 stars, 1.8% rated 2 stars, and 0% rated 1 star. Database-driven analysis revealed 91% of users were satisfied with the platform's career predictions.

## VII. DISCUSSION

The experimental results confirm the efficacy and architectural robustness of the Sankalp system. The internal AI logic contributes a minimal mean latency of ~350 ms (driven by precomputed SBERT embeddings and fast Q-table lookups), while total end-to-end latency of 1.25 seconds accommodates real-time multimodal interaction.

The ablation study (Table 5) explicitly demonstrates that the hybrid fusion architecture is responsible for superior performance. The Knowledge Graph provided the most significant gain in Top-1 precision (71.3% to 75.9%), and the RL layer further improved adaptability. The high User Satisfaction Score of 4.71/5.0 reflects the value of the VADER-enabled empathetic interaction and the inclusive multilingual voice/text experience.

### A. Current Limitations

- Rule-Model Rigidity: Rule-based filters may over-constrain personalization in edge cases where user profiles fall outside predefined conditions, affecting flexibility for novel or ambiguous inputs.

- LLM Baseline Omission: Direct comparison with GPT-4/Gemini was not conducted due to computational cost, lack of deterministic explainability, and ethical compliance (IRB/GDPR) concerns around non-auditable external APIs.

## VIII. FUTURE WORK

- Formal Expert Validation: Large-scale A/B evaluation with certified career counselors using Cohen's Kappa (k) for inter-rater reliability.
- GPT-based Integration: Integration of large-scale language models for deeper causal reasoning, contextual conversation continuity, and nuanced career storytelling.
- Mobile Deployment: Android application with offline mode (basic recommendations + cached profiles), multilingual voice interaction, and push notifications for job and roadmap updates.
- Gamified Interaction Layer: Career quizzes, mission-style roadmap unlocking, and achievement mechanics to increase engagement among school-age users.
- Scalability: Optimization of back-end infrastructure for massive scalability with efficient load balancing for Naukri API requests and cloud-based voice processing.

## IX. REPRODUCIBILITY

To promote transparency and full reproducibility, the core components of the Sankalp framework will be released upon publication. The open-access GitHub repository will contain: multi-agent orchestration logic, the hybrid recommendation engine (SBERT module, Rule-Based Engine, RL fusion layer), the anonymized Curated Ground Truth (CGT) dataset, and the Knowledge Graph schema.

## X. ETHICS AND RESPONSIBLE AI

### A. Fairness and Bias Mitigation

Sankalp uses multilingual SBERT embeddings and interpretable rule-based logic to minimize algorithmic bias. Unlike opaque black-box systems, the transparent Knowledge Graph and scoring mechanism allow users to understand why a career was suggested. Support for regional languages reduces urban and English-language biases that typically exclude underserved communities.

### B. Privacy and Compliance

All participants provided informed consent prior to participation. User profile data is stored in PostgreSQL with AES-256 encryption. No personally identifiable information is shared with third parties. Data retention is strictly limited to the study duration. Sankalp complies with GDPR principles and India's Information Technology Act.

### C. Transparency and User Autonomy

Users can accept, reject, or request alternative career suggestions at any point, reinforcing human agency. All scoring and recommendation pathways are auditable via the deterministic rule engine and Knowledge Graph. This commitment to responsible AI positions Sankalp as a standard for ethical, scalable, and user-centered AI in education.

## XI. CONCLUSION

This paper presented Sankalp, an AI-powered, multilingual, and emotionally aware career guidance system tailored for India's diverse and large-scale educational landscape. By integrating deterministic rule-based reasoning with dynamic semantic understanding (SBERT) and a Reinforcement Learning adaptive layer, Sankalp delivers highly personalized and accessible career guidance. Its modular multi-agent design supports real-time voice/text input, full multilingual support (English, Hindi, Kannada), and explainable AI via a visual Knowledge Graph.

Validated against a Curated Ground Truth dataset, Sankalp achieved a Top-3 Hit Rate of 90.5% (95% CI: [88.1%, 92.9%]) — significantly superior to all baselines — with a core logic latency of 350 ms and a high User Satisfaction Score of 4.71/5.0. These results validate the strength of the hybrid design in both precision and usability.

Beyond technological novelty, Sankalp aligns with India's NEP 2020 objectives and the United Nations Sustainable Development Goals (SDG 4: Quality Education; SDG 8: Decent Work and Economic Growth), providing a scalable, intelligent solution that addresses educational inequities and prepares youth for a dynamic workforce.

## REFERENCES

[1] O. O. Ayeni et al., 'AI in education: A review of personalized learning and educational technology,' GSC Adv. Res. Rev., vol. 18, no. 2, pp. 261-271, Feb. 2024.

[2] R. Jadhav et al., 'NLP-powered data-driven career decision methodology based on machine learning,' in Proc. ICOECA, Apr. 2024, pp. 353-358.

[3] B. Avinash et al., 'AI driven career guidance and recommendation,' Int. Res. J. Modernization Eng. Technol. Sci., vol. 7, no. 4, pp. 1-5, Apr. 2025.

[4] A. Birajdar et al., 'AI-powered career counselling,' Int. Res. J. Eng. Technol., vol. 11, no. 11, p. 543, Nov. 2024.

[5] V. E. Kulugh et al., 'Artificial intelligence-powered personalised career guidance system,' Dutse J. Pure Appl. Sci., vol. 11, no. 1c, pp. 92-104, Apr. 2025.

[6] T. K. Arshad et al., 'VocaVisionary: A career guidance chatbot,' in Proc. 2nd Int. Conf. Adv. Inf. Technol. (ICAIT), Jul. 2024, pp. 1-6.

[7] N. El-Khalili et al., 'AI-powered career guidance system,' in Proc. 1st ICCIAA, Apr. 2025, pp. 1-6.

[8] S. V. Dewasurendra et al., 'Cutting-edge AI-driven higher educational and career advisory platform,' in Proc. ICSES, Dec. 2024, pp. 1-7.

[9] R. Waikar et al., 'Student career guidance using generative AI,' in Proc. ICAIQSA, Dec. 2024, pp. 1-6.
[10] B. Hou et al., 'KG-PLPPM: A knowledge graph-based personal learning path planning method,' arXiv:2408.15620, 2025.
[11] M. A. Akber et al., 'Personality prediction based on contextual feature embedding SBERT,' in Proc. IEEE TENSYMP, Sep. 2023, pp. 1-5.
[12] S. B. Kolli et al., 'VADER-RLA: A reinforcement learning-augmented sentiment analysis model,' J. Theor. Appl. Inf. Technol., vol. 103, no. 10, p. 4462, May 2025.
[13] 'India needs 1.4 million career counsellors to maintain a globally acceptable ratio,' The Financial Express, 2018.
[14] J. L. Holland, Making Vocational Choices: A Theory of Vocational Personalities and Work Environments, 3rd ed., PAR, 1997.
[15] D. E. Super, 'A life-span, life-space approach to career development,' J. Vocational Behav., vol. 16, no. 3, pp. 282-298, Jun. 1980.
[16] R. W. Lent et al., 'Toward a unifying social cognitive theory of career and academic interest,' J. Vocational Behav., vol. 45, no. 1, pp. 79-122, Aug. 1994.