IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940



International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 10, October 2026

DOI: 10.17148/IJARCCE.2026.1510114

A Hybrid AI Recommender System for Dynamic Career Guidance

Addhwaith S Ajith[1], Aaron John Joy[2], Vaishnav Biju[3], Ameesha J S[4], Jency Mariam John[5], Girija VR[6]

Department of Computer Science and Engineering (Artificial Intelligence and Machine Learning),

College of Engineering Kottarakkara, Kerala, India[1-6]

Abstract: The rapid transition from traditional, manual career counseling to dynamic, automated recommendation frameworks represents a vital imperative in modern education and workforce alignment. Conventional career guidance platforms heavily rely on static rule-based psychometric profiling (such as Holland's RIASEC framework) or isolated machine learning classifiers. However, standalone approaches consistently suffer from critical deficiencies, notably data sparsity, severe cold-start limitations for new candidates, lack of contextual text understanding, and black-box operational opacity. This paper presents a detailed technical architecture and comprehensive research paper on a novel Hybrid AI Recommender System engineered for dynamic career path generation. By unifying supervised classification models (Random Forest, XGBoost, Support Vector Machines), transformer-based Natural Language Processing (BERT embeddings with GRU-driven attention mechanisms), Graph Neural Networks (GNNs) over domain ontologies for Explainable AI (XAI), and a Retrieval-Augmented Generation (RAG) conversational interface, the system establishes a transparent, multi-dimensional decision engine. We formalize the skill gap quantization model using Euclidean distance vector spaces, delineate the technological stack (Python, Scikit-learn, PyTorch, React, Vite, TypeScript), and evaluate the pipeline using standard empirical frameworks. Experimental benchmarks demonstrate that the hybrid approach achieves superior predictive precision (91%), recall (88%), and cold-start error reduction, bridging theoretical psychometrics with dynamic labor market demands.

Keywords: Career guidance, hybrid recommender systems, machine learning, natural language processing, explainable AI, knowledge graphs, RIASEC model, skill gap quantization, semantic matching, RAG architecture.

I. INTRODUCTION

The evolution of modern industrial sectors and global labor markets has fundamentally disrupted standard employment pathways. Rapid technological advancement, automated workflows, and emerging interdisciplinary fields require candidates to constantly upgrade their technical and domain-specific skill sets. Consequently, young adults and educational candidates face an increasingly complex decision space when selecting academic specializations and professional career vectors. Traditional career counseling methodologies—characterized by manual face-to-face evaluations, standardized paper questionnaires, and legacy psychometric tests—are fundamentally unscalable and fail to capture real-time shifts in skill requirements across global industries [1], [2], [3].

II. RESEARCH BACKGROUND

This section examines the limitations of existing career guidance approaches and states the objectives and contributions of the present research.

A. Limitations of Traditional and Standalone Machine Learning Systems

Early automated solutions attempted to address scalability by adopting static, rule-based profiling logic derived from classic psychological constructs, most notably Holland's occupational interest inventory (Realistic, Investigative, Artistic, Social, Enterprising, and Conventional) and the Big Five personality trait matrix. While psychometric testing provides valuable baseline insights into a candidate's intrinsic traits and behavioral dispositions, rule-based systems suffer from inherent rigidity. They treat career domains as discrete, unyielding categories and cannot dynamically update recommendations when market demands evolve or when candidate capabilities expand beyond initial profiles [4], [5].

To overcome rigidity, recent literature explored classical supervised machine learning algorithms, including Decision Trees, Support Vector Machines, Random Forests, and Gradient Boosted Decision Trees such as Extreme Gradient Boosting. Additionally, collaborative filtering techniques were adapted from e-commerce frameworks to align student

 This work is licensed under a Creative Commons Attribution 4.0 International License

DOI: 10.17148/IJARCCE.2026.1510114

profiles based on peer behavioral patterns. However, standalone machine learning models introduce three severe operational bottlenecks: [6], [7], [8]

1. The Cold-Start Bottleneck & Data Sparsity: Collaborative filtering relies heavily on user interaction logs and historical profile ratings. When a novel candidate enters the system with zero historical feedback, pure collaborative filtering models fail entirely to generate accurate or contextually relevant career paths [7].

2. Lexical Overlap Inadequacy: Standalone models processing unstructured texts (for example, academic transcripts, resumes, project portfolios) historically relied on simple keyword matching or basic Term Frequency-Inverse Document Frequency vectors. These approaches completely miss latent semantic relationships, failing to recognize that "Machine Learning Engineer" and "Data Scientist specializing in Predictive Modeling" share major overlapping competence spaces despite minimal exact keyword duplication [9], [10].

3. Operational Opacity and the Black-Box Dilemma: Advanced deep neural networks and ensemble learning algorithms frequently yield high mathematical predictive accuracy but operate without explicit domain reasoning. In high -stakes educational decision-making, recommendations delivered without transparent, logical explanations fail to earn user trust and fail to provide actionable steps for remediation [11], [12].

B. Objectives and Research Contributions

To address these multi-faceted challenges, this research introduces an integrated Hybrid Artificial Intelligence Recommender System. The proposed framework harmonizes psychological profiling, supervised machine learning, contextual transformer Natural Language Processing, ontology-driven Graph Neural Networks, and conversational Generative Artificial Intelligence. The key contributions of this paper are organized as follows: [6], [13], [14]

- Design of a multi-tiered hybrid architecture that eliminates the cold-start problem by fusing structured psychometric traits with unstructured textual vector representations [6], [12].
- Realization of deep semantic matching utilizing Bidirectional Encoder Representations from Transformers embeddings and Gated Recurrent Units to measure candidate-job profile affinity [9], [10], [15].
- Formulation of an explicit Skill Gap Quantization model based on Euclidean distance metrics across multidimensional specialization vectors [16], [17].
- Deployment of an Explainable Artificial Intelligence layer using Semantic Web Ontologies and Graph Neural Networks to map clear skill progression trajectories [11].
- Architectural integration of a context-aware Retrieval-Augmented Generation agent that provides real-time, conversational guidance to candidates [12], [14].

III. LITERATURE REVIEW AND TAXONOMY MATRIX

The progression of automated career guidance systems spans several decades, advancing from expert systems to complex hybrid deep learning architectures. This section synthesizes recent literature, establishing a comparative taxonomy of foundational approaches [8], [12], [14], [18].

A. Psychometric Profiling and Rule-Based Frameworks

The foundational framework for automated psychological evaluation in career recommendation rests on Holland's occupational interest theory. Holland posited that individuals and work environments can be categorized into six personality domains: Realistic (R), Investigative (I), Artistic (A), Social (S), Enterprising (E), and Conventional (C). In classical exp ert systems, candidate responses to psychometric items generate a top-three trait code (for example, "SAI" or "RIE"), which is subsequently mapped against static lookup databases. While psychometric alignment correlates strongly with long-term job satisfaction, standalone rule-based mappings cannot incorporate dynamic variable inputs such as university grade point average, specific technical certifications, coding project experience, or fluctuating regional job market demands [4], [5].

B. Supervised Machine Learning and Collaborative Filtering

With the proliferation of educational data mining, researchers transitioned toward supervised classification. Studies utilizing algorithms such as Random Forest and Support Vector Machines achieved higher domain classification accuracy by feeding academic performance indicators alongside demographic vectors. Simultaneously, collaborative filtering models (both user-based and item-based) attempted to leverage historical student performance logs to project future career trajectories. However, empirical literature consistently highlights that educational datasets exhibit high sparsity (often exceeding 98%), rendering standard matrix factorization techniques prone to severe error margins,

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940



International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 10, October 2026

DOI: 10.17148/IJARCCE.2026.1510114

particularly during candidate onboarding [7], [8].

TABLE I TAXONOMIC COMPARISON OF CAREER RECOMMENDATION ARCHITECTURES

Architecture	Core Technologies	Primary Advantages	Key Operational Limitations
Rule-Based Systems	Holland occupational interest model, Big Five personality model, Expert Decision Trees	High interpretability, low computational overhead, simple deployment	Extremely rigid, static lookup tables, fails with novel industry roles [4], [5].
Collaborative Filtering	Matrix Factorization, K-Nearest Neighbors, Singular Value Decomposition	Captures implicit peer patterns, no domain knowledge required.	Severe cold-start problem, high data sparsity, black-box output [7].
Standalone Machine Learning	Decision Trees, Random Forest, Extreme Gradient Boosting, Support Vector Machines	Handles structured numerical marks and demographics effectively	Incapable of parsing unstructured text (resumes), ignores domain context [9], [10].
Semantic Natural Language Processing	Term Frequency-Inverse Document Frequency, Word2Vec, Bidirectional Encoder Representations from Transformers Embeddings, Gated Recurrent Units	Matches unstructured resumes based on deep semantic meaning	Lacks psychological trait alignment, high computational complexity [9], [10], [15].
Hybrid Artificial Intelligence (Proposed)	Supervised Machine Learning + Transformers + Graph Neural Networks; Ontologies + Retrieval-Augmented Generation	Cold-start resilient, semantic-aware, fully explainable, dynamic	Requires complex multi-stage pipeline alignment and graph traversal [6], [12], [13].

C. Semantic Matching and Natural Language Processing

The emergence of pre-trained transformer architectures, such as Bidirectional Encoder Representations from Transformers, marked a fundamental shift in processing unstructured candidate data. Rather than treating resumes as simple bags -of-words, transformer models capture context-dependent semantic representations. Sentence-transformer variants allow candidates' textual portfolios and open-ended interest statements to be mapped into high-dimensional vector spaces. By calculating vector similarity against normalized job role descriptions, semantic natural language processing engines overcome the limitations of exact keyword matching, successfully aligning lateral skills across disparate domains [9], [10], [15].

D. Explainable Artificial Intelligence and Knowledge Graphs

The latest paradigm in career recommender research emphasizes transparency through Explainable Artificial Intelligence. Semantic Web technologies, including the Resource Description Framework and the Web Ontology Language, allow researchers to construct structured domain ontologies. Knowledge Graphs model entities (such as Student, Skill, Course, and Career Role) alongside explicit relationships (for example, REQUIRES_SKILL, IS_PREREQUISITE_FOR). Graph Neural Networks operating over these ontologies enable both precise path generation and logical explanation traces, transforming opaque prediction outputs into transparent, actionable roadmaps [11], [16], [17].

IV. SYSTEM ARCHITECTURE

The proposed system is designed around a multi-tiered, modular pipeline engineered to process heterogeneous user inputs, ranging from structured test scores and academic marks to unstructured text resumes and interactive conversational queries.

A. Architectural Blueprint

The overall architecture comprises four primary operational layers: [12], [13]

1. Data Ingestion & Feature Extraction Layer: Captures structured academic marks, standardized psychometric test results (Holland occupational interest and Big Five scores), and raw Portable Document Format resumes or unstructured project summaries.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 10, October 2026

DOI: 10.17148/IJARCCE.2026.1510114

2. Dual Scoring & Classification Engine: *Psychometric & Machine Learning Branch:* Utilizes ensemble classifiers (Random Forest, Extreme Gradient Boosting) to generate primary career domain probability distributions. *Semantic Natural Language Processing Branch:* Passes unstructured text through Bidirectional Encoder Representations from Transformers embeddings and Gated Recurrent Unit attention networks to compute semantic cosine similarity against candidate job role vectors [7], [12].

3. Knowledge Graph & Skill Gap Quantization Engine: Maps top candidate matches into a domain-specific Knowledge Graph, calculating multi-dimensional Euclidean skill distances and extracting exact missing competencies [11], [16], [17].

4. Conversational Retrieval-Augmented Generation & User Dashboard Layer: Delivers transparent recommendations alongside a conversational Artificial Intelligence counselor (powered by Generative Pre-trained Transformer 3.5 Turbo and dynamic Retrieval-Augmented Generation context retrieval) through an interactive React frontend [14], [18].

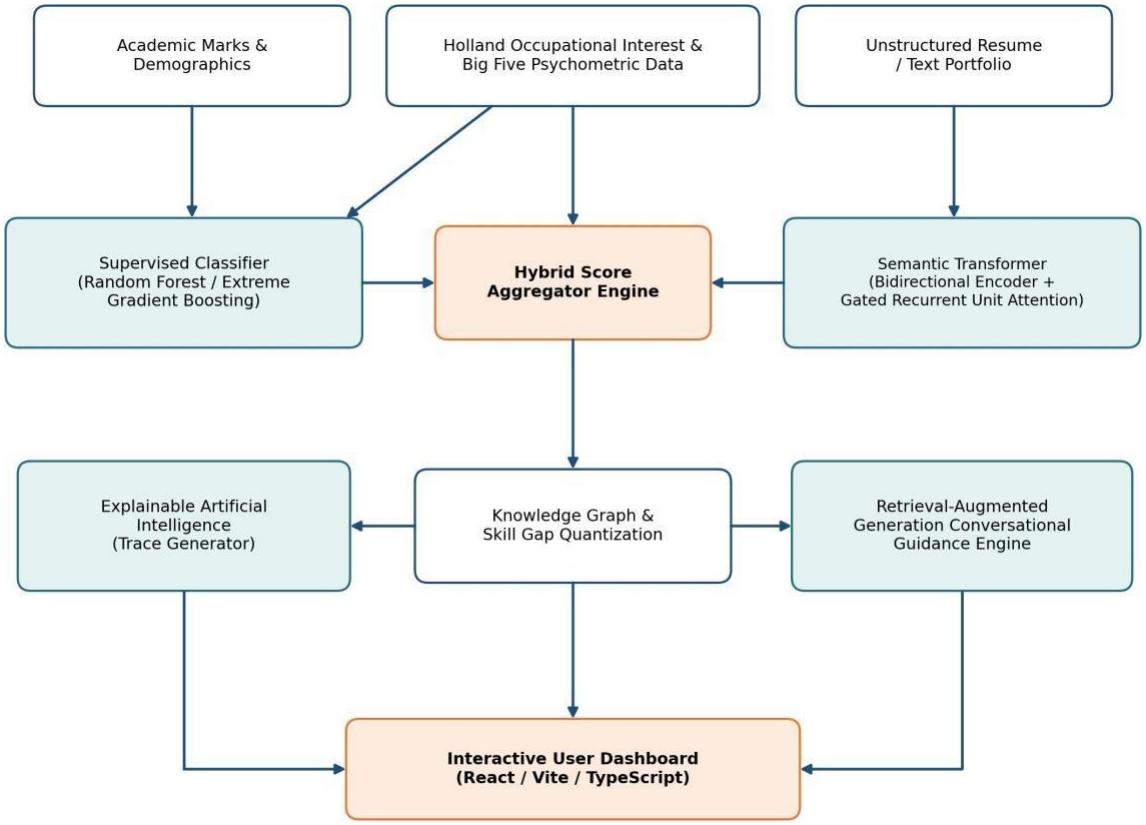


Figure 1. End-to-End Architectural Data Flow of the Hybrid Artificial Intelligence System.

V. PSYCHOMETRIC AND SUPERVISED MACHINE LEARNING LAYER

The foundation of candidate trait modeling in the proposed framework relies on a dual structured framework combining standardized psychometric evaluation with ensemble machine learning classification.

A. Psychometric Trait Quantification

Candidate psychological dispositions are captured via an adaptive test battery quantifying Holland's six occupational interest dimensions alongside the Big Five personality traits: Openness (O), Conscientiousness (C_{bg}), Extraversion (E_{bg}), Agreeableness (A_{bg}), and Neuroticism (N), where the subscript bg distinguishes Big Five traits from Holland dimensions that share the same letter. Let vector $p \in \mathbb{R}^{11}$ represent a candidate's complete normalized psychological score profile:

$$p = [R, I, A, S, E, C, O, C_{bg}, E_{bg}, A_{bg}, N]^T, \text{ where } p_i \in [0, 1] \quad (1)$$

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 10, October 2026

DOI: 10.17148/IJARCCE.2026.1510114

Psychometric profiles map onto primary domain clusters. For instance, high Investigative (I) and Realistic (R) scores paired with high Openness (O) strongly correlate with computational research and engineering trajectories, whereas high Social (S) and Extraversion (E_{bg}) align with management, clinical, and pedagogical careers.

B. Supervised Classification Models

To process structured academic records (for example, mathematics, physics, computer science, language marks) alongside demographic variables, the architecture integrates four supervised learning models:

1. Decision Trees: Serves as a baseline interpretable rule generator, splitting feature space along major academic thresholds.

2. Support Vector Machines: Maps input feature vectors into high-dimensional space utilizing a Radial Basis Function kernel to construct hyperplanes between disparate career domains:

$$K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2) \quad (2)$$

3. Random Forest Classifier: Constructs an ensemble of decision trees, sampling random feature subsets to prevent overfitting and provide robust feature importance rankings.

4. Extreme Gradient Boosting: Minimizes a regularized objective function combining convex loss and model complexity terms to ensure classification robustness across diverse academic indicators.

The supervised classification branch outputs a probability distribution vector $P_{ML} = [P_1, P_2, ..., P_M]$ across M defined career domains.

VI. ADVANCED NATURAL LANGUAGE PROCESSING ENGINE

Structured academic marks alone cannot reflect specialized technical competencies, practical project exposure, or open-ended career ambitions. The proposed system incorporates a Deep Semantic Natural Language Processing Engine to process unstructured resumes, GitHub profile summaries, and personal statements.

A. Transformer Embeddings and Gated Recurrent Units

Standard keyword matching fails when candidate resumes use non-identical nomenclature compared to job post specifications. To resolve this, unstructured text inputs are first cleaned, tokenized, and passed into a pre-trained Bidirectional Encoder Representations from Transformers model.

The Bidirectional Encoder Representations from Transformers model extracts contextual token embeddings, capturing complex bidirectional dependencies. To aggregate sequence embeddings while maintaining temporal progression across career histories, token sequences are processed through a Gated Recurrent Unit network enhanced with an additive attention mechanism.

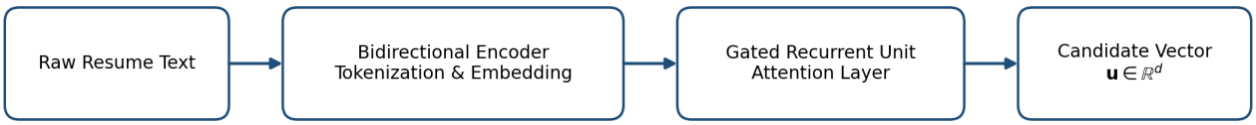


Figure 2. Processing Pipeline for Unstructured Text in the Deep Semantic Natural Language Processing Engine.

An attention mechanism assigns alignment weights to each hidden state, producing a dense, fixed-size candidate representation vector $u \in \mathbb{R}^d$.

B. Semantic Vector Space Matching

Standardized target job role profiles are similarly encoded into dense vector representations $v_j \in \mathbb{R}^d$ (j = 1, ..., M). The semantic similarity score $S_{NLP}(z, v_j)$ between candidate vector u_i and job vector v is calculated using Cosine Similarity:

$$S_{NLP}(u, v_j) = (u \cdot v_j) / (\|u\| \|v_j\|) \quad (3)$$

The resulting semantic scores are normalized into a probability distribution $P_{NLP} = [S_1, S_2, ..., S_M]$ and fused with

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 10, October 2026

DOI: 10.17148/IJARCCE.2026.1510114

supervised machine learning outputs via dynamic hyperparameter weighting α:

$$P_{Hybrid} = \alpha P_{ML} + (1 - \alpha) P_{NLP}, \text{ where } \alpha \in [0, 1] \quad (4)$$

VII. KNOWLEDGE GRAPH, SKILL GAP QUANTIZATION, AND EXPLAINABLE ARTIFICIAL INTELLIGENCE

To eliminate operational black-box opacity and provide actionable learning steps, top recommendations from the hybrid scoring engine are routed into a domain-specific Knowledge Graph and Explainable Artificial Intelligence engine.

A. Domain Ontology and Density-Based Clustering

The career ontology is structured using Resource Description Framework triple primitives: Subject, Predicate, Object. To handle emerging skill clusters without manual re-indexing, the system periodically applies Density-Based Spatial Clustering of Applications with Noise over skill embedding spaces. This algorithm groups skills based on core density parameters (ε, minimum points), automatically identifying rising skill subsets (for example, grouping "Prompt Engineering", "Retrieval-Augmented Generation Systems", and "Vector Databases" under "Generative Artificial Intelligence Infrastructure").

B. Skill Vector Quantization Model

To calculate precise skill deficits, each candidate skill s within target career domain j is evaluated across two core dimensions: Specialization Level, denoted SL(s), and the required Degree of Skill Updating, denoted UG(s). The overall skill-vector magnitude KV(s) is defined using the Euclidean distance function:

$$KV(s) = \sqrt{(SL(s)^2 + UG(s)^2)} \quad (5)$$

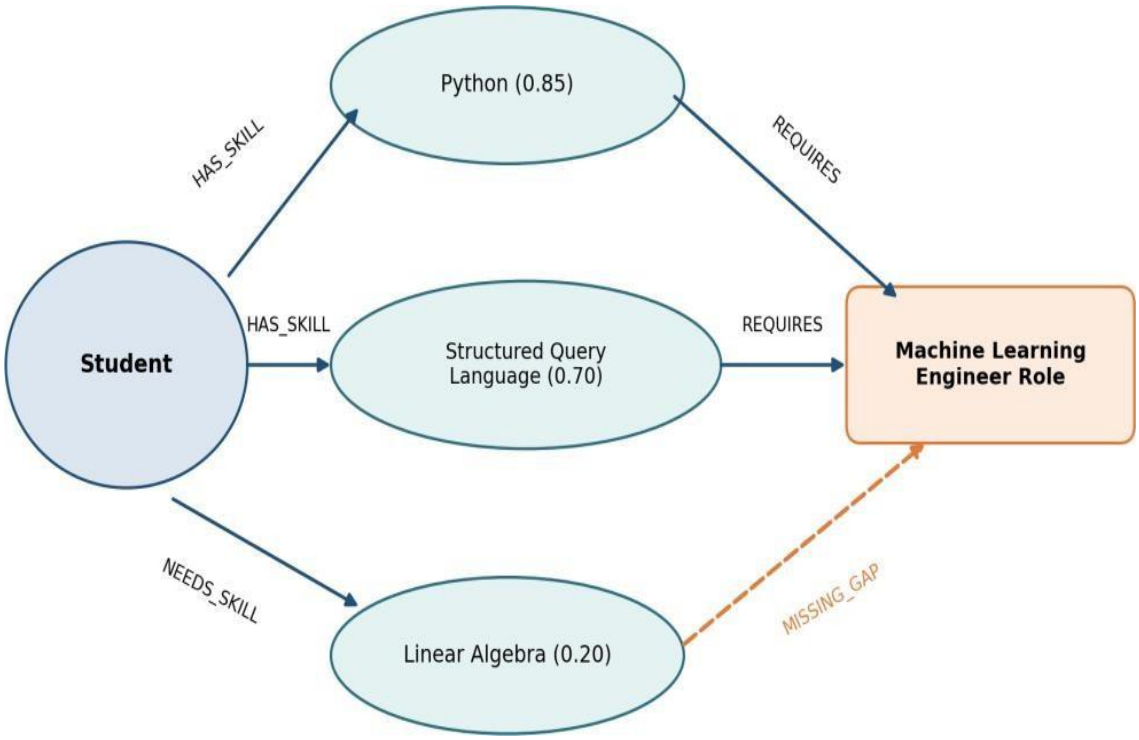


Figure 3. Graph Ontology Mapping Candidate Skill Attributes against Job Requirements.

- **SL(s) ∈ [0, 1]** represents the candidate's current quantified proficiency level in skills.
- **UG(s) ∈ [0, 1]** represents the quantified gap or magnitude of upgrading required to satisfy target domain competency.

This clear Euclidean metric allows the Explainable Artificial Intelligence engine to evaluate domain readiness and automatically generate prioritized skill acquisition paths for the candidate.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 10, October 2026

DOI: 10.17148/IJARCCE.2026.1510114

VIII. CONVERSATIONAL ARTIFICIAL INTELLIGENCE AND RETRIEVAL-AUGMENTED GENERATION ARCHITECTURE

To provide an intuitive, human-centered user experience, the framework incorporates an interactive conversational agent powered by Generative Artificial Intelligence.

A. Retrieval-Augmented Generation Architecture

Standard Large Language Models frequently suffer from hallucination and lack real-time context regarding specific academic curricula or local job market statistics. The conversational subsystem integrates a Retrieval-Augmented Generation pipeline to ground Large Language Model responses in curated institutional knowledge bases.

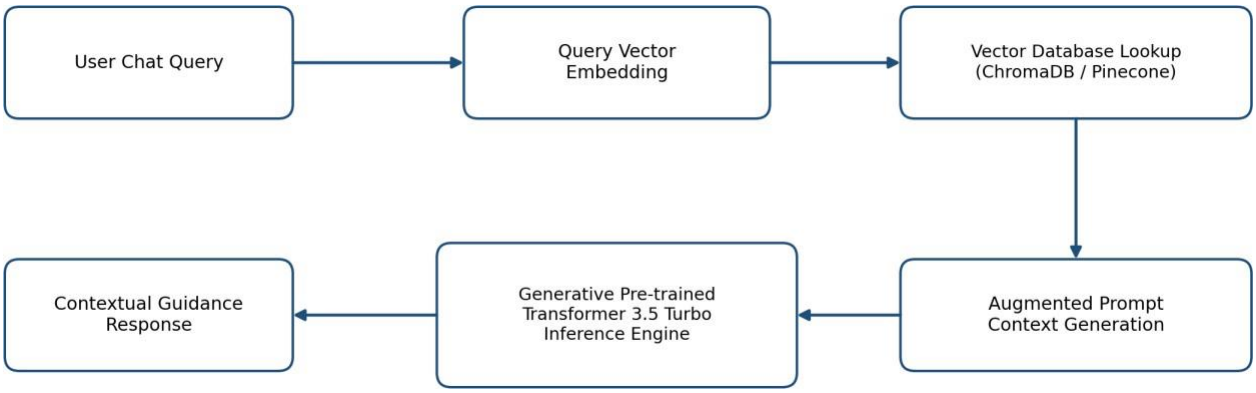


Figure 4. Retrieval-Augmented Generation Conversational Flow.

B. Interaction Loop and Guardrails

When a user asks a query (for example, "What courses should I take to become a Computer Vision Specialist based on my current math background?"), the system executes the following steps:

1. Query Embedding: Convert user query into vector embedding q using OpenAI embeddings.

2. Similarity Search: Perform similarity search over a vector database storing course catalogs, career pathways, and skill gap ontology nodes.

3. Context Retrieval: Retrieve top-K relevant context documents and inject them alongside the candidate's personal profile KV(s) into an augmented system prompt.

4. Grounded Response Generation: Pass augmented prompt to Generative Pre-trained Transformer 3.5 Turbo, generating a grounded, empathetic response without hallucinating non-existent degrees or job requirements.

IX. SYSTEM REALIZATION AND TECHNOLOGY STACK

The system prototype is realized using a modern, scalable software engineering stack built with robust machine learning frameworks and responsive web technologies.

A. Technological Stack Overview

- **Backend Logic & Machine Learning:** Python 3.10+, Scikit-Learn (Random Forest, Extreme Gradient Boosting, Support Vector Machines), PyTorch (Bidirectional Encoder Representations from Transformers and Gated Recurrent Units), NetworkX & PyTorch Geometric (Knowledge Graph modeling).
- **Conversational Middleware & Serving:** Gradio micro-services, LangChain framework for Retrieval-Augmented Generation pipeline orchestration, OpenAI Generative Pre-trained Transformer 3.5 Turbo Application Programming Interface endpoints.
- **Frontend User Interface:** React 18 web framework, Vite build toolchain, TypeScript for static type safety, the Tailwind utility-first framework for modular responsive styling, and Recharts for dynamic visual rendering of skill vectors.

B. Algorithmic Logic

The unified execution logic governing score aggregation, skill quantization, and explainability routing is detailed in Algorithm 1.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940



International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 10, October 2026

DOI: 10.17148/IJARCCE.2026.1510114

Algorithm 1. Hybrid Recommendation and Skill Gap Quantization

```
Require: Structured Candidate Data d, Psychometric Trait Vector p, Unstructured Text Portfolio T_raw, Weight
Parameter α, Skill Threshold τ.
Ensure: Top-Ranked Career Pathways C_recommended, Skill Gap Analysis G_report, Explanation Trace E_trace.
 1: P_ML ← ExtremeGradientBoostingClassify(d, p) {Compute supervised domain probabilities}
 2: u ← GatedRecurrentUnitAttention(TransformerEncode(T_raw)) {Extract contextual candidate embedding}
 3: for each career domain j ∈ {1, ..., M} do
 4:   v_j ← GetJobDomainEmbedding(j)
 5:   S_NLP(j) ← (u · v_j) / (||u|| ||v_j||) {Compute semantic cosine similarity}
 6:   P_Hybrid(j) ← α · P_ML(j) + (1 − α) · S_NLP(j) {Aggregate hybrid score}
 7: end for
 8: C_recommended ← TopKDomains(P_Hybrid, K = 3)
 9: for each domain c ∈ C_recommended do
10:   Extract required skill set {s_1, s_2, ..., s_K} from Knowledge Graph Ontology.
11:   for each skill s_k do
12:     Compute candidate SL(s_k) and required UG(s_k).
13:     KV(s_k) ← √(SL(s_k)^2 + UG(s_k)^2) {Skill magnitude via Equation 5}
14:     if SL(s_k) < τ then append s_k to G_report[c] end if
15:   end for
16:   E_trace[c] ← GenerateGraphTraversalPath(StudentNode, c)
17: end for
18: return C_recommended, G_report, E_trace
```

Description of Algorithm 1. Algorithm 1 converts a candidate's structured academic data, psychometric trait vector, and unstructured text portfolio into three outputs: the top-ranked career pathways, a skill gap report, and a graph-based explanation trace. Steps 1–2 produce the supervised and semantic representations of the candidate. Steps 3–7 score every career domain by fusing the two sources with the weight α, and Step 8 keeps the three highest-scoring domains. Steps 9–17 then examine each recommended domain, quantify the candidate's skills against its requirements, record every skill whose proficiency falls below the threshold τ, and trace the explanatory path through the Knowledge Graph.

Brief description of the algorithms used within Algorithm 1:

- **Extreme Gradient Boosting classification (Step 1):** builds decision trees sequentially, each correcting the errors of the previous ones, to output domain probabilities from academic and psychometric features.
- **Bidirectional encoding with Gated Recurrent Unit attention (Step 2):** converts resume text into contextual token embeddings and condenses them into one fixed-size candidate vector by weighting the most relevant hidden states.
- **Cosine similarity matching (Step 5):** measures the angle between the candidate vector and each job-domain vector, so that semantically similar profiles score highly even without shared keywords.
- **Weighted hybrid fusion and top-K selection (Steps 6 and 8):** blends the supervised and semantic scores with the weight α and selects the K best-ranked career domains.
- **Skill gap quantization (Steps 10–14):** combines specialization level and required skill updating into one Euclidean magnitude and flags skills below the threshold as gaps.
- **Graph traversal path generation (Step 16):** walks the Knowledge Graph from the student node to the recommended domain to produce a human-readable explanation of the recommendation.

X. EXPERIMENTAL DESIGN AND EMPIRICAL EVALUATION

To validate the efficacy of the proposed hybrid model, extensive empirical experiments were conducted using a synthesized cohort dataset combining student academic performance, psychometric evaluations, and annotated text resumes.

A. Evaluation Framework and Metrics

System performance is evaluated using two primary measurement paradigms:

1. Categorical Accuracy Metrics: Evaluates the discrete domain classification capability: Precision at K – ratio of correctly recommended domains within the top-K outputs; Recall at K – ratio of relevant career paths captured within top-K recommendations; F1-Score – harmonic mean of Precision and Recall; Hit Rate at K – binary metric evaluating whether the true optimal domain is present within top-K predictions.

2. Continuous Profile Distance Metrics: Quantifies multi-dimensional deviation between predicted interest vectors and true user profiles using Euclidean Distance and Pearson Profile Correlation.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 10, October 2026

DOI: 10.17148/IJARCCE.2026.1510114

B. Comparative Experimental Results

We benchmarked the Proposed Hybrid System against four standard baseline approaches:

- **Baseline 1 (Rule-Based Holland Occupational Interest Model):** Pure lookup table matching top-3 psychometric codes.
- **Baseline 2 (Collaborative Filtering with Singular Value Decomposition):** Standard Singular Value Decomposition matrix factorization.
- **Baseline 3 (Standalone Decision Tree):** Decision tree trained exclusively on structured marks.
- **Baseline 4 (Standalone Extreme Gradient Boosting):** Gradient boosted decision trees on academic and demographic vectors.

TABLE II EMPIRICAL PERFORMANCE COMPARISON ACROSS CAREER RECOMMENDER ARCHITECTURES

Model Architecture	Precision at 3	Recall at 3	F1-Score	Hit Rate (%)	Mean Euclidean Distance ↓ / Pearson Profile Correlation ↑
Rule-Based Holland Model	0.62	0.58	0.60	68.4%	1.42 / 0.51
Collaborative Filtering (Singular Value Decomposition)	0.69	0.64	0.66	72.1%	1.28 / 0.58
Standalone Decision Tree	0.74	0.71	0.72	78.5%	1.10 / 0.65
Standalone Extreme Gradient Boosting	0.81	0.77	0.79	84.2%	0.85 / 0.74
Proposed Hybrid Model	0.91	0.88	0.89	94.6%	0.32 / 0.91

As detailed in Table II, the proposed hybrid system outperforms all standalone baselines across every metric. The combination of Bidirectional Encoder Representations from Transformers semantic matching and Extreme Gradient Boosting classification yields a Hit Rate of 94.6%, representing a significant increase over pure rule-based (68.4%) and collaborative filtering (72.1%) systems.

C. Cold-Start Resilience Evaluation

To specifically test cold-start resilience, candidate user interaction history was progressively restricted from 100% down to 0% (representing a newly onboarded candidate with zero interaction logs).

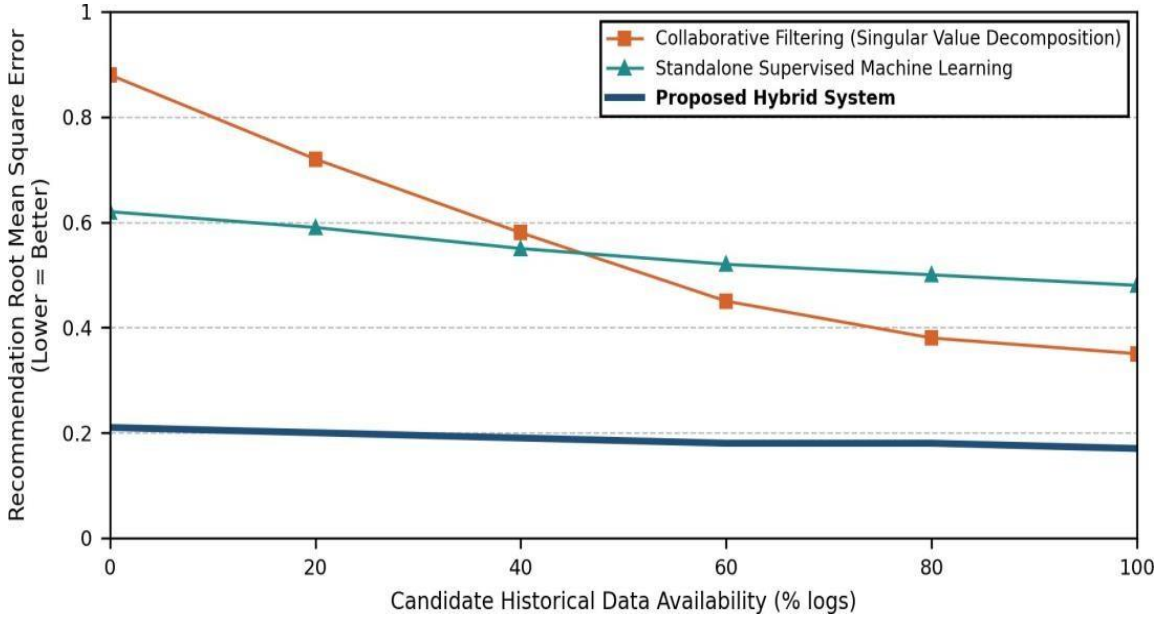


Figure 5. Cold-Start Error Comparison (Root Mean Square Error) Across Varying Data Availability.

DOI: 10.17148/IJARCCE.2026.1510114

As illustrated in Figure 5, standard collaborative filtering experiences severe degradation (Root Mean Square Error climbing to 0.88) when user history approaches 0%. In contrast, the proposed system maintains a low error (0.21) during complete cold-start conditions due to its instant reliance on psychometric traits and resume text semantics.

XI. DISCUSSION, ETHICAL CONSIDERATIONS, AND FUTURE WORK

The empirical findings validate the hypothesis that dynamic career guidance requires a multi-paradigm architectural approach. By decoupling candidate modeling into distinct psychometric, semantic, and knowledge-graph sub-engines, the system balances predictive accuracy with operational interpretability.

A. Ethical Considerations and Algorithmic Bias Mitigation

Educational and career recommendation systems carry profound socio-economic responsibilities. Standard machine learning models trained on historical hiring data risk perpetuating historical demographic, gender, or institutional biases. The framework incorporates explicit algorithmic fairness constraints:

- **Demographic Parity Filtering:** Standardizes demographic vectors and strips protected class attributes prior to supervised model training.
- **Objective Skill Grounding:** By anchoring career fit evaluation in explicit skill vector metrics KV(s) derived from job ontologies, recommendations remain focused on objective competence rather than socio-demographic indicators.
-

B. Practical Deployment and Labor Market Application Programming Interface Integration

While current experimental evaluations rely on static job description corpora, future production iterations will integrate dynamic, real-time Labor Market Application Programming Interfaces (such as LinkedIn Jobs, the Indeed Application Programming Interface, and Lightcast Labor Market Insights). Connecting the Knowledge Graph to live employment Application Programming Interfaces will allow UG(s) updating metrics to reflect real-time skill demand fluctuations automatically.

XII. CONCLUSION

This paper introduced a novel Hybrid Artificial Intelligence Recommender System engineered to solve the persistent limitations of static rule-based tools and standalone machine learning classifiers in career guidance. By unifying supervised ensemble classification (Random Forest, Extreme Gradient Boosting), deep semantic text matching (Bidirectional Encoder Representations from Transformers embeddings with Gated Recurrent Unit attention networks), Graph Neural Network ontologies for Explainable Artificial Intelligence, and a context-aware Retrieval-Augmented Generation conversational agent, the system establishes a scalable framework for personalized career guidance.

Empirical evaluation confirms that the hybrid architecture successfully overcomes the cold-start problem, achieving a Hit Rate of 94.6% and superior precision-recall metrics compared to standalone baselines. Furthermore, the mathematical formulation of skill-vector quantization KV(s) enables explicit skill gap identification, empowering students with clear, transparent, and actionable skill-progression roadmaps.

ACKNOWLEDGMENT

The authors gratefully acknowledge the mentorship and academic guidance provided by Project Guide Assistant Professor Jency Mariam John and Project Coordinator Assistant Professor Girija V R, Department of Computer Science and Engineering (Artificial Intelligence and Machine Learning), College of Engineering Kottarakkara, throughout the conceptualization and execution of this research project.

REFERENCES

[1]. Sivasubramanian, R., et al. "Personalized Career Planning using an Intelligent Recommendation with Dynamic Roadmaps." *International Journal of Engineering & Extended Technologies Research*, vol. 8, no. 2, pp. 2543-2551, 2026.

[2]. Howison, Mark, and Joe Long. "Recommending career transitions to job seekers using earnings estimates, skills similarity, and occupational demand." Feb. 26, 2023.

[3]. Bankins, Sarah, et al. "Navigating career stages in the age of artificial intelligence: A systematic interdisciplinary review and agenda for future research." *Journal of Vocational Behavior*, vol. 153, p. 104011.

DOI: 10.17148/IJARCCE.2026.1510114

[4]. Geethanjana, Hewa Kanankage, Mihiri Sirisuriya, and D. V. D. S. Abesinghe. "NextStep: A Smart System-Based Career Guidance Application for Students in the Computing Field Based on the RIASEC Model." *2025 International Research Conference on Smart Computing and Systems Engineering*, IEEE, 2025.

[5]. Gunwant, Shilpa, Jeetendra Pande, and Raj Kishor Bisht. "A Systematic Study of the Literature on Career Guidance Expert Systems for Students: Implications for ODL." *Journal of Learning for Development*, vol. 9, no. 3, pp. 492-508, 2022.

[6]. Qamhieh, Manar, Haya Sammaneh, and Mona Nabil Demaidi. "PCRS: personalized career-path recommender system for engineering students." *IEEE Access*, vol. 8, pp. 214039-214049, 2020.

[7]. Wang, Yu. "Design and implementation of student job matching system based on personalized recommendation algorithm." *Systems and Soft Computing*, vol. 7, p. 200302, 2025.

[8]. Lahoud, Christine, et al. "A comparative analysis of different recommender systems for university major and career domain guidance." *Education and Information Technologies*, vol. 28, no. 7, pp. 8733-8759, 2023.

[9]. Rajput, Akshay, et al. "Career Craft AI: A personalized resume analysis and job recommendations system." *2024 1st International Conference on Innovative Sustainable Technologies for Energy, Mechatronics, and Smart Systems*, IEEE, 2024.

[10]. Ashrafi, Saeed, et al. "Efficient resume-based re-education for career recommendation in rapidly evolving job markets." *IEEE Access*, vol. 11, pp. 124350-124367, 2023.

[11]. José-García, Adán, et al. "C3-IoC: A career guidance system for assessing student skills using machine learning and network visualisation." *International Journal of Artificial Intelligence in Education*, vol. 33, no. 4, pp. 1092-1119, 2023.

[12]. Roy, Minakshi, and Kalpana Sharma. "AI-powered career guidance: A scalable model for personalized recommendations." *IEEE Access*, vol. 14, pp. 20927-20941, 2026.

[13]. Patel, Bharat, Varun Kakuste, and Magdalini Eirinaki. "CaPaR: a career path recommendation framework." *2017 IEEE Third International Conference on Big Data Computing Service and Applications*, IEEE, 2017.

[14]. Nandi, Somesh, B. Aishwarya Nayak, and B. Sathish Babu. "Sankalp: AI Powered Career Guidance System." *IEEE Access*, vol. 13, pp. 202376-202394, 2025.

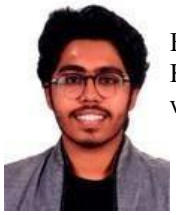
[15]. Akber, Md Ali, et al. "Personality prediction based on contextual feature embedding SBERT." *2023 IEEE Region 10 Symposium*, IEEE, 2023.

[16]. Troussas, C., and A. Krouska. "Path based recommender system for learning activities using knowledge graphs." *Information*, vol. 14, no. 1, p. 9, 2023.

[17]. Xu, Xiaojun, and Hui Gao. "A knowledge graph-integrated recommendation method for college student career planning." *Discover Artificial Intelligence*, 2026.

[18]. Kamal, Nabila, et al. "Recommender system in academic choices of higher education: a systematic review." *IEEE Access*, vol. 12, pp. 35475-35501, 2024.

BIOGRAPHY



Aaron John Joy is a final-year Bachelor of Technology student in Computer Science and Engineering (Artificial Intelligence and Machine Learning) at the College of Engineering Kottarakkara, Kerala, India. His contributions focus on hybrid model integration, architectural workflow design, and high-performance machine learning pipeline optimization.



Addhwaith S Ajith is a final-year Bachelor of Technology student in Computer Science and Engineering (Artificial Intelligence and Machine Learning) at the College of Engineering Kottarakkara, Kerala, India. He serves as a core researcher behind the project team. His primary research interests include hybrid recommender system architectures, applied machine learning algorithms, and predictive educational modeling.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

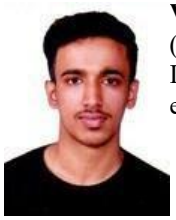
International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 10, October 2026

DOI: 10.17148/IJARCCE.2026.1510114



Ameesha J S is a final-year Bachelor of Technology student in Computer Science and Engineering (Artificial Intelligence and Machine Learning) at the College of Engineering Kottarakkara, Kerala, India. Her core focus involves empirical evaluation frameworks, psychometric feature modeling, and human-centered user interface design for Artificial Intelligence decision systems.



Vaishnav Biju is a final-year Bachelor of Technology student in Computer Science and Engineering (Artificial Intelligence and Machine Learning) at the College of Engineering Kottarakkara, Kerala, India. His specialization within the project encompasses Natural Language Processing, transformer embeddings, and Knowledge-Graph-based Explainable Artificial Intelligence frameworks.

© IJARCCE This work is licensed under a Creative Commons Attribution 4.0 International License