



RAG-Enhanced Career Counseling System

Retrieval-Augmented Generation for Personalized Career Guidance

Team No: **PROJECT_79**

Project Mentor : Dr. Soumi Paul

Zubair Khan 1231

Branch : **CSE**

Saib Idris 1248

Institution : **Techno India**

Anshu Kumar - 1251

University, Kolkata

Aman Mishra - 1269

2025-26_Final YR Project_79

The Crisis: Shortage of Career Counselors

Global Shortage Crisis

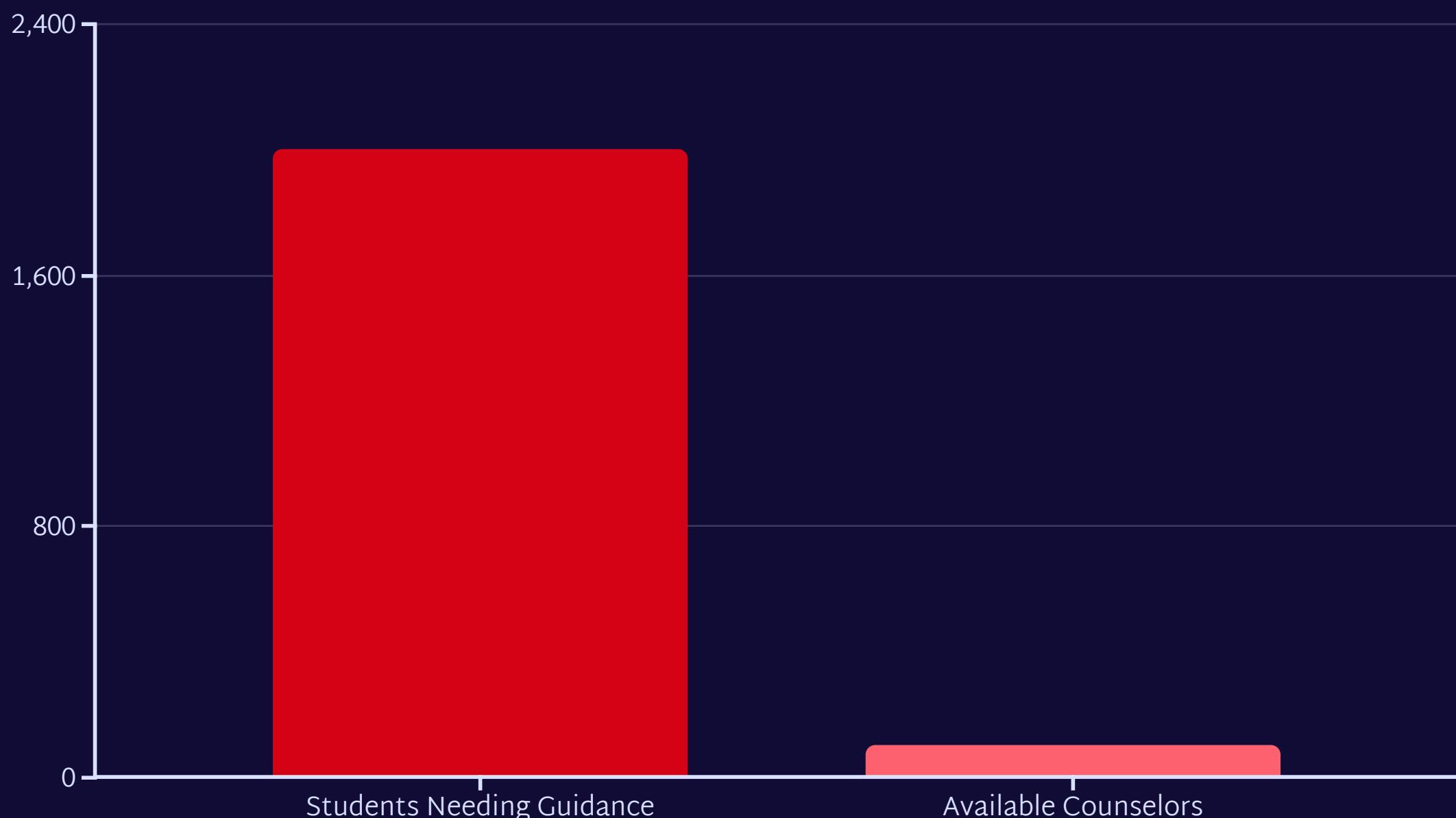
- Career counselors are increasingly scarce in educational institutions
- Developing nations face 1:2,000+ counselor-to-student ratio
- 70% of students report confusion about career decisions

Information Asymmetry

- Students lack access to real-time job market data
- Outdated career resources and static information
- Generic advice that doesn't match individual skills
- Inequitable access between urban vs rural students

Why RAG is the Solution

- RAG enables LLMs to access domain-specific and updated information not available in training data
- Democratizes access to quality career guidance 24/7
- Provides personalized recommendations grounded in factual data



India's [student-to-counselor](#) ratio is about 1:3,000—far higher than the globally recommended ratio of 1:250. Such a gap highlights the **career counsellor shortage India**. Research indicates that over 93% of students know ten careers or less from more than 250 possible options.

Research Foundation: RAG in Career Guidance



RAG for Career Guidance (2025)

- ✓ NHSJS study shows RAG outperforms baseline LLM on empathy and relevance
- ✓ RAG improves LLM responses by retrieving relevant data from knowledge bases
- ☒ GAP: Limited to high school context only



Real-time Data Integration

- ✓ JMIR AI study: 92% answer relevancy with 9,500+ users using RAG
- ✓ Hybrid search combining keyword, semantic, and metadata filtering
- ☒ GAP: No career-specific job market API integration



Fairness in AI Systems

- ✓ Nature study provides 27 fairness metrics framework
- ✓ Science Direct: fairness considerations in recruitment
- ☒ GAP: No fairness-first career counseling system



Explainable AI (XAI)

- ✓ SHAP framework provides transparent explanations
- ☒ GAP: Never applied to career recommendations with RAG

OUR CONTRIBUTION:

- First RAG system combining O*NET/ESCO taxonomies with real-time job APIs
- Hybrid retrieval with knowledge graphs + vector databases
- SHAP-based explainability for transparent recommendations
- Fairness-first architecture ensuring demographic parity

Problem Statement: The RAG Solution Need

Current Reality (Global Scale)

- Career Counseling Market: \$5B+ growing 12% annually
- Counselor Shortage: 1:2,000+ ratio in developing nations
- Student Confusion: 70% report decision paralysis
- AI in EdTech: \$10B+ growing 20% annually
- Result: Students make career decisions blindly

Current Manual Process Issues

[Static Resources] → [Generic Advice] → [No Follow-up] → [Skill Mismatches]

(Outdated) (One-size-fits-all) (Unemployment)



Why RAG Addresses These Problems

Real-time Updates

<6 hour data freshness

Cost-effective

No retraining needed

Factually Grounded

Citations from trusted sources

Infinitely Scalable

Serves thousands simultaneously

24/7 Available

Always-on guidance

Technical Challenges Addressed by RAG

Challenge 1: Data Currency

Problem: LLM training data is static with knowledge cut-off dates

Current: Career advice based on outdated market conditions

RAG Solution: Real-time job market data updates via APIs

Target: <6 hour data freshness from LinkedIn, Glassdoor

Challenge 2: Factual Accuracy

Problem: LLMs may generate answers without factual basis (hallucinations)

Current: Students receive unreliable career guidance

RAG Solution: RAG provides LLMs with specific knowledge and steers away from making up answers

Target: >90% answer relevancy with source citations

Challenge 3: Domain Specificity

Problem: General LLMs lack career counseling domain expertise

Current: Generic advice not tailored to career progression

RAG Solution: ESCO taxonomy integration (960 occupations, 13,485 skills)

Target: Domain-optimized career path recommendations

Challenge 4: Transparency & Trust

Problem: Students don't trust black-box recommendations

Current: No explanation for career suggestions

RAG Solution: SHAP-based explainability with source citations

Target: Transparent reasoning for each recommendation

Challenge 5: Bias & Fairness

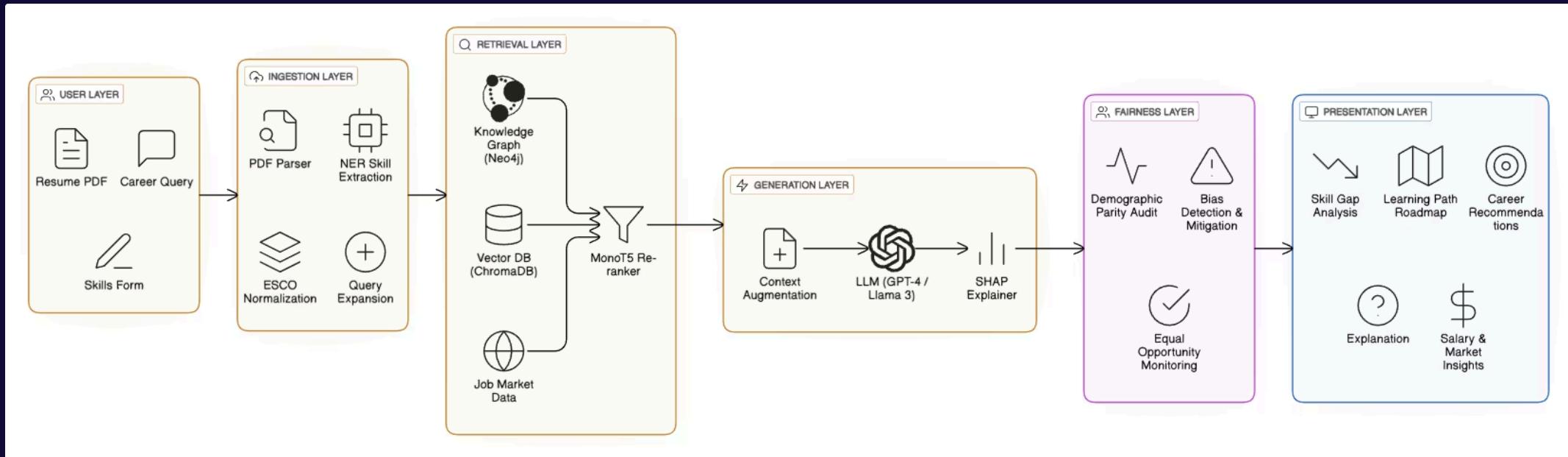
Problem: AI systems can perpetuate demographic biases

Current: Unfair recommendations affecting underrepresented groups

RAG Solution: Fairness-first architecture with bias monitoring

Target: ±3pp demographic parity across gender/ethnicity

RAG-Enhanced System Architecture



Our RAG architecture addresses career counseling challenges by combining LLM generative capabilities with external knowledge retrieval. The system integrates three critical data sources: standardized occupational taxonomies, real-time job market APIs, and educational resources, creating a comprehensive knowledge base for factually grounded career recommendations.

RAG vs Traditional Approaches

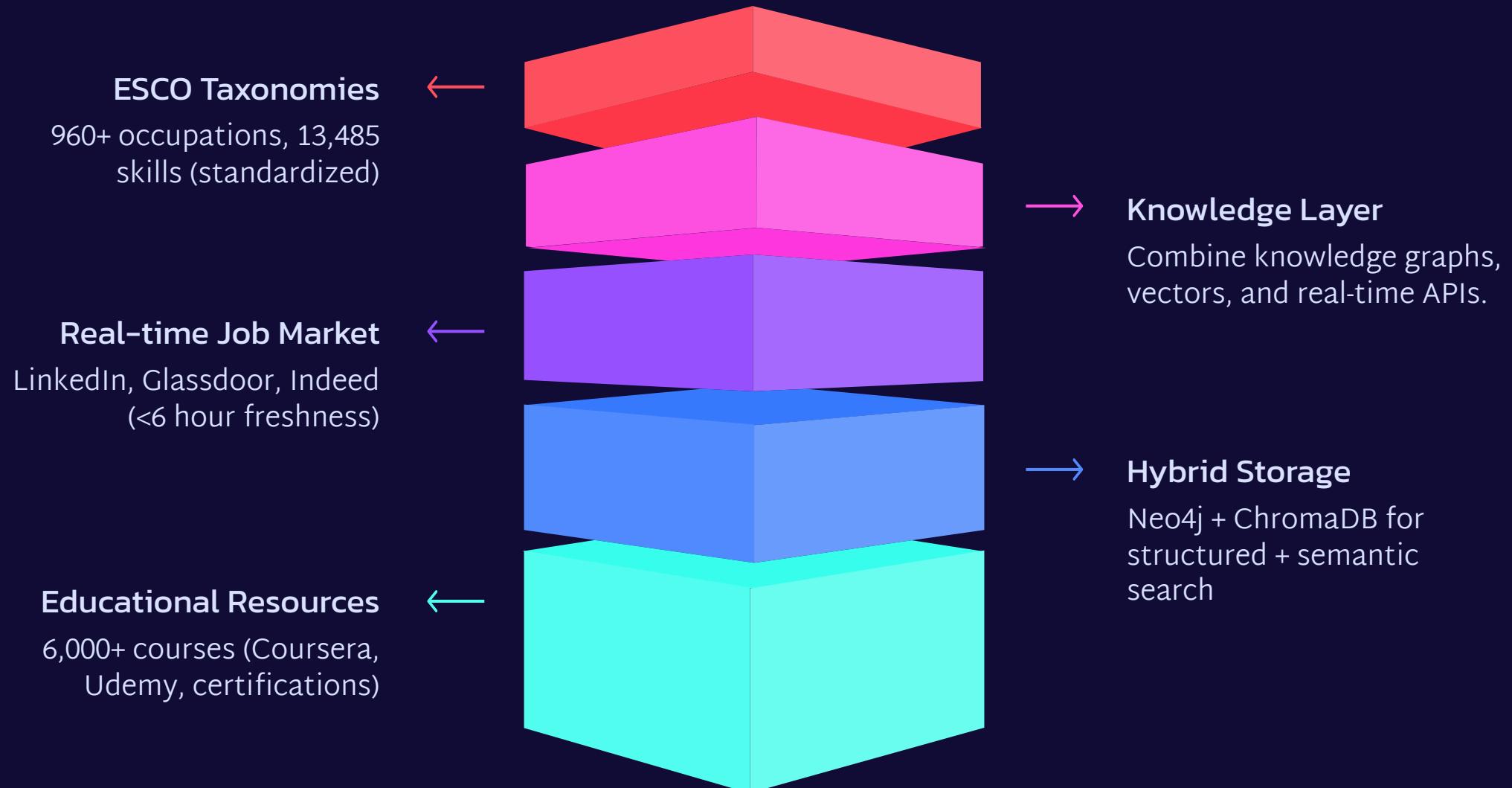
Feature	Traditional Systems	Our RAG System
Data Freshness	✗ Static/Outdated	✓ <6 hours via APIs
Knowledge Source	✗ Training data only	✓ Multi-source RAG
Factual Grounding	✗ Hallucination prone	✓ Source citations
Personalization	✗ Generic advice	✓ Profile-based RAG
Explanations	✗ Black box	✓ SHAP + Citations
Fairness Audit	✗ No monitoring	✓ Demographic parity

What Makes Our RAG Solution Revolutionary

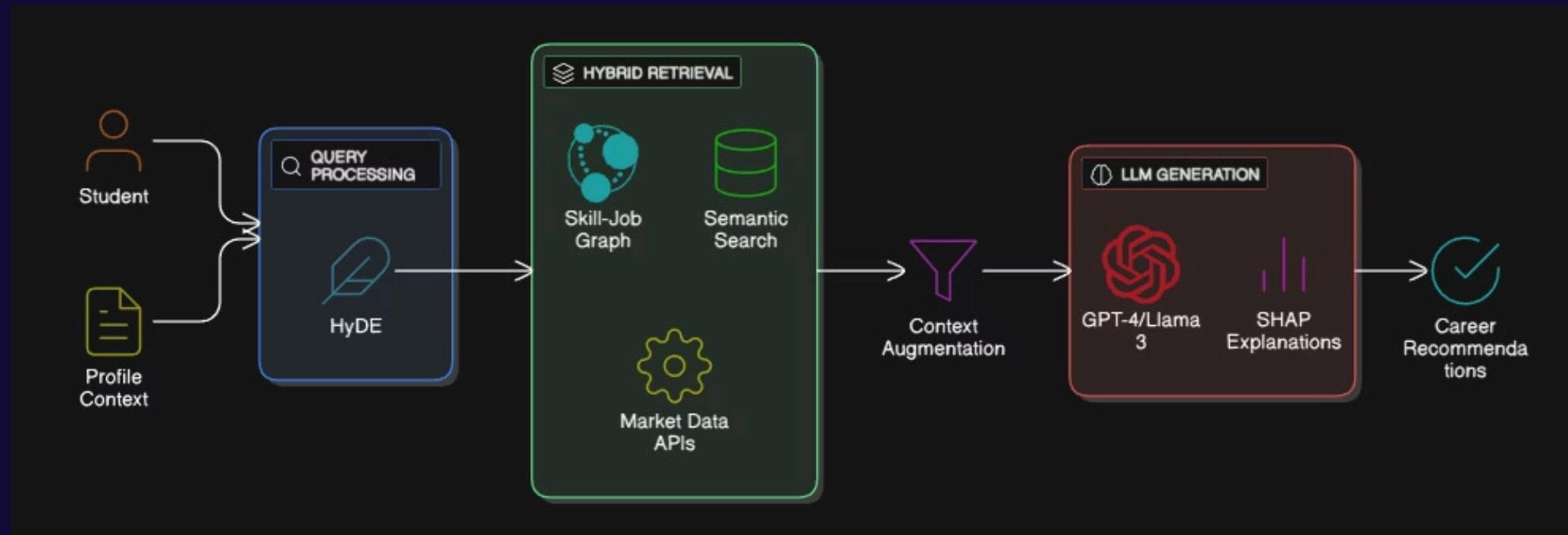


Component 1: Multi-Source Knowledge Base

RAG Data Sources Architecture



Component 2: RAG Generation Pipeline



Real-World Scenario: Data Analyst (3 yrs, SQL/Python) → Wants Data Science role in 6 months, \$2K budget

01

Query Processing

Expand user intent with HyDE Engine for semantic variants.

02

Hybrid Retrieval

Gather data from multiple powerful knowledge sources.

03

Context Augmentation

Intelligently re-rank retrieved documents by relevance (MonoT5).

04

LLM Generation

Synthesize concrete, personalized action plan (fine-tuned GPT-4).

Step 2 Details: Hybrid Retrieval - Multi-Source Knowledge

Neo4j Knowledge Graph

- Skill-job relationships
- Analyst→DS success: **82%**
- Salary growth: **+8.3%**

ChromaDB Vector Database

- Semantic course matching
- ML Specialization: **0.94** match
- Duration: **3 months**

Job Market APIs

- Real-time opportunities
- **2,340** DS positions
- Median salary: **\$128K**

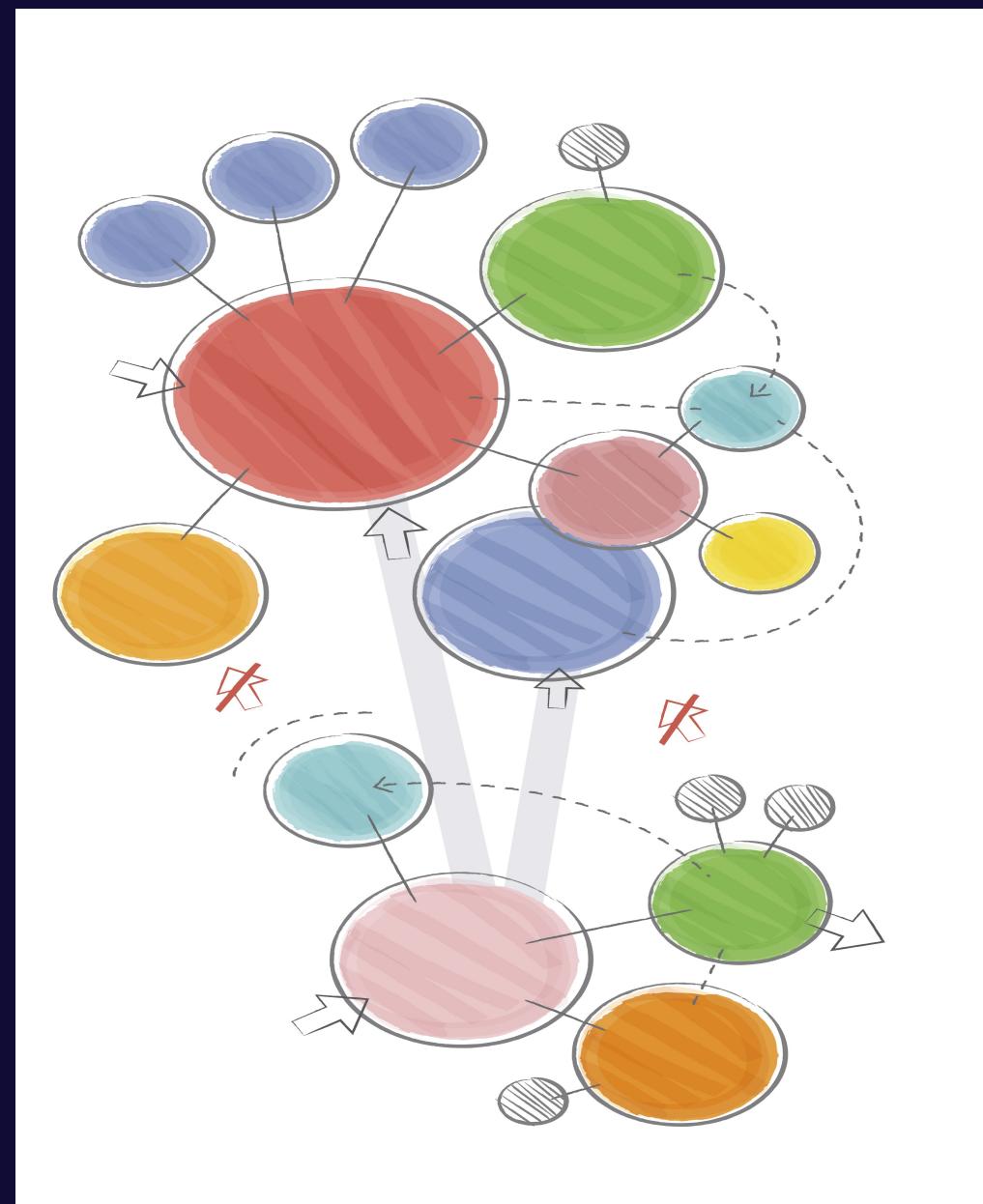
The system delivers grounded, current, and accurate recommendations—every time.

Component 3: Neo4j Knowledge Graph for Career Intelligence

Knowledge Graph Structure

Structured career intelligence database:

- **960 occupations** (ESCO taxonomy—EU standard)
- **13,485 skills** (standardized skill definitions)
- **50,000+ relationships** (job ↔ skill, course ↔ skill)



Graph Relationships

Nodes (Entities):

- **Occupations:** {Data Scientist, Product Manager, ML Engineer}
- **Skills:** {Python, Leadership, Statistics, Deep Learning}
- **Courses:** {Deep Learning Specialization, Leadership 101}
- **Progressions:** {SDE → Senior SDE → Engineering Manager}

RAG-Optimized Relationships:

- (Data Scientist) -REQUIRES→ (Python) [confidence: 0.95]
- (Python) -PREREQUISITE→ (Programming Fundamentals)
- (Deep Learning Course) -TEACHES→ (Neural Networks)
- (ML Engineer) -PROGRESSES_TO→ (Senior ML Engineer)

RAG Query Example: "I'm a Software Engineer, want to become Data Scientist"

1

Knowledge Graph Query

```
MATCH (ds:Occupation {name: "Data Scientist"})-[:REQUIRES]->(skill:Skill)
RETURN skill.name
```

Retrieved: [Python, Statistics, Machine Learning, Data Visualization]

2

Skill Gap Analysis

Current skills: [Python, SQL, Git]

Missing skills: [Statistics, Machine Learning, Data Visualization]

3

Learning Path Generation

1. Statistics for Data Science (8 weeks)
2. Machine Learning Fundamentals (12 weeks)
3. Data Visualization with Python (4 weeks)

4

RAG Context Integration

Knowledge graph results + vector search + job market APIs = comprehensive career roadmap

Thank You for Your Attention

We sincerely appreciate you taking the time to learn about our RAG-Enhanced Career Counseling System. Your engagement and interest are highly valued.

Github link: <https://github.com/Zubair576335/career-counselling-ai>