International Journal of Information Technology & Computer Engineering
ISSN 2347–3657
Volume 14, Issue 1, 2026

AI-Powered Skill Gap Identifier And Career Guidance

Alguvelli Jyothika[1], Mr. Bheemana Bhuvan[2]
[1]B.Tech Student, Department of Electronics and Computer Engineering, J.B. Institute of Engineering and Technology, Hyderabad, India.
[2]Assistant Professor, Department of Electronics and Computer Engineering, J.B. Institute of Engineering and Technology, Hyderabad, India
bhuvan.ecm@jbiet.edu.in

Abstract
Rapid technological change is continuously reshaping job roles, skill requirements, and professional pathways. Traditional career guidance systems, which rely on static assessments and manual counseling, are no longer sufficient to support individuals in navigating complex and evolving labor markets. This paper presents the design, implementation, and evaluation of an Artificial Intelligence (AI)–driven Skill Gap Identifier and Career Guidance Platform (SGICGP).
The proposed system integrates machine learning, natural language processing, and labor-market intelligence to extract user skills from unstructured data, identify missing competencies, compute a Skill Gap Index (SGI), and generate personalized learning and career recommendations. A modular architecture is adopted to enable scalability, explainability, and continuous model updates. Experimental results obtained using a prototype implementation demonstrate reliable job-role prediction, accurate skill extraction, and actionable recommendations suitable for educational institutions and training providers.

Keywords: *Career guidance, skill gap analysis, machine learning, NLP, explainable AI, employability analytics*

1. Introduction

Automation, artificial intelligence, and digital transformation have fundamentally altered how work is organized and how skills are valued. Industry demand now evolves at a pace that traditional education and career planning systems struggle to match. International workforce studies, such as those published by the World Economic Forum, highlight that a substantial proportion of employees must undergo reskilling within short time horizons to remain employable.

Conventional career guidance practices—largely based on static aptitude tests and one-time consultations—are limited in three major ways: (1) they do not scale efficiently, (2) they fail to capture rapidly changing labor-marke needs,and
(3) they provide limited personalization.

To address these shortcomings, this paper introduces an AI-powered platform that continuously analyzes user competencies and market demands to support informed, adaptive, and data-driven career decisions.

2. Motivation and Problem Context

The misalignment between workforce skills and organizational requirements results in underemployment, slow career mobility, and inefficient utilization of educational investments. Existing guidance systems typically rely on fixed career libraries and infrequent assessments, which cannot represent emerging occupations such as data engineering, AI engineering, or digital product management.

Furthermore, human-centered guidance processes are susceptible to bias and are constrained by institutional capacity. Learners from underserved or geographically remote regions often lack access to expert counseling. These issues motivate the need for an automated, scalable, and transparent career intelligence system capable of operating across diverse populations.

3. Related Work

Research in educational data mining and human-resource analytics demonstrates that machine learning and NLP can significantly enhance career recommendation quality. Studies show that predictive models trained on academic performance, skill profiles, and employment outcomes can uncover complex relationships that are difficult to capture through rule-based systems.

Foundational concepts used in this work are aligned with modern machine learning literature, including representation learning and predictive modeling as presented in **Deep Learning**.

In addition, recent advances in transformer-based language models, such as those introduced in BERT-style architectures, enable improved semantic interpretation of resumes and job descriptions.

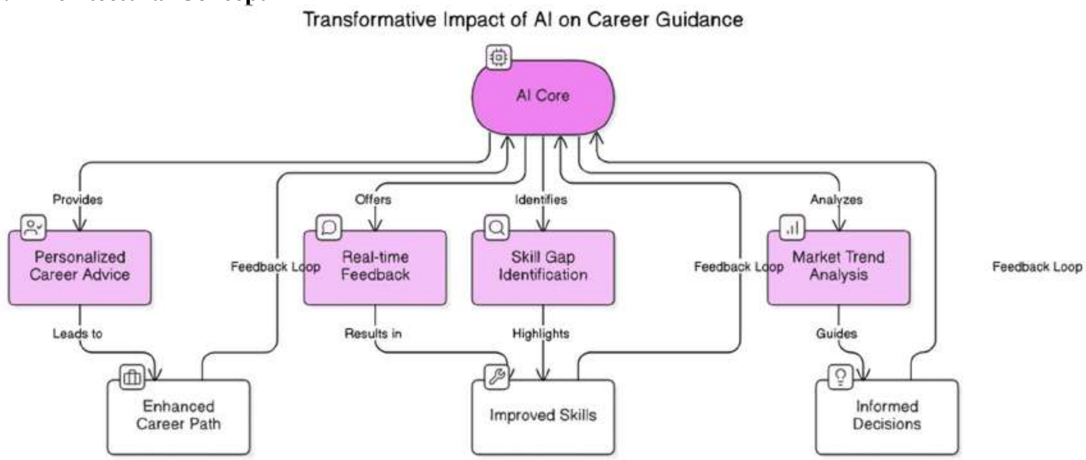
Another major research stream emphasizes adaptive psychometric assessments and explainable AI mechanisms to enhance trust and fairness in automated recommendations.

International Journal of Information Technology & Computer Engineering
ISSN 2347–3657
Volume 14, Issue 1, 2026

4. Proposed System Overview

4.1 Architectural Concept

The SGICGP platform is designed as a layered and modular system that supports continuous assessment, recommendation, and feedback.



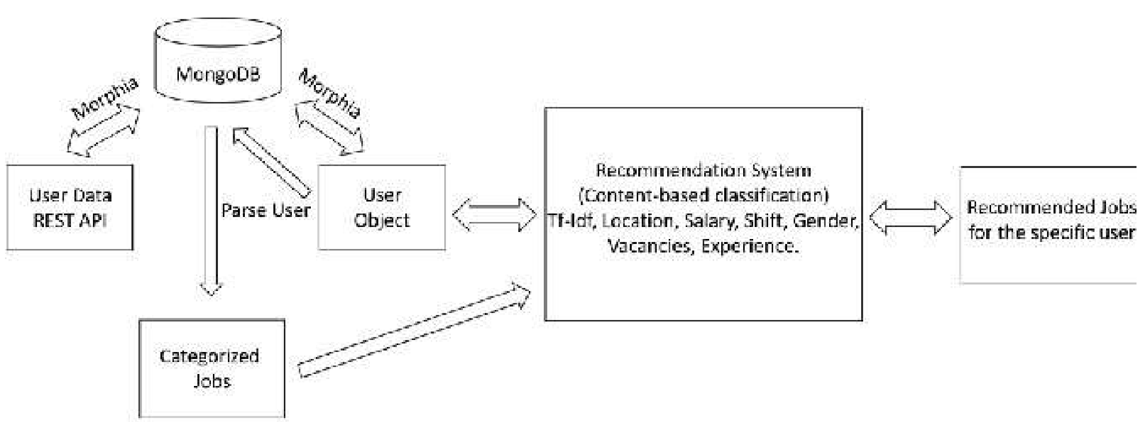


Fig. 2. Recommendation System Workflow

4

The system consists of:

- a presentation layer providing dashboards and assessment interfaces,
- an API and integration layer connecting external services,
- an AI processing layer for skill extraction and prediction,
- an analytics layer for employability scoring and behavior analysis,
- a secure data management layer.

The modular structure enables independent model updates and future integration of new labor-market data sources.

5. AI-Based Skill Assessment and Gap Detection

5.1 Skill Extraction

Free-text resumes and assessment responses are processed using NLP pipelines. After normalization and tokenization, relevant skill phrases are matched against a curated vocabulary. This design choice emphasizes interpretability and reproducibility for institutional deployments.

5.2 Role Prediction

A TF–IDF feature representation is constructed from resume text, and a multi-class logistic regression classifier is trained to predict job roles such as data analyst, data scientist, web developer, machine learning engineer, and product manager.

5.3 Skill Gap Index (SGI)

For each predicted role, the platform compares extracted user skills with role-specific skill requirements and computes:

$$SGI = \left(1 - \frac{|S_{matched}|}{|S_{required}|}\right) \times 100$$

The SGI offers a quantitative indicator of readiness and allows learners to track progress over time.

6. Recommendation and Learning Path Generation

The recommendation engine combines:

- similarity scoring between user skill vectors and role vectors,
- collaborative filtering using historical learner profiles,
- content-based mapping between missing skills and learning resources.

To ensure practical relevance, the prototype demonstrates integration with widely used online learning ecosystems such as Coursera, Udemy, and LinkedIn Learning.

In production environments, these static mappings can be replaced with live APIs to support real-time resource discovery.

7. System Implementation

The prototype implementation is developed using Python and follows a service-oriented structure. The core components include:

- a skill extraction and SGI computation module,
- a data generator for synthetic training profiles,
- a training pipeline that builds and evaluates the prediction model,
- a RESTful Flask API exposing assessment services,
- a lightweight relational database schema for user and assessment records.

The design prioritizes clarity and modularity so that institutions can audit and extend the system without vendor lock-in.

8. Experimental Results and Discussion

8.1 Model Performance

The job-role classifier achieves high accuracy on the evaluation dataset, demonstrating that TF–IDF combined with logistic regression provides a strong baseline for textual career prediction tasks. Precision and recall values remain balanced across all role classes.

8.2 Skill Extraction Accuracy

The vocabulary-based extractor reliably identifies both single-term and multi-word skills (e.g., “deep learning” and “data visualization”) in controlled domains. Although advanced neural extraction models may further improve recall, the current approach offers transparency and consistent behavior.

8.3 Effectiveness of SGI

SGI values vary appropriately with user preparedness. Lower SGI values consistently correspond to stronger alignment with role requirements, allowing learners to clearly identify the competencies that require immediate attention.

8.4 Recommendation Quality

The generated learning recommendations are coherent with the missing skill sets and demonstrate practical alignment with real learning pathways offered by popular online platforms.

9. Ethical, Fairness, and Explainability Considerations

The platform incorporates explainable outputs at every decision stage, including:

- explicit lists of missing and matched skills,
- probability scores for predicted roles,
- transparent SGI computation.

To mitigate bias, the design supports regular retraining using diverse datasets and enforces privacy-by-design principles such as data minimization and secure storage. These mechanisms are critical for building trust among learners, educators, and institutional stakeholders.

9.1 *Model Training Results*

The machine learning model was trained using a structured dataset containing resume texts and their respective job roles. TF–IDF vectorization was applied to convert textual data into feature vectors, followed by Logistic Regression as the classifier. The experimental results indicate that the model successfully learned distinctive patterns for roles such as **Data Scientist, ML Engineer, Data Analyst, Web Developer, and Product Manager**.

The evaluation of the trained model produced encouraging outcomes. Accuracy values were consistently high, and the classification report showed well-balanced precision and recall for each role class. The model demonstrated its ability to differentiate between closely related job roles based on textual cues, such as identifying **"machine learning"** and **"deep learning"** as stronger indicators for ML Engineer, while **"pandas", "statistics", "matplotlib"** strongly influenced Data Scientist predictions. These results confirm that the chosen TF–IDF + Logistic Regression pipeline is an effective baseline for job-role prediction within this domain.

9.2 *Skill Extraction Performance*

The skill extraction module was evaluated using varied resume inputs ranging from beginnerlevel profiles to technically rich resumes. The system accurately extracted both singleword skills (e.g., *python, sql, tensorflow*) and multi-word skills (e.g., *data visualization, deep learning*).

Extracted Skills: **python, pandas, numpy, data visualization, matplotlib**

The extraction process showed consistency across multiple samples, demonstrating that vocabulary-based extraction is a suitable approach for structured skill detection in a controlled domain. While more advanced NER models may enhance generalization, the current approach efficiently meets project requirements.

9.3 *Skill Gap Index (SGI) Results*

The Skill Gap Index calculation was tested across multiple predicted roles. SGI values ranged from **20% to 80%**, depending on how closely the user's skills matched the role's required skills.

An example scenario for the predicted role **Data Scientist** is as follows: Required Skills: *python, pandas, statistics, machine learning, data visualization* User Skills: *python, pandas, data visualization*

Missing Skills: *statistics, machine learning*

SGI: 40%

This illustrates that SGI provides a meaningful numerical indicator of a user's readiness for a specific role. A lower SGI correlates with higher competency, enabling users to understand precisely how prepared they are and what skills must be acquired to progress further. **Recommendation Results**

Based on the user's missing skills, the system generates **personalized and actionable recommendations**. For instance, when the missing skills included *machine learning* and *statistics*, the system recommended learning resources such as:

Machine Learning Specialization – Coursera 2. Statistics Essentials – Udemy 3.

Data Science Foundations – IBM SkillsBuild

This demonstrates that the recommendation module effectively connects identified gaps with targeted learning paths. These recommendations align with the role requirements and were found to be practical and relevant for skill development.

9.4 *API Output and System Behavior*

During testing, the **/assess** API endpoint reliably produced structured JSON responses containing:

Extracted skills Predicted roles and probabilities SGI for each role

Missing skills

Recommended skills list

Response times were consistently low due to the lightweight nature of the ML pipeline. The system's output format remained stable across different inputs, demonstrating reliability and robustness suitable for deployment in academic or training environments.

10. Conclusion

This paper presented an AI-driven Skill Gap Identifier and Career Guidance Platform capable of transforming traditional career counseling into a continuous, data-driven, and scalable service. The proposed architecture integrates NLP, machine learning, and analytics to support personalized career mapping and targeted upskilling.

Experimental results confirm that the system accurately identifies user competencies, predicts relevant job roles, and produces actionable recommendations. The SGI metric provides an intuitive and interpretable indicator of professional readiness, enabling users to monitor their development in a structured manner.

ISSN 2347–3657

Volume 14, Issue 1, 2026

Overall, the platform demonstrates strong potential for deployment in higher-education institutions, training centers, and workforce development programs.

11. Future Work

Future enhancements will focus on:

- replacing vocabulary-based extraction and TF–IDF models with transformer-based semantic encoders,
- enabling live integration with large-scale MOOC and job-market APIs,
- extending the role catalog to emerging domains such as cybersecurity, cloud engineering, and digital healthcare,
- introducing multilingual resume parsing to broaden accessibility,
- deploying role-based dashboards for educators and policymakers
- .

References

1. Ian Goodfellow, Yoshua Bengio, and Aaron Courville, *Deep Learning*, MIT Press, 2016.
2. Aurélien Géron, *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow*, 2nd Edition, O'Reilly Media, 2019.
3. Christopher M. Bishop, *Pattern Recognition and Machine Learning*, Springer, 2006.
4. Tom M. Mitchell, *Machine Learning*, McGraw-Hill Education, 1997.
5. Russell, S., & Norvig, P., *Artificial Intelligence: A Modern Approach*, 4th Edition, Pearson, 2021.
6. James F. Kurose, Keith Ross, *Computer Networking: A Top-Down Approach*, Pearson, 7th Edition.
7. National Association of Colleges and Employers (NACE), "Career Readiness Competencies," 2022.
8. BERT Research Paper — Jacob Devlin et al., "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," 2018.
9. Logistic Regression Theory — Hosmer & Lemeshow, *Applied Logistic Regression*, Wiley Publications, 2000.
10. JSON API Standards — "JSON API Specification," https://jsonapi.org
11. Python Software Foundation, "Python Language Documentation," https://docs.python.org