Journal of Universal Computer Science, vol. 32, no. 2 (2026), 286-302
submitted: 26/4/2025, accepted: 27/8/2025, appeared: 28/2/2026

JobMatcher: Multi-Layer Personalized and Inclusive Job Recommendations

Mashael M. Alsulami
(Taif University, Taif, KSA
https://orcid.org/0000-0003-2298-8278, mashael.s@tu.edu.sa)

Kholoud Althobaiti
(Taif University, Taif, KSA
https://orcid.org/0000-0002-7299-7095, kholod.k@tu.edu.sa)

Haneen Algethami
(Taif University, Taif, KSA
https://orcid.org/0000-0002-7582-4480, hmgethami@tu.edu.sa)

Abstract: Job recommendation systems play a critical role in matching individuals with relevant career opportunities based on their skills and experiences. However, many existing systems struggle to balance precision and contextual relevance, leading to mismatches in job recommendations. In this paper we introduce JobMatcher, a multilayered recommendation system that integrates a well established technique, cosine similarity and KNN clustering with ChatGPT based evaluation. Initial recommendations are generated through content-based filtering and refined via clustering similar job descriptions aligned with user profiles by seniority and trajectory. To enhance contextual accuracy, GPT 3.5 turbo was prompted to act as an expert evaluator, scoring top recommendations based on skill relevance and career fit using structured and unbiased prompts. In a user study with seven domain experts and ten user profiles, system-selected jobs scored significantly higher (mean = 3.43 *compared to* 2.99 for KNN clustering, $p = 0.0035$), with moderate inter-rater agreement (Kendall's W = 0.417). JobMatcher bridges algorithmic filtering with human like evaluation, offering a scalable, intelligent solution for improved job matching.

Keywords: Recommendation systems, job recommendations, content based recommendations, ChatGPT as evaluator
Categories: H.3.1, H.3.2, H.3.3, H.3.7, H.5.1
DOI: 10.3897/jucs.157024

1 Introduction

Job recommendation systems play an important role in connecting job seekers with opportunities that correspond to their abilities, qualifications, experience, and professional ambitions. Not only that, but also the demand for intelligent, flexible, and exact job-matching systems has escalated as the employment market grows more dynamic [Mandalapu et al. , 2023]. Additionally, balancing technical relevance in a recommendation system with the real-world applicability of recruitment must be more inclusive of a wide range of features [Tran , 2023]. Especially when the aim is to find prospective hires more suitable to the current job market.

Traditional methods such as content-based filtering and collaborative filtering [Ricci et al. , 2021] have established the basis for numerous job recommendation systems, providing systematic methods for aligning prospective hires with job descriptions. Such approaches succeeded in ensuring job recommendations match their abilities and qualifications; however, they have failed to include relevant job opportunities and inclusive job preferences. For example, traditional content-based filtering effectively identifies textual alignment between a user's profile and job descriptions. However, it often struggles to capture higher-level contextual nuances, such as career trajectory, industry shifts, or skill adaptability [Javed et al. , 2021; Jin et al. , 2023; Mehrabi et al. , 2021], while collaborative filtering suffers from cold-start problems and lacks explainability [Nassar et al. , 2020]. Hybrid models could improve accuracy but often require complex integration and lack real-time adaptability [Bhowmick et al. , 2021]. Furthermore, prior systems rarely incorporate human-like judgment or domain-specific alignment, which leads to mismatches in job recommendations.

In order to include these features, an approach that combines diverse recommendation techniques with human-like evaluative insights is required [Al-slaity , 2021]. The goal is to ensure that job suggestions are contextually relevant and practically feasible. Therefore, this study proposes JobMatcher, a multilayered job recommendation system that combines two layers of content-based filtering, including cosine similarity, and KNN search algorithm with AI-driven evaluation using ChatGPT [OpenAI , 2024]. This study therefore set out to assess the effect of including real-world features such as career progression, seniority alignment, or role-specific expectations on the reliability of the recommendation system. In addition, this study evaluates the performance of ChatGPT as an adaptive refinement mechanism in generating final job recommendations. To this end, the specific objectives of this study are:

- Develop a multilayered job recommendation system that integrates content-based filtering and the K-Nearest Neighbors (KNN) algorithm to generate personalized and inclusive job matches.
- Enhance recommendation quality using ChatGPT as an intelligent filtering mechanism to refine job suggestions based on contextual relevance, skill alignment, and career trajectory.
- Evaluate the system's effectiveness by comparing ChatGPT's selections with KNN-based recommendations and human expert assessments to ensure reliability and alignment with real-world hiring practices.

This multilayered approach mimics human decision-making in job searching. Just as a recruiter first searches for technical matches before assessing a prospective hire's career trajectory and aspirations, a job seeker typically begins by identifying positions that align with their current skills and experience. They then refine their search by considering long-term career growth, industry trends, and personal aspirations. By structuring the recommendation process in a similar manner—starting with technical relevance and then incorporating higher order career factors, this system provides job recommendations that balance immediate fit with future career potential. This study makes a major contribution by demonstrating that system generated recommendations are in alignment with expert ratings compared to the initial results from the content-based layer. Thus, the integration of ChatGPT offers a realistic evaluation of job alignment, facilitating the connection between algorithmic suggestions and expert confirmation. The novel contributions of this study include:

- The design of a multi-layered framework that sequentially applies content-based filtering including KNN, and a large language model (LLM) based evaluation.
- The use of LLM, specifically ChatGPT (GPT-3.5-turbo), to simulate expert judgment in evaluating job suitability.
- A comprehensive evaluation pipeline involving human experts to validate LLM-assisted job recommendations.

Following the introduction, the remainder of the paper is organized as follows: Section 2 reviews related work. Section 3 describes the method used in this study with a walkthrough example in Section 4. The experimental study and its results are presented in Sections 5 and 6 along with the study limitations 7. Finally, the paper concludes with Section 8.

2 Related Work

Recommendation systems have evolved significantly, leveraging different methodologies to enhance personalization and relevance. These systems are broadly categorized into content-based filtering, collaborative filtering, knowledge-based systems, and hybrid approaches [Ricci et al. , 2021]. Among these, content-based filtering has played a crucial role in job recommendation systems by emphasizing semantic similarity measures. This method enables systems to suggest jobs that align with users' skill sets, broadening the scope of job opportunities beyond exact matches to related competencies. For instance, Domeniconi et al. [2016] utilized LinkedIn profile data to compare user skills with job postings, demonstrating improved recommendation accuracy. However, challenges remain, particularly regarding computational efficiency and the ability to account for career trajectory shifts.

To address limitations in content-based filtering, collaborative filtering has been integrated to personalize recommendations further. A multi-criteria collaborative filtering model proposed by Nassar et al. [2020] who incorporated item interactions to refine job matches, leading to significant improvements in accuracy and user satisfaction. Additionally, clustering-based collaborative filtering techniques, such as K-means clustering, have been employed to manage large datasets efficiently. By grouping similar users and leveraging collective preferences, these approaches enhance scalability and recommendation diversity [Lee et al. , 2024]. However, traditional collaborative filtering faces cold-start issues, limiting its effectiveness for users with minimal historical data.

Hybrid recommendation systems have emerged as a solution by combining the strengths of content-based and collaborative filtering approaches. These models have demonstrated effectiveness across various domains, including e-commerce [Krishna et al. , 2025] and education [Butmeh and Abu-Issa , 2024]. In job recommendation contexts, researchers have explored hybrid models integrating clustering techniques with content-based filtering. For example, Bhowmick et al. [2021] and Kamdar and Mehta [2023] proposed a method where users are first grouped based on preferences, followed by cosine similarity-based content filtering within clusters. This approach enhances recommendation precision while mitigating sparsity and cold-start issues.

LLMs have introduced new paradigms in recommendation evaluation and generation. The work by Brown et al. [2020] highlighted the capabilities of LLMs to generalize across diverse evaluation settings without specialized training. LLMs possess emergent properties such as contextual awareness and adaptive decision-making, which make them valuable tools for refining recommendations [Wei et al. , 2022; Gao et al. , 2023]. Recent studies have explored how LLMs, particularly ChatGPT, can improve recommendation tasks through role-playing prompts and expert simulations. By simulating domain experts, these models can assess job recommendations beyond traditional metrics [Dai , 2023].

LLMs have also been explored for generating recommendations directly. Several studies have examined how transformer-based architectures can dynamically personalize recommendations based on historical user interactions [Zhao et al. , 2024; Chen et al. , 2023]. The ability of LLMs to capture complex patterns in user behavior makes them promising for adaptive recommendation systems. Furthermore, Gao et al. [2023] introduced a ChatGPT-based recommendation framework that demonstrated significant improvements in ranking relevance and diversity compared to traditional filtering techniques.

Incorporating LLMs into recommendation generations and evaluations has yielded promising results. For instance, an evaluation pipeline was proposed to measure ChatGPT's effectiveness in Top-N recommendations, cold-start cases, and ranking tasks [Di Palma et al. , 2023]. The findings indicated that ChatGPT significantly outperformed baselines, demonstrating its ability to provide contextually relevant rerankings. Similarly, a study that investigated LLM-driven user modeling, revealing improvements in personalized relevance assessments. However, challenges such as potential biases and the need for standardized evaluation frameworks remain [Wang et al. , 2024a].

The integration of LLMs with hybrid recommendation models presents a novel direction for job recommendation systems. By combining content-based filtering with KNN clustering for initial job retrieval and employing ChatGPT as an adaptive filtering mechanism, our approach aligns with recent advancements in AI-driven recommendation frameworks. The role of a LLM, specifically ChatGPT in refining job matches based on career trajectory and industry-specific factors builds upon prior research in AI-assisted recommendations [Liu et al. , 2023; Dai , 2023; Haque et al. , 2022]. This integration not only enhances recommendation relevance but also improves interpretability by providing explanations for job suggestions, a critical aspect in career decision-making processes [Mehrabi et al. , 2021].

In summary, prior research underscores the effectiveness of hybrid models in overcoming the limitations of content-based and collaborative filtering techniques. The incorporation of LLMs such as ChatGPT introduces a new dimension to recommendation refinement, ensuring context-aware, realistic career-aligned job suggestions. Our methodology extends these advancements by integrating the KNN algorithm with LLM-driven expert evaluation, providing a robust and scalable framework for job recommendation.

3 Multilayered Job Recommendation Framework

The Multilayered Job Recommendation Framework enhances job matching by combining content-based filtering with AI-driven refinement. The first layer uses initial textual matching and KNN-based job expansion to retrieve relevant jobs based on cosine similarity, ensuring both precision and diversity. The second layer, ChatGPT-based filtering, refines recommendations by considering career growth, seniority alignment, and industry-specific factors. The overall architecture of the proposed framework is illustrated in Figure 1.

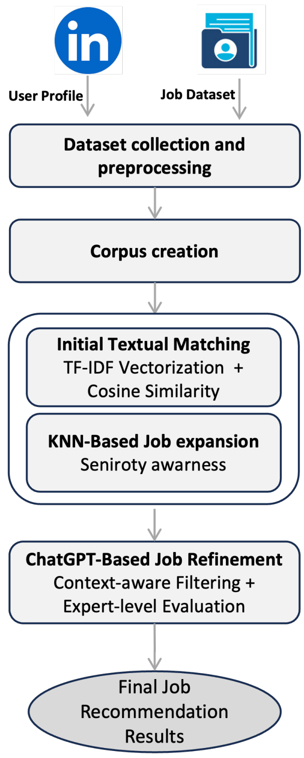

Figure 1: The workflow of the proposed multilayered job recommendation system.

3.1 Dataset Collection and Pre-processing

This research study utilized two primary datasets: the Monster Job Descriptions Dataset [PromptCloud and DataStock , 2020] with 1,489 job descriptions and the LinkedIn Profiles Dataset containing 762 user profiles. The Monster dataset included detailed job information such as titles, descriptions, company offering positions, salary information, and other relevant job-related data. The LinkedIn dataset encompassed users' work history, positions held, skills, and professional backgrounds. Both datasets were processed in CSV format using data processing tools, ensuring consistency and readiness for analysis.

To ensure high-quality recommendations, a comprehensive preprocessing workflow was applied. Only English-language job descriptions and profiles were retained for consistency. User profiles with fewer than two years of experience or lacking job history were excluded to ensure meaningful career context and allow for reliable similarity-based matching. URLs and irrelevant metadata were removed from both datasets, minimizing noise and focusing on essential textual content. These preprocessing steps ensured that the system operated on clean, structured, and contextually relevant data, improving the accuracy of job matches [Ladani and Desai , 2020].

3.2 Content-Based Recommendation Layer

The first layer applies content-based filtering using cosine similarity, which efficiently identifies the single best-matching job for each user based on textual overlap between job descriptions and user profiles. This ensures that the starting point is a highly relevant job rather than relying on broad clustering. Once this anchor job is established, the KNN algorithm expands the search by retrieving the top five most similar jobs, considering both textual similarity and career progression factors like seniority alignment.

3.2.1 Initial Textual Matching

The recommendation process begins with text-processing techniques that identify job recommendations most aligned with LinkedIn profiles based on skill relevance. The TF-IDF method [Qaiser and Ali , 2018] was used to convert the textual content of both job descriptions and user profiles into structured numerical vectors, emphasizing important terms while minimizing common or uninformative words. This process resulted in a high-dimensional TF-IDF matrix with a shape of $(2504, 14743)$, where 2,504 represents the combined number of documents and 14,743 reflects the vocabulary size extracted from the corpus.

To manage this complexity and enhance computational efficiency, the TF-IDF matrix was reduced using Truncated Singular Value Decomposition (SVD)[Wang et al. , 2024b], reducing each vector to 300 components that preserved the most salient semantic features. Cosine similarity [Gunawan et al. , 2018] which was then applied to these reduced 300-dimensional vectors to quantify the similarity between user profiles and job descriptions. This similarity score, ranging from -1 to 1, allowed the system to rank job recommendations by how well they matched each user's skills and experience.

The initial job recommendations were generated by extracting the relevant row from the similarity matrix for each LinkedIn profile and ranking the job listings in descending order of similarity. By focusing on the top-matching entries, the system ensured that users received recommendations that were not only textually relevant but also meaningful in the context of their professional background.

The final recommendation dataset was structured to include each user's ID, current position, recommended job title, and the corresponding cosine similarity score.

In this step, a preliminary set of job recommendations was successfully generated by analyzing the textual content of job descriptions and LinkedIn profiles. After processing 1,489 job listings from the Monster dataset, cosine similarity scores were calculated, yielding a similarity matrix for 300 processed components. Table 1 presents the top five recommended jobs along with their similarity scores.

Each LinkedIn profile received a recommended job title, with an average similarity score of 0.56, indicating a strong alignment between profiles and job descriptions.

While this initial filtering provides strong text-based matches, it does not introduce broader contextual factors such as seniority awareness and career trajectory.

3.2.2 KNN-Based Job Expansion

To increase diversity in the job suggestions generated by the first layer, this layer refines and expands the initial recommendations by applying the K-Nearest Neighbors (KNN) algorithm [Ge et al. , 2024] as a content-based filtering [Sawant et al. , 2023]. This layer ensures that users are presented with a broader range of relevant job opportunities. At the outset, we extracted seniority levels for both users and jobs. Job titles and user

Recommended Job Title	Similarity Score
Information Security Leader	0.88
SAP BODS and I.S.	0.83
Requirements Sharepoint, WebSphere	0.83
Senior Executive or Assistant Manager Import Export	0.82
Online Bidding - Pre Sales	0.79

* The cosine similarity scores reflect the degree of similarity between the job title and the given position.

Table 1: Top 5 Cosine Similarity Scores

positions were categorized into four seniority levels: Junior, Mid-Level, Senior, and Executive, using a keyword-based approach. This classification ensured that seniority was incorporated into the feature set, allowing the system to generate job recommendations that align with users' career stages.

We then vectorized both job descriptions and user profiles using TF-IDF. The job matrix was constructed using key features that capture different aspects of a job, including job title, job description, required skills, qualifications, and job seniority. These attributes provided a comprehensive representation of job postings, ensuring that matches were based on both content similarity and role alignment. Similarly, a user profile matrix was built to reflect both current and recommended career paths. This matrix was constructed using three core components: the user's previous job titles, their extracted seniority level, and the recommended job resulting from the initial matching stage only if its cosine similarity score was 0.5 or higher. This step ensured that only highly relevant first-layer recommendations contributed to the second stage. By combining both current and recommended jobs, this approach aligned new job recommendations with past experiences and projected career paths, ensuring that users were not confined to their existing roles but guided toward realistic career advancements.

Once the job and user matrices were constructed, the KNN model was applied to perform direct job-user matching using cosine similarity as the distance metric. For each user, the system identified the top 5 most similar jobs, ranking them based on how closely their textual features aligned with the user's profile. The choice of $K = 5$ was empirically validated by comparing recommendation diversity and seniority alignment across $K \in \{3, 5, 7, 10\}$. $K = 5$ provided the best trade-off between computational cost, diversity of results, and relevance based on preliminary qualitative inspection. Unlike traditional clustering approaches that group users into predefined categories, this approach continuously searches for the best individual matches, ensuring that recommendations are personalized at a granular level. The computed cosine distances were converted into similarity scores, where higher values indicated a stronger match between the user's profile and the job description.

While KNN-based matching ensured technical alignment but lacked contextual awareness, often missing career progression and industry shifts.

3.2.3 Evaluating the content-based recommendations

We evaluated the effectiveness of the KNN-based recommendation system by examining seniority matching statistics, similarity scores, and the diversity of recommended job titles. The seniority matching analysis reveals that, on average, users receive approximately 2.86 out of 5 job recommendations matching their seniority level. Notably, 26.15% of users received five perfect matches, demonstrating the system's capability to align recommendations with user seniority. Conversely, 13.85% of users received no seniority-aligned recommendations, which may indicate challenges in finding matches for certain profiles due to niche expertise or unconventional career paths. This suggests a potential area for improvement.

The similarity score distribution further supports the robustness of our model. With a mean similarity score of 0.57 and a median of 0.56, the majority of recommendations are reasonably aligned with user profiles. The relatively low standard deviation (0.14) indicates consistency in recommendation quality. We also examined the diversity of job recommendations to ensure that users receive varied opportunities rather than repetitive listings. The results indicate that most users receive five unique job recommendations, demonstrating that our model provides a broad range of suggestions rather than cycling through similar job titles. This enhances user exposure to different opportunities, improving the overall effectiveness of the system.

3.3 ChatGPT-based Recommendation Layer

To bridge algorithmic precision with contextual awareness, we employed a role-playing approach in which GPT-3.5-turbo functioned as an expert evaluator, assessing the suitability of recommended jobs resulting from KNN based on a user's skills and previous position. This builds on expert-based assessments, a widely used approach in recommender system evaluations [Han et al. , 2023]. While expert evaluations are traditionally conducted by human specialists, recent studies have explored the use of LLMs, specifically GPT-3.5-turbo (via ChatGPT), for simulating expert decision-making in recommendation tasks [Liu et al. , 2023; Dai , 2023; Haque et al. , 2022]. By leveraging GPT-3.5-turbo (via ChatGPT) in this capacity, we aimed to filter job recommendations to ensure that only the most relevant options were retained, providing an additional measure of recommendation quality.

Our approach relies on prompting rather than fine-tuning the language model. Prompting allows LLMs to be adapted to specific tasks through structured instructions in natural language, eliminating the need for extensive retraining [Zhao et al. , 2024]. Unlike fine-tuning, which requires specialized model adjustments, our method maintains the model's general-purpose nature while guiding its decision-making through detailed prompts. However, since ChatGPT is not explicitly fine-tuned for job evaluation tasks, it may not fully grasp the subtleties of complex job descriptions. To mitigate this, we designed structured input prompts to improve its contextual understanding and evaluation quality. Additionally, we excluded demographic information such as nationality and gender to prevent biases in the evaluation process [Mehrabi et al. , 2021], ensuring that assessments were based purely on skill alignment and career relevance.

To refine job recommendations, we integrated GPT-3.5-turbo (via ChatGPT) as a filtering layer to enhance the initial KNN-based results. The system retrieved user profiles, including their skills, previous job titles, and a list of jobs recommended by the KNN model. ChatGPT was then tasked with analyzing the top five recommended job descriptions and selecting the three most suitable positions based on skill relevance and career alignment. ChatGPT was then prompted with the instruction:

```
You are a job recommendation assistant. You are provided a list of
5 recommended jobs, select the most three suitable jobs from the
list for the user based on their following profile.
```

```
  User ID: {user_id}
  Skills: {user_skills}
  Current Job: {current_job}
  - {job_title_1}: {job_description_1}
  - {job_title_2}: {job_description_2}
  - {job_title_3}: {job_description_3}
  - {job_title_4}: {job_description_4}
  - {job_title_5}: {job_description_5}
Which job is the best fit?
```

ChatGPT assessed job relevance by processing this structured input and refined the recommendations to better match users' career trajectories. By bridging algorithmic precision with contextual awareness, this approach added an intelligent filtering mechanism to complement the KNN model.

3.3.1 Analysis of ChatGPT Results

We conducted several key analyses to evaluate GPT-3.5-turbo effectiveness in refining job recommendations after the KNN layer.

Initially, we examined the jobs suggested by ChatGPT to assess their alignment with the provided list. Our findings revealed that in 6 out of 65 cases, ChatGPT did not strictly adhere to the given job list, either selecting only one job from the provided options or suggesting additional jobs beyond those recommended by KNN. This observation raised questions about the consistency of ChatGPT's selection process and the underlying criteria it used. To further understand this behavior, we queried ChatGPT for clarification. The responses indicated that rather than strictly following the top-ranked KNN recommendations, ChatGPT prioritized semantic relevance and job alignment with the user's skills and current position. This suggested that ChatGPT applies additional reasoning beyond numerical similarity scores.

To quantify the extent of this deviation, we compared ChatGPT's selections with the top three job recommendations from KNN. The agreement rate between ChatGPT's selections and the top three KNN recommendations was found to be 60% on average, Figure 2. This result demonstrates that while ChatGPT considers KNN recommendations, it does not rigidly follow them, frequently opting for selections based on a broader interpretation of relevance. Consequently, this refinement layer introduces an element of flexibility that may enhance the quality of recommendations beyond strict similarity-based ranking.

Next, we compared the similarity scores of ChatGPT's selections against those of KNN's recommendations. The average similarity score of KNN-selected jobs was found to be 0.1465, while the average similarity score of ChatGPT-selected jobs was 0.1474. This small yet positive improvement suggests that ChatGPT does not significantly compromise similarity despite its alternative selection approach. Instead, it maintains a comparable level of similarity while integrating additional qualitative aspects into its decision-making. We investigated the similarity further by examining whether ChatGPT consistently selects the highest similarity jobs from KNN's recommendations. The results indicated that ChatGPT does not always prioritize the jobs with the highest similarity score– only 13.85% of the recommended jobs are in alignment with KNN's highest recommended jobs.

These findings highlight an essential trade-off in job recommendation systems: a strict adherence to numerical similarity versus a more nuanced approach that incorpo-

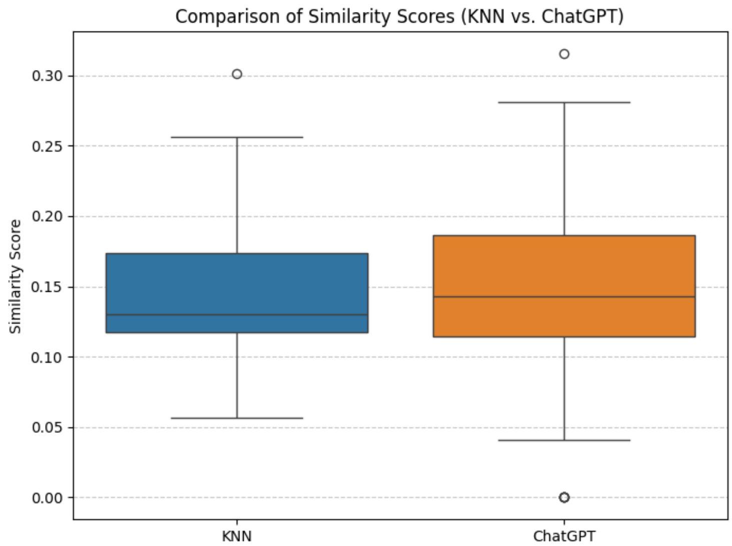

Figure 2: Comparison of Similarity Scores: Top 5 Jobs Recommended by KNN vs. Top 3 Jobs Selected by GPT-3.5-turbo (via ChatGPT)

rates contextual job relevance. ChatGPT acts as a refinement layer that enhances the recommendation process by balancing similarity with semantic understanding.

4 A Walkthrough Example of the Multi-Layered JobMatcher

To understand how the system would behave in real-world scenarios, we will walk through an example of the recommendations provided to User18 in our dataset. User18 has a strong background in software development, with experience spanning over a decade in system analysis, design, implementation, and deployment. Their current role as an Analyst Programmer involves working on monitoring and control systems, while their previous experience focused on content services for mobile devices. Their skill set is diverse, encompassing `.NET`, `C#`, `Mono`, `C++`, **SCADA**, **Java**, `Spring`, `Hibernate`, and web technologies such as `ASP.NET MVC`, `JavaScript`, and `HTML`. Additionally, expertise in databases like `MySQL`, Microsoft SQL Server, and SSAS 2008 reflects strong data-handling capabilities. The combination of enterprise application development, backend programming, and database management positions User18 as a versatile developer suited for both software engineering and technical leadership roles.

The first step of the first layer of the recommendation process applies content-based filtering using cosine similarity measurement to generate an initial job suggestion for the user. In this case, the highest-ranked job is *Sr. Software Engineer / Tech Lead*, with a similarity score of 0.74. This strong match indicates a high degree of overlap between the user's skills—primarily in ".NET, C#, ASP.NET, SQL Server, and Java"—and the job's technical requirements. Given the user's extensive background in software development, including experience in SCADA, Java Spring, Hibernate, and web technologies, this recommendation appears highly relevant. However, this step only considers textual similarity without factoring in job seniority.

In the KNN step, we refine the job recommendations by factoring in job seniority and broader contextual considerations. This leads to a revised list of five recommended jobs, ranked by similarity, shown in Table 2.

Recommended Job Title	Job Seniority	Similarity score
JSF Tech Lead	Senior	0.21
Sr. Software Engineer / Tech Lead	Senior	0.21
Software Engineer	Mid-Level	0.21
Sr. Software QA Engineer	Senior	0.16
Sr. Software Development Engineer	Senior	0.14

Table 2: Top 5 Recommended Job After Revising by KNN

The presence of one mid-level role (Software Engineer) alongside multiple senior-level positions (e.g., JSF Tech Lead, Sr. Software Engineer / Tech Lead, Sr. Software QA Engineer, Sr. Software Development Engineer) indicates that the model primarily considers the user suitable for senior positions, with a single mid-level recommendation as a potential alternative. The inclusion of JSF Tech Lead as a senior-level role suggests that while it aligns with the user's general expertise, its specific technology stack (JSF) might not be an ideal match, resulting in a moderate similarity score.

In the final stage, ChatGPT acts as an expert evaluator by reviewing the top five recommendations and selecting the three most suitable jobs based on a holistic assessment of skills, job descriptions, and career alignment:

1. Sr. Software Engineer / Tech Lead – Strongest match due to direct alignment with the user's expertise in .NET, C#, ASP.NET, SQL Server, and MVC framework.
2. Software Engineer – Selected due to its emphasis on UI development, leveraging the user's skills in HTML, JavaScript, and ASP.NET MVC.
3. Sr. Software QA Engineer – While slightly outside the user's core development expertise, it includes Java, SQL, and testing tools like Selenium, which overlap with the user's background.

Notably, ChatGPT excluded the JSF Tech Lead and Sr. Software Development Engineer roles, likely because JSF does not strongly align with the user's primary skill set, and the Sr. Software Development Engineer role may have been deemed less relevant compared to the alternatives. The inclusion of a QA-focused role suggests that ChatGPT recognizes the user's broader technical background and the potential for skill transferability, even though QA is not explicitly mentioned in the user's experience.

5 User Study

A user study was conducted to assess the accuracy and relevance of job recommendations based on users' skills and current job titles to evaluate the effectiveness of our job recommendation system. Ethical approval was obtained through the Taif University Ethical Committee, and all participants provided informed consent before participating in the study.

5.1 Participants Recruitment

All participants in the study work in the Information Technology field and belong to the information services and data processing industry. Participants were recruited through Prolific [Douglas et al. , 2023], an online research platform that provides access to a diverse and pre-screened pool of respondents. Prolific ensures high-quality data collection by verifying participant backgrounds and engagement levels. This approach helped enhance the validity of the evaluation by aligning participant backgrounds with the job recommendations being assessed.

5.2 Study Procedure and Design

Participants were presented with a structured evaluation task, which required them to assess the suitability of job recommendations for multiple user profiles. Each participant reviewed a total of five job recommendations per user profile and rated them accordingly. To further refine the screening process, user profiles in this survey were selected from a specific domain, ensuring that participants had relevant expertise within the field of study. Each user profile contains the user's skills, current job titles, and five recommended job titles. Participants were instructed to assess each recommendation based on its alignment with the user's expertise and career trajectory. To standardize the evaluation, participants rated each recommended job on a five-point Likert scale:

– **5 - Excellent:** The recommendation is highly relevant and represents a strong potential career progression.

– **4 - Good:** The job aligns well with the skills and experience of the user.

– **3 - Moderate:** There is some overlap, but the job would require notable upskilling.

– **2 - Weak:** There is minimal alignment and the job requires significantly different skills.

– **1 - Poor:** The recommendation is entirely unrelated to the user's skills and current job.

The first three recommendations in each user profile are the final system-generated recommendations, while the last two are considered non-recommended jobs in the KNN-based results. The evaluation process was designed to take approximately four minutes per participant.

Participants also provided qualitative feedback to further contextualize their ratings and identify potential areas for system improvement. In addition, a brief demographic questionnaire was included at the end of the study to collect information on the age, gender, education level, and job title of the participants.

All data collected were anonymized to protect participant confidentiality. No personally identifiable information was stored and responses were kept secure. Participation in the study was entirely voluntary and the participants had the right to withdraw at any time without penalty.

6 Results

The study aimed to compare system-generated recommendations against non recommended jobs derived from the KNN layer to determine whether the system effectively provided more relevant job suggestions. The evaluation was designed to measure both expert agreement with recommendations and the consistency of expert ratings across different user profiles. Additionally, the study sought to answer two key questions: *How do participants rate the system-generated recommendations compared to non-recommended ones? Is there a statistically significant difference between ChatGPT-recommended and non-recommended jobs?* and *Is there a suitable agreement between participants?*

A total of seven experts assessed job recommendations across ten user profiles, with each profile containing five job recommendations. The first three recommendations per profile were system-generated (the three jobs selected by the second layer of our system), while the last two were classified as non-recommended jobs (the remaining two). Participants rated each recommendation on a five-point Likert scale, evaluating its relevance based on the user's skills and career trajectory. The expert ratings were analyzed to determine the overall effectiveness of the system's recommendations and to identify any statistically significant differences between the ratings of recommended and non-recommended jobs.

Table 3 provides insights into the demographic and professional background of the individuals involved in evaluating the job recommendation system.

Gender	Age Group	Education Level	Current Job Title
Male	25 - 30	Bachelor's degree	Manager
Female	25 - 30	Bachelor's degree	IT
Female	25 - 30	Bachelor's degree	Recruiter
Female	25 - 30	Graduate or professional degree	Risk Analyst
Male	25 - 30	Bachelor's degree	IT TECHNICIAN
Male	41 - 50	Bachelor's degree	Senior Software Engineer
Male	25 - 30	High school diploma or equivalent	Support Engineer

*The participants were categorized by gender, age group, education level, and current job title to better understand their demographics.

Table 3: Participants Description

The evaluation specifically involved experts working in fields related to information technology such as data analysis, software engineering, information security, and information processing. This structured screening ensured that the job recommendation system was assessed by individuals with the necessary domain knowledge, leading to a more informed and contextually relevant evaluation of the recommendations generated.

To determine whether the system successfully prioritizes jobs that align with expert opinions, we compared the system-generated recommendations (the three jobs selected by the second layer of our system) with non-recommended jobs (the remaining two). The findings suggest that the system-generated recommendations were consistently rated higher by experts (mean = 3.43, std = 1.35, count = 210) compared to non-recommended jobs (mean = 2.99, std = 1.42, count = 140), as shown in Figure 3. This distinction highlights the system's ability to identify and prioritize jobs that align more closely with expert preferences, reinforcing its potential utility in job-matching applications.

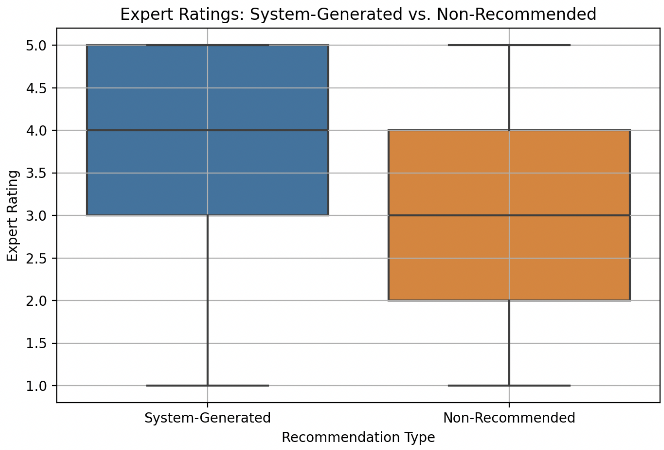

Figure 3: Expert ratings of system-generated jobs recommendations vs. non-recommended jobs

We then investigated whether the observed difference in ratings was statistically significant by applying an independent t-test. The results showed a t-statistic of 2.94 and a p-value of 0.0035, along with the 95% confidence interval for the difference in means: [0.15, 0.73], confirming that system-generated recommendations were rated significantly higher than non-recommended jobs.

We then measured the agreement among participants using Kendall's W, a statistical measure of concordance that evaluates the consistency of ordinal ratings across multiple raters [Bulut and Amasyali , 2017]. The analysis yielded a moderate agreement level (Kendall's W = 0.417), suggesting that while individual ratings varied, there was a meaningful level of consistency in how participants evaluated the job recommendations.

7 Limitations

Despite the promising results, this study has several limitations that may affect the generalizability of its findings. First, the dataset used was restricted to English-language job postings and LinkedIn-style user profiles, primarily within the IT domain. While this focus ensured consistency in evaluation, it may limit applicability to other languages or industries. Second, the seniority classification was based on title-based heuristics rather than organizational hierarchies, which may introduce mismatches in some cases. However, by incorporating multiple profile elements (e.g., previous job history, skill tags), we attempted to reduce the impact of isolated title-based misclassifications. Third, although LLMs such as GPT-3.5-turbo provide sophisticated contextual reasoning, they operate as black-box systems and may produce recommendations that are not fully explainable. To mitigate this, we used carefully structured prompts and avoided demographic bias by excluding sensitive attributes like gender and nationality. Lastly, while the human expert evaluation involved qualified raters, the small sample size (N=7) may limit the breadth of perspectives. Still, the observed inter-rater agreement (Kendall's W = 0.417) and statistical significance (p = 0.0035) support the reliability of the results.

8 Conclusion and Future work

This study investigated the integration of generative AI into content-based job recommendation systems, specifically utilizing a large language model, GPT-3.5-turbo as an intelligent refinement layer. By integrating content-based filtering with KNN algorithm, the system refines recommendations beyond textual similarity, incorporating career progression and seniority alignment. Additionally, AI-driven evaluation serves as an adaptive refinement mechanism, providing structured assessments to improve recommendation quality.

The results indicate that ChatGPT effectively refines job recommendations by filtering out mismatches and prioritizing roles that align with users' career trajectories. While traditional content-based methods generate initial job matches based on textual similarity, the inclusion of an AI-driven adaptive layer enhances qualitative judgment. This methodology ensures that recommendations are not only algorithmically relevant but also contextually meaningful, offering a significant advancement in job recommendations.

While the proposed system can be extended to applications such as career counseling tools and job portals, where personalized and adaptable recommendations are crucial, aligning AI-driven assessments with human expertise remains an ongoing challenge, necessitating further refinement and domain-specific calibration. Future research should focus on improving evaluation consistency and exploring broader applications across dynamic job markets. This study highlights the potential of multi-layered recommendation systems in balancing technical efficiency with real-world applicability, ultimately improving job-matching accuracy while reducing evaluation overhead.

Acknowledgements

The authors would like to acknowledge the Deanship of Graduate Studies and Scientific Research, Taif University for funding this work. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. During the preparation of this work, the authors used ChatGPT-4o in order to proofread the text. After using this tool, the authors reviewed and edited the content as needed and took full responsibility for the content of the publication.

References

[Al-slaity 2021]. Al-slaity, A. N. (2021). *Beyond Recommendation Accuracy: A Human-Like Recommender System*. PhD thesis, University of Ottawa.

[Bhowmick et al. 2021]. Bhowmick, H., Chatterjee, A., and Sen, J. (2021). Comprehensive Movie Recommendation System. *arXiv preprint arXiv:2112.12463*.

[Brown et al. 2020]. Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., et al. (2020). Language Models Are Few-Shot Learners. *Advances in Neural Information Processing Systems*, 33:1877–1901.

[Bulut and Amasyali 2017]. Bulut, F. and Amasyali, M. F. (2017). Locally Adaptive K Parameter Selection for Nearest Neighbor Classifier: One Nearest Cluster. *Pattern Analysis and Applications*, 20(2):415–425.

[Butmeh and Abu-Issa 2024]. Butmeh, H. and Abu-Issa, A. (2024). Hybrid Attribute-Based Recommender System for Personalized E-Learning with Emphasis on Cold Start Problem. *Frontiers in Computer Science*, 6:1404391.

[Chen et al. 2023]. Chen, L., Zaharia, M., and Zou, J. (2023). FrugalGPT: How to Use Large Language Models While Reducing Cost and Improving Performance. *arXiv preprint arXiv:2305.05176*.

[Dai 2023]. Dai, S. (2023). Uncovering ChatGPT's Capabilities in Recommender Systems. In *Proceedings of the 17th ACM Conference on Recommender Systems*, pages 1126–1132.

[Di Palma et al. 2023]. Di Palma, D., Biancofiore, G. M., Anelli, V. W., Narducci, F., Di Noia, T., and Di Sciascio, E. (2023). Evaluating ChatGPT as a Recommender System: A Rigorous Approach. *arXiv preprint arXiv:2309.03613*.

[Domeniconi et al. 2016]. Domeniconi, G., Moro, G., Pagliarani, A., Pasini, K., and Pasolini, R. (2016). Job Recommendation from Semantic Similarity of LinkedIn Users' Skills. In *International Conference on Pattern Recognition Applications and Methods*, volume 2, pages 270–277. SciTePress.

[Douglas et al. 2023]. Douglas, B. D., Ewell, P. J., and Brauer, M. (2023). Data Quality in Online Human-Subjects Research: Comparisons Between MTurk, Prolific, CloudResearch, Qualtrics, and SONA. *PLOS ONE*, 18(3):1–17.

[Gao et al. 2023]. Gao, Y., Sheng, T., Xiang, Y., Xiong, Y., Wang, H., and Zhang, J. (2023). Chat-rec: Towards interactive and explainable LLMS-augmented recommender system. *arXiv preprint arXiv:2303.14524*.

[Ge et al. 2024]. Ge, Y., Li, Y., Chen, X., Zhai, Z., Yao, T., Wang, T., and Mei, T. (2024). Mora: Parameter efficient multi-modal ranking adapter for retrieval-augmented generation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 20804–20813. IEEE.

[Gunawan et al. 2018]. Gunawan, D., Sembiring, C. A., and Budiman, M. A. (2018). The implementation of cosine similarity to calculate text relevance between two documents. *Journal of Physics: Conference Series*, 978:012120.

[Han et al. 2023]. Han, K., Wang, Y., Chen, H., Chen, X., Guo, J., Liu, Z., Tang, Y., Xiao, A., Xu, C., Xu, Y., Yang, Z., Zhang, Y., and Tao, D. (2023). A survey on vision transformer. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(1):87–110.

[Haque et al. 2022]. Haque, S., Eberhart, Z., Bansal, A., and McMillan, C. (2022). Semantic similarity metrics for evaluating source code summarization. *IEEE International Conference on Program Comprehension*, pages 36–47.

[Javed et al. 2021]. Javed, U., Shaukat, K., Hameed, I. A., Iqbal, F., Alam, T. M., and Luo, S. (2021). A Review of Content-Based and Context-Based Recommendation Systems. *International Journal of Emerging Technologies in Learning (iJET)*, 16(3):274–306.

[Jin et al. 2023]. Jin, D., Wang, L., Zhang, H., Zheng, Y., Ding, W., Xia, F., and Pan, S. (2023). A Survey on Fairness-Aware Recommender Systems. *ACM Computing Surveys*, 55(1):1–38.

[Kamdar and Mehta 2023]. Kamdar, A. and Mehta, I. (2023). Movie Recommendation System Using Composite Ranking. In *Proceedings of the conference Intelligent Systems and Machine Learning*. ICISML 2022. Springer.

[Krishna et al. 2025]. Krishna, E. S. P., Ramu, T. B., Chaitanya, R. K., Ram, M. S., Balayesu, N., Gandikota, H. P., and Jagadesh, B. N. (2025). Enhancing E-commerce recommendations with sentiment analysis using MLA-EDTCNet and collaborative filtering. *Scientific Reports*, 15:6739.

[Ladani and Desai 2020]. Ladani, D. J. and Desai, N. P. (2020). Stopword Identification and Removal Techniques on TC and IR Applications: A Survey. *2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS)*, 466–472.

[Lee et al. 2024]. Lee, Y. J., Swayamdipta, S., and Smith, N. A. (2024). Understanding the Effectiveness of Contrastive Data Selection Methods for Pretraining Large Language Models. In *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics (NAACL)*. Association for Computational Linguistics.

[Liu et al. 2023]. Liu, Y., Li, Y., Xie, Y., Zhao, W. X., and Wen, J.-R. (2023c). GPTScore: Evaluate as You Desire. *Information Processing and Management*, 60(2):103240.

[Mandalapu et al. 2023]. Mandalapu, S. R., Narayanan, B., and Putheti, S. (2023). Job Recommendation System Using Deep Reinforcement Learning (DRL). *International Journal on Recent and Innovation Trends in Computing and Communication*, 11(10s):621–630.

[Mehrabi et al. 2021]. Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., and Galstyan, A. (2021). A Survey on Bias and Fairness in Machine Learning. *ACM Computing Surveys*, 54(6):1–35.

[Nassar et al. 2020]. Nassar, N., Jafar, A., and Rahhal, Y. (2020). A Novel Deep Multi-Criteria Collaborative Filtering Model for Recommendation System. *Knowledge-Based Systems*, 187:104811.

[OpenAI 2024]. OpenAI (2024). GPT-3.5-Turbo and GPT-4 Large Language Models. OpenAI Technical Documentation Available at: https://platform.openai.com/docs/models.

[PromptCloud & DataStock 2020]. PromptCloud & DataStock (2020). Monster Job Listings Dataset 2020. https://www.kaggle.com/datasets/promptcloud/monster-job-listings-dataset-2020. Accessed: 2024-10-12.

[Qaiser and Ali 2018]. Qaiser, S. and Ali, R. (2018). Text Mining: Use of TF-IDF to Examine the Relevance of Words to Documents. *International Journal of Computer Applications*, 181:25–29.

[Ricci et al. 2021]. Ricci, F., Rokach, L., and Shapira, B. (2021). *Recommender Systems Handbook*. Springer.

[Sawant et al. 2023]. Sawant, A. M., Kunekar, P., Taywade, R., Tiwari, A., Sharma, R., Shinde, M., and Roy, A. (2023). Content-Based Filtering Using K-Nearest Neighbor. *2023 5th Biennial International Conference on Nascent Technologies in Engineering (ICNTE)*, pages 1–6.

[Tran 2023]. Tran, T. H. A. (2023). Explainable Artificial Intelligence in Job Recommendation Systems. Master's thesis, University of Twente.

[Wang et al. 2024a]. Wang, J., Lu, H., Liu, Y., Ma, H., Wang, Y., Gu, Y., Zhang, S., Han, N., Bi, S., Baugher, L., Chi, E. H., and Chen, M. (2024). LLMs for User Interest Exploration in Large-scale Recommendation Systems. In *Proceedings of the 18th ACM Conference on Recommender Systems (RecSys)*, pages 872–877.

[Wang et al. 2024b]. Wang, X., Zheng, Y., Wan, Z., and Zhang, M. (2024). SVD-LLM: Truncation-Aware Singular Value Decomposition for Large Language Model Compression. *arXiv preprint arXiv:2403.07378*.

[Wei et al. 2022]. Wei, J., Tay, Y., Bommasani, R., Raffel, C., Zoph, B., Borgeaud, S., Yogatama, D., Bosma, M., Zhou, D., Metzler, D., et al. (2022). Emergent Abilities of Large Language Models. *Transactions on Machine Learning Research*.

[Zhao et al. 2024]. Zhao, Z., Fan, W., Li, J., Liu, Y., Mei, X., Wang, Y., Wen, Z., Wang, F., Zhao, X., Tang, J., and Li, Q. (2024). Recommender Systems in the Era of Large Language Models (LLMs). *IEEE Transactions on Knowledge and Data Engineering*, 36(11):6889–6907.