Using Large Language Models and Recruiter Expertise for Optimized Multilingual Job Offer – Applicant CV Matching

Hamit Kavas[1,2] , **Marc Serra-Vidal**[2] , **Leo Wanner**[1,3]
[1]NLP Group, Pompeu Fabra University
[2]Adevinta Spain
[3]Catalan Institute for Research and Advanced Studies (ICREA)
hamit.kavas@upf.edu, marc.serrav@adevinta.com, leo.wanner@upf.edu

Abstract

In the context of the increasingly globalised economy and labour market, recruitment agencies face the challenge to deal with a magnitude of job offers and job applications written in a variety of languages, formats, and styles. Quite often, this leads to a suboptimal evaluation of the CVs of job seekers with respect to their relevance to a job offer. To address this challenge, we propose an interactive system that follows the "human-in-the-loop" approach, actively involving recruiters in the job offer – applicant CV matching. The system uses a fine-tuned state-of-the-art classification model that aligns job seeker CVs with labels of the *European Skills, Competences, Qualifications and Occupations* taxonomy to propose an initial match between job offers with the CVs of job candidates. This match is refined in sequential LLM driven-interaction with the recruiter, which culminates in CV relevance scores and reports that justify them.

1 Introduction

Globalization of the economy is paired with the globalization of the labour market, in which employers publish their job offers and job seekers pursue opportunities not only locally, but also internationally, across countries and even continents. This leads to a great variety of languages, formats, levels of detail as well as to an inconsistent use of terms and skill descriptions in job offers and job applications alike. Under these circumstances, the work of the recruiting and employment agencies becomes a challenge. Most of them use either the *European Skills, Competences, Qualifications and Occupations* (ESCO) taxonomy,[1] or its US equivalent O*Net[2] to align job offers with job applications. Proposals on automatic alignment of job offers with job applications also draw upon ESCO or O*Net [Colucci *et al.*, 2003; Zhang *et al.*, 2024; Clavié and Souli'e, 2023]. However, tree-like taxonomies such as ESCO or O*Net do not distingish between different flavours of an occupation. For instance, depending on the field of application (stock market analysis, education evaluation campaign, etc.), the profile of the occupation 'data analyst' might look rather differently, and so will do the corresponding job offers.

To address this challenge and improve the accuracy of the match between job offers and the CVs of job seekers, we advance the research in the area in two directions. [3] Firstly, we fine-tune the state-of-the-art ESCOXLM-R+ model for the classification of job seeker CVs with respect to the ESCO taxonomy [Kavas *et al.*, 2023] on a dataset containing English and Spanish job offers and job seeker CVs in order to identify candidates whose CVs are relevant to the ESCO occupation label(s) mentioned in the job offer under consideration. From the filtered CVs, the mentions of the available skills and competencies are extracted using [Nguyen *et al.*, 2024]'s Large Language Model (LLM)-based *Skill Extraction*. Secondly, we implement the *Manual Correction System* (MCS) [Cai *et al.*, 2023] for further CV filtering and match correction by involving the expert recruiters into the job offer – candidate CV matching procedure: The recruiter can mark via an interactive interface the relevance of each of the extracted skills and competencies and thus indicate the prioritization of each skill or competence. Subsequently, two-step Mixtral LLM-based [Jiang *et al.*, 2024] interactions evaluate candidates based on the relevant skills / competencies and recruiter comments, assigning relevance scores to each candidate. Lastly, another Mixtral LLM interaction within this sequential chain utilizes these scores to generate a summary report.

2 Related Work

Applicant Tracking Systems (ATSs) are used by recruiters to manage job applications. ATSs vary in their capabilities, each focusing on different aspects of recruitment, such as application tracking, candidate evaluation, and engagement throughout the hiring process [Koh *et al.*, 2023].

Some ATSs provide comprehensive reporting and analytics, whose outcome is fed into recruitment strategies. Integration with other *Human Resources* technologies is also a key feature since it enables seamless workflows across different recruitment tools. Others focus on *Candidate Relationship Management*, enhancing communication with appli-

[1]https://esco.ec.europa.eu/en/classification
[2]https://www.onetonline.org/

[3]For a demonstration of our system, consult https://youtu.be/Gom5hr8umf8.

Model	InfoJobs Spanish	Kaggle English
GTE-base embeddings	0.18	0.86
Multilingual e5-large embeddings	0.32	0.89
ESCOXLM-R+	0.95	0.98

Table 1: Model Performance Evaluation using Recall@5

cants, while others allow for customization and development in order to tailor the recruitment process and enhance employer branding, which involves promoting a company as an employer of choice to prospective hires. Examples of commercial ATSs are [Dover, 2023; HireVue, 2023; Lever, 2023; Greenhouse, 2023], and [Workable, 2023].

AI technologies are also increasingly used for recruitement. They include the use of generative models for job recommendations and generation of personalized CVs for improved job matching. Thus, [Zheng *et al.*, 2023] and [Zinjad *et al.*, 2024] focus on generating job descriptions from candidates' CVs and tailoring CVs to specific job offers, respectively. Techniques like data augmentation and contrastive learning by [Yu *et al.*, 2024] and synthetic job posting generation by [Magron *et al.*, 2024] further refine the matching process. Additionally, advancements in skill extraction and entity linking, as proposed by [Nguyen *et al.*, 2024] respectively [Zhang *et al.*, 2024] support the alignment of skills with job requirements, indirectly enhancing matching accuracy.

3 Our Approach

Our system utilizes the Mixtral 8x7B LLM [Jiang *et al.*, 2024] from mistral.ai[4] for its demonstrated ability to outperform previous models including GPT-3.5 Turbo, Claude-2.1, Gemini Pro, and Llama 2 70B - chat model on human benchmarks. The Mistral model is deployed using an Ollama Docker image [5], facilitating local operation and easy deployment across various environments. The entire application is designed to run on hardware equipped with an NVIDIA RTX 3090Ti GPU. The backend is developed in Python with FastAPI, and is also containerized in Docker for modularity and ease of maintenance. The frontend leverages React with TypeScript, styled using Tailwind CSS, and is similarly containerized. In what follows, we describe the methodology and the realization of the individual modules of our system.

3.1 Selecting the Right Classifier

To address the problem of multilingual classification of job offers and job seeker CVs, we explored the use of multilingual embeddings and of the state-of-the-art multilingual classification model ESCOXLM-R+ [Kavas *et al.*, 2023]. We observed that when fine-tuned on an English-Spanish dataset (see Section 3.2 below), the ESCOXLM-R+ model achieves a superior performance. Table1 compares the performance of two types of multilingual embeddings (GTE-base embeddings [Li *et al.*, 2023] and Multilingual e5-large embeddings [Wang *et al.*, 2024a]) against the fine-tuned ESCOXLM-R+ model, using Recall@5 as metric. The test datasets consist of 27,002 English documents from the same Kaggle dataset referenced in [Kavas *et al.*, 2023], and 28,629 Spanish documents, sourced from the InfoJobs database.[6] The classification labels employed in our analysis are derived from the ESCO taxonomy. The performance is evaluated on datasets from InfoJobs in Spanish and Kaggle in English.

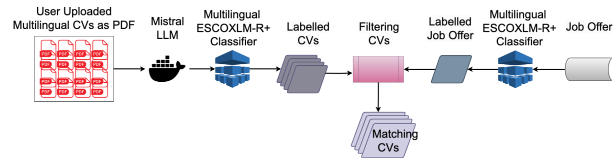

Figure 1: Matching process

3.2 Job Offer–Candidate CV Matching

In accordance with the outcome of the above analysis, we integrated the ESCOXLM-R+ model into our system; see Figure 1. To ensure that recruiters are presented with CVs that are most relevant to the job offer in question, we fine-tune the multilingual ESCOXLM-R+ model to systematically classify English, Spanish, or Catalan CVs and the job offer written in English with respect to the ESCO-taxonomy. For this purpose, we curated a training dataset comprising 63,161 documents in English and 50,521 in Spanish. For the fine-tuning procedure, we adopted a configuration of 10 epochs and a batch size of 8, utilizing the AdamW optimizer with a learning rate of 2e-5. Consistent with the referenced methodology, our training regimen included a warm-up phase, integral to the learning rate scheduling.

Each CV may contain several occupation experiences of the job seeker. Therefore, each of these experiences is classified by the fine-tuned ESCOXLM-R+ model with respect to the ESCO-taxonomy. The outcome of the ESCOXLM-R+ classification are ESCO occupation labels, tagged with the matching probability of a candidate experience / the job offer. The matching probability of an experience is aggregated together with the duration of the experience and its recentness, which quantifies how long the experience dates back, to a relevance weight of this experience, which is then abstracted into one of the three categories ('No', 'Mid', and 'High') of the matching score of the corresponding CV with the job offer. In formal terms, the matching procedure can be presented as follows:

Given a collection of job applicant CVs, $\mathcal{C} = \{CV_1, CV_2, ..., CV_n\}$, each containing a set of experiences $\mathcal{E}_i = \{E_{i1}, E_{i2}, ..., E_{im}\}$ for CV_i, and a job offer characterized by a set of required ESCO occupation labels $\mathcal{J} = \{J_1, J_2, ..., J_k\}$, with $k = 5$, the process of matching CVs to the job offer based on the relevance of their experiences proceeds as follows:

Each experience E_{ij} within CV_i is classified using the multilingual ESCOXLM-R+ model to determine its alignment with any of the job offer's assigned occupation labels J_k. This is represented by a binary function $I(E_{ij}, J_k)$, which is '1' if E_{ij} aligns with any $J_k \in \mathcal{J}$, and ´0' otherwise. Each experience E_{ij} mentioned in a job seeker's CV is

[4]https://mistral.ai/

[5]https://github.com/ollama/ollama

[6]https://www.infojobs.com/

assessed with respect to its relevance to the requirements in the job offer. The relevance is expressed by a weight $W(E_{ij})$, calculated as a function of the experience's duration and recentness and the normalized class probability $P(E_{ij}, J_k)$ assigned by the multilingual ESCOXLM-R+ model:

$$W(E_{ij}) = f(\text{duration}(E_{ij}), \text{recentness}(E_{ij})) \cdot P(E_{ij}, J_k)$$

where f represents the averaging function, encapsulating the importance of the experience's duration and its temporal proximity to the present, and $P(E_{ij}, J_k)$ adjusts the weight based on the model's confidence in the experience's relevance to the job classification.

The cumulative score $S(CV_i)$ for each CV_i is the sum of the weights of all experiences that are relevant to the job offer, expressed as:

$$S(CV_i) = \sum_{j=1}^{m} W(E_{ij}) \cdot I(E_{ij}, \mathcal{J})$$

This score represents the overall relevance of CV_i to the job offer, aggregating the weighted relevance of individual experiences that match the job's requirements.

Based on $S(CV_i)$, a matching score is assigned to each CV_i as follows:

- **No Match**: If $S(CV_i) < T_{\text{low}}$.
- **Mid Match**: If $T_{\text{low}} \leq S(CV_i) < T_{\text{High Match}}$.
- **High**: If $S(CV_i) \geq T_{\text{high}}$,

where $T_{\text{low}} = 25\%$ and $T_{\text{high}} = 75\%$ are predefined thresholds that delineate the categories of 'No match', 'Mid(ddle relevance)', and 'High (relevance)', respectively.

3.3 Skill Extraction and Candidate Evaluation

To evaluate and score candidate CVs based on required qualifications in the job offer, we utilize a sequential LLM chain consisting of four steps. **1.** building on [Nguyen *et al.*, 2024]'s study, we employ in-context learning to extract skills and competencies from job descriptions, exploiting few-shot learning capabilities of the Mixtral LLM through five-examples. Specifically, we adopt the extraction-style technique described in [Nguyen *et al.*, 2024], which allows the extraction of skills with longer sentence length. **2.** The extracted skills and competencies are then presented to recruiters via an interactive interface, which not only displays the extracted information, but also allows recruiters to label any skill or competency as "irrelevant" and to add specific comments for each of them. **3.** Another Mixtral LLM utilizes the recruiter's feedback to assess each candidate's suitability for the offer in view of their background, experience, and skills. This assessment includes the LLM's evaluation of the alignment between the candidate's skills and the job requirements, taking into account the importance and relevance of each skill as annotated by recruiters. If a recruiter designates any skill as irrelevant, that particular skill is subsequently omitted from the evaluation process. **4.** Finally, the same LLM assigns scores to each candidate's CVs, considering both the extracted skills and the recruiters' annotations. Scores range from 1 to 10, calculated

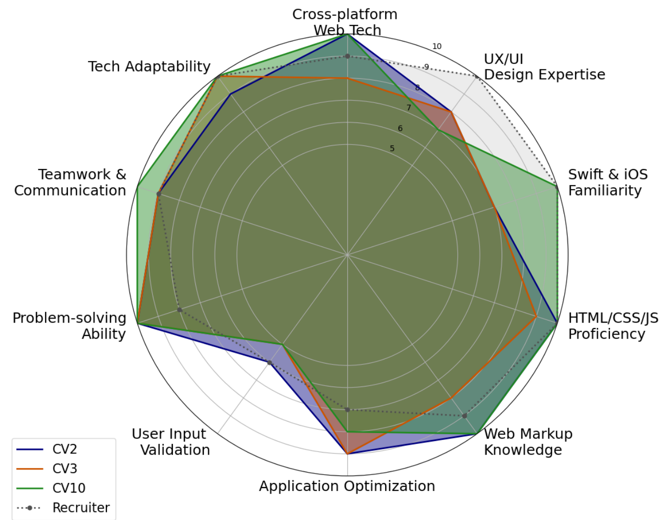

Figure 2: Example scoring results from our interface, demonstrating the evaluation of candidates based on skills and recruiter feedback.

to reflect each candidate's suitability and the recruiter's preferences, as illustrated in Figure 2.

To ensure accuracy and reduce potential bias in the evaluation, the scoring process is repeated 3 times for each candidate, with the final score being an average of these iterations.

3.4 Retrieval Augmented Generation

To provide the Mixtral LLM with domain-specific information and to mitigate hallucinations in the course of the assessment of the the experiences mentioned in candidate CVs, we employ the *Retrieval Augmented Generation* (RAG) method [Lewis *et al.*, 2020], which combines an information retrieval component with a generative model. Utilizing a vector store generated through the multilingual embeddings model [Wang *et al.*, 2024b], we encode 3015 distinct occupation definitions extracted from the ESCO taxonomy as retrieval context for the Mixtral LLM. We use sequential LLM interactions to conduct a comprehensive evaluation of the descriptions and related skills of each experience in a two-stage process. In the initial chain, the Mixtral LLM assesses the candidate's experiences and provides analyses, utilizing the vectorized ESCO definitions as a knowledge base. Subsequently, in the second chain, another Mixtral LLM conducts a comparative analysis. It generates a summary for each experience and integrates relevant job offer descriptions into the report for direct comparison. The final report provides a conclusive analysis of how well the candidate's experiences align with the requirements outlined in the job offers.

4 Summary and Future Work

We presented an advanced LLM-based system for optimized interactive job offer – job seeker CV matching. In the future, we plan to continue to experiment with fine-tuned LLMs and class embeddings for multilingual job offer / CV classification.

Ethical Statement

We acknowledge the potential for hallucination and bias in generative models utilized within recruitment processes, as noted by [Koh *et al.*, 2023]. To mitigate these concerns, we have opted not to assign numerical scores to candidates directly. Instead, our approach involves multiple calls to a language model with the purpose to diminish system bias, along with the implementation of Retrieval Augmented Generation (RAG) techniques that shall help reduce the risk of hallucination. Furthermore, the system prompts that guide the Mixtral LLM are designed to ensure that any negative aspects noted in user profiles are based strictly on verifiable data, adopting an observation-based narrative rather than inferential conclusions. Importantly, the feedback procedure embedded within our system ensures that the model adjusts to the distinct requirements of each user. This adaptability reflects personal preferences, distinguishing it from systemic bias and underscoring its role in personalizing the evaluation process.

References

[Cai *et al.*, 2023] Zefan Cai, Baobao Chang, and Wenjuan Han. Human-in-the-loop through chain-of-thought. *ArXiv*, abs/2306.07932, 2023.

[Clavié and Souli'e, 2023] Benjamin Clavié and Guillaume Souli'e. Large language models as batteries-included zero-shot esco skills matchers. *ArXiv*, abs/2307.03539, 2023.

[Colucci *et al.*, 2003] Simona Colucci, Tommaso Di Noia, Eugenio Di Sciascio, Francesco Donini, Marina Mongiello, and Marco Mottola. A formal approach to ontology-based semantic match of skills descriptions. *J. UCS*, 9:1437–1454, 01 2003.

[Dover, 2023] Dover. Home page. https://www.dover.com/, 2023. Accessed: 20-02-2024.

[Greenhouse, 2023] Greenhouse. Home page. https://www.greenhouse.com/uk, 2023. Accessed: 20-02-2024.

[HireVue, 2023] HireVue. Home page. https://www.hirevue.com/, 2023. Accessed: 20-02-2024.

[Jiang *et al.*, 2024] Albert Q. Jiang, Alexandre Sablayrolles, Antoine Roux, Arthur Mensch, Blanche Savary, Chris Bamford, Devendra Singh Chaplot, Diego de Las Casas, Emma Bou Hanna, Florian Bressand, Gianna Lengyel, Guillaume Bour, Guillaume Lample, L'elio Renard Lavaud, Lucile Saulnier, Marie-Anne Lachaux, Pierre Stock, Sandeep Subramanian, Sophia Yang, Szymon Antoniak, Teven Le Scao, Théophile Gervet, Thibaut Lavril, Thomas Wang, Timothée Lacroix, and William El Sayed. Mixtral of experts. *ArXiv*, abs/2401.04088, 2024.

[Kavas *et al.*, 2023] Hamit Kavas, Marc Serra-vidal, and Leo Wanner. Job offer and applicant cv classification using rich information from a labour market taxonomy. *SSRN Electronic Journal*, 2023.

[Koh *et al.*, 2023] Nam Ho Koh, Joseph Plata, and Joyce Chai. Bad: Bias detection for large language models in the context of candidate screening. *ArXiv*, abs/2305.10407, 2023.

[Lever, 2023] Lever. Home page. https://www.lever.co/, 2023. Accessed: 20-02-2024.

[Lewis *et al.*, 2020] Patrick Lewis, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, Mike Lewis, Wen-tau Yih, Tim Rocktäschel, Sebastian Riedel, and Douwe Kiela. Retrieval-augmented generation for knowledge-intensive nlp tasks. In *Proceedings of the 34th International Conference on Neural Information Processing Systems*, NIPS'20, Red Hook, NY, USA, 2020. Curran Associates Inc.

[Li *et al.*, 2023] Zehan Li, Xin Zhang, Yanzhao Zhang, Dingkun Long, Pengjun Xie, and Meishan Zhang. Towards general text embeddings with multi-stage contrastive learning. *ArXiv*, abs/2308.03281, 2023.

[Magron *et al.*, 2024] Antoine Magron, Anna Dai, Mike Zhang, Syrielle Montariol, and Antoine Bosselut. Jobskape: A framework for generating synthetic job postings to enhance skill matching. *ArXiv*, abs/2402.03242, 2024.

[Nguyen *et al.*, 2024] Khanh Cao Nguyen, Mike Zhang, Syrielle Montariol, and Antoine Bosselut. Rethinking skill extraction in the job market domain using large language models. *ArXiv*, abs/2402.03832, 2024.

[Wang *et al.*, 2024a] Liang Wang, Nan Yang, Xiaolong Huang, Linjun Yang, Rangan Majumder, and Furu Wei. Multilingual e5 text embeddings: A technical report. *ArXiv*, abs/2402.05672, 2024.

[Wang *et al.*, 2024b] Liang Wang, Nan Yang, Xiaolong Huang, Linjun Yang, Rangan Majumder, and Furu Wei. Multilingual e5 text embeddings: A technical report. *ArXiv*, abs/2402.05672, 2024.

[Workable, 2023] Workable. Home page. https://www.workable.com/, 2023. Accessed: 20-02-2024.

[Yu *et al.*, 2024] Xiao Yu, Jinzhong Zhang, and Zhou Yu. Confit: Improving resume-job matching using data augmentation and contrastive learning. *ArXiv*, abs/2401.16349, 2024.

[Zhang *et al.*, 2024] Mike Zhang, Rob van der Goot, and Barbara Plank. Entity linking in the job market domain. *ArXiv*, abs/2401.17979, 2024.

[Zheng *et al.*, 2023] Zhi Zheng, Zhaopeng Qiu, Xiao Hu, Likang Wu, Hengshu Zhu, and Hui Xiong. Generative job recommendations with large language model. *ArXiv*, abs/2307.02157, 2023.

[Zinjad *et al.*, 2024] Saurabh Bhausaheb Zinjad, Amrita Bhattacharjee, Amey Bhilegaonkar, and Huan Liu. Resumeflow: An llm-facilitated pipeline for personalized resume generation and refinement. *ArXiv*, abs/2402.06221, 2024.