AlexU-NLP at TalentCLEF 2025: Curriculum-Driven Hybrid Retrieval for Multilingual Job Title Matching

Notebook for the TalentCLEF Lab at CLEF 2025

Rana Barakat[1,*], Omar Mokhtar[1,*], Marwan Torki[1] and Nagwa Elmakky[1]

[1]*Computer and Systems Engineering Department, Alexandria University, Egypt*

Abstract

This paper describes our approach for TalentCLEF 2025 Task A, focusing on multilingual and cross-lingual job title matching. The core challenge lies in the inherent brevity and ambiguity of job titles across different languages (English, Spanish, and German) and professional sectors. Our methodology employs a curriculum learning strategy to fine-tune an embedding model, gradually exposing it to more complex data involving job titles and their descriptions. We further enhance retrieval performance through a hybrid system combining semantic search with BM25 keyword matching, followed by a multilingual cross-encoder reranker. Experimental results on the validation set demonstrate the effectiveness of our phased training approach and hybrid retrieval, achieving a top average mAP of 56% on the validation set and an average mAP of 53% on the test set.

Keywords

Embedding Models, Hybrid Retrieval, Curriculum Learning, Large Language Models, Human Resources, TalentCLEF

1. Introduction

The modern workplace has undergone a profound transformation in recent years, driven by technological innovation, globalization, and shifting social dynamics. Technological advancements—particularly in artificial intelligence and natural language processing—are reshaping how companies source, assess, and manage human capital. At the same time, the globalization of the workforce, enabled by remote hiring and digital collaboration tools, has introduced new complexities in matching candidates to job roles across linguistic and cultural boundaries. These developments require intelligent systems capable of handling large-scale, multilingual data while maintaining the semantic integrity of role descriptions and candidate profiles.

A central challenge in this space is the variability and ambiguity of job titles. Job titles are often brief, under-specified, and highly context-dependent.

Furthermore, different organizations frequently use distinct terms to describe similar roles. For instance, the positions of "Software Engineer," "Backend Developer," and "Platform Engineer" may share significant overlap in responsibilities, yet differ in naming conventions based on organizational or regional preferences. This terminological inconsistency becomes even more pronounced in multilingual contexts, where translation, cultural nuance, and domain-specific jargon further complicate the task of semantic alignment.

Task A of TalentCLEF 2025 [1] addresses this problem by focusing on multilingual and cross-lingual job title matching across English, Spanish, and German. The objective is to retrieve and rank relevant job titles for a given query title, leveraging both linguistic and contextual understanding. Effective solutions to this task must reconcile lexical variation, cross-language equivalence, and domain specificity while remaining computationally efficient and scalable.

In this notebook, we present AlexU-NLP's approach to Task A, which combines curriculum learning, hybrid retrieval, and re-ranking to address the multifaceted challenges of job title normalization

CLEF 2025 Working Notes, 9 – 12 September 2025, Madrid, Spain

*Corresponding author.

†These authors contributed equally.

es-rana.moustafa2024@alexu.edu.eg (R. Barakat); es-omar.mokhtar2019@alexu.edu.eg (O. Mokhtar)

and retrieval. Our methodology employs a curriculum-based fine-tuning strategy for a multilingual embedding model, progressively introducing more complex data—from isolated job titles to rich title-description pairs—thereby enhancing the model's capacity to learn meaningful semantic representations. To maximize retrieval performance, we adopt a hybrid strategy that integrates dense semantic search with BM25-based sparse retrieval, followed by a multilingual cross-encoder re-ranker.

This system not only achieves strong performance on the TalentCLEF 2025 validation and test sets but also demonstrates practical relevance in real-world Human Capital Management (HCM) scenarios. By improving the robustness and adaptability of job title matching systems, our approach supports more accurate talent identification and enhances the alignment between workforce capabilities and organizational needs in a multilingual, cross-sector labor market.

2. Related Work

In the evolving landscape of job recommendation systems, recent research has focused on enhancing the semantic understanding of job titles and descriptions to improve candidate-job matching. Zbib et al.[2] introduced an unsupervised method that learns job title similarities by leveraging noisy skill labels, demonstrating effectiveness in text ranking and job normalization tasks.

Complementing this, Laosaengpha et al. [3] proposed a Job Description Aggregation Network (JDAN) that derives job title representations directly from job descriptions, bypassing the need for explicit skill extraction and achieving superior performance over traditional skill-based approaches. Addressing multilingual challenges, Zhang et al.[4] developed ESCOXLM-R, a multilingual language model pre-trained on the ESCO taxonomy across 27 languages, which achieved state-of-the-art results on various job-related tasks.

Furthermore, Deniz et al. [5] combined unsupervised and contrastive learning techniques to create a multilingual job title encoder, enhancing cross-lingual job recommendation capabilities. These advancements collectively contribute to more accurate and inclusive job matching systems in a global context.

3. Dataset

The corpus for Task A [6] comprises job titles in English, Spanish, and German, spanning various job domains and professional sectors.

- Training Data: Provided as 15,000 pairs of related job titles per language (English, Spanish, German).
- Validation Data: Structured into three distinct files per language: queries, corpus elements, and qrels (query relevance assessments). This set contains 100 query job titles per language, each with a list of related job titles from the corpus. A knowledge base of 2,500 unique job titles per language serves as the corpus for retrieval tasks within the validation set.
- Test Data: A background set comprising 5,000 job titles. The evaluation is conducted on a subset of the background set, that will be a gold standard corpus of 100 job titles in each language.

4. Methodology

4.1. Data Augmentation

To enhance the contextual understanding of job titles during model fine-tuning and inference, we implemented a two-pronged data augmentation strategy. First, for enriching the training dataset, we used the European Skills, Competences, Qualifications and Occupations (ESCO) taxonomy [7]. We utilized this resource to source authentic job descriptions corresponding to the titles in our training set. The integration of these descriptions furnished the model with rich, real-world contextual information,

which is vital for effective fine-tuning. Second, to address the common challenge of missing descriptions for corpus entries during the inference phase, we employed the `Qwen3-14B` large language model (LLM) [8]. This model was tasked with generating synthetic yet contextually plausible job descriptions for each title within the inference corpus using zero-shot prompting. This ensured that every entry in our retrieval corpus, both for validation and testing, consisted of a title paired with a description, a step whose impact is quantified in our results (Section 6.4).

4.2. Embedding Model and Fine-tuning Rationale

We selected the `multilingual-E5-large-instruct` model [9] as our core embedding backbone due to its strong performance on multilingual tasks and its instruction-tuned architecture, which is beneficial for understanding task-specific nuances. Effectively fine-tuning such models for our task requires careful consideration, as simplistic fine-tuning strategies can present certain challenges. For instance, training exclusively on brief job title pairs risks overfitting to lexical patterns, limiting the model's ability to generalize to semantically equivalent but lexically diverse titles. Conversely, directly incorporating lengthy job descriptions from the start might lead the model to become overly reliant on this rich contextual data, diminishing its focus on the job titles themselves and impacting performance when such descriptions are absent or of variable quality. To navigate these pitfalls and foster a more balanced and robust learning process, we adopted a curriculum learning strategy, detailed in the subsequent section.

4.3. Curriculum Learning Implementation

We implemented a curriculum learning approach that incrementally increases the complexity of the training data. This allows the model to first establish a robust understanding of job title semantics before integrating the richer contextual information from job descriptions. A key aspect of our methodology is its emphasis on cross-lingual learning. For instance, for each English job title, our training data included not only pairs with its relevant English title and description but also pairs with translations of both the title and its description into German and Spanish. This systematic exposure to semantically equivalent information across languages is designed to encourage the alignment of embeddings in the multilingual space, thereby mitigating language-specific clustering. The curriculum was structured into several stages. These are briefly described in the following subsections, while full details regarding their specific configurations can be found in Section 5.4.

4.3.1. Initial Stages

The curriculum commenced with training on symmetric (job title, job title) pairs. These initial stages prioritized monolingual data, with the model being trained on pairs within each language (en-en, es-es, de-de) before progressing to cross-lingual title pairs (en-es, en-de). The primary objective of these early stages was to preserve and refine the pretrained model's intrinsic ability to align job titles based purely on their semantics, forming a strong foundation for subsequent learning.

4.3.2. Intermediate Stages

Following the initial alignment, the curriculum gradually introduced asymmetric pairs of (job title, job title + job description). This was done in a controlled manner, balancing these richer contextual pairs with the title-only examples from the preceding stages. Similar to the early stages, this phase also began with same-language pairs before incorporating cross-lingual pairs. This part of the curriculum acted as a form of soft domain adaptation, injecting more extensive contextual cues to help disambiguate job titles and discourage overfitting to short, potentially ambiguous title tokens, while simultaneously reinforcing the symmetric retrieval structure learned earlier.

4.3.3. Final Stages

The concluding stages of the curriculum placed a strong emphasis on the asymmetric (job title, job title + job description) format. This was intended to allow the model to fully adapt to the anticipated real-world inference conditions, where job descriptions are expected to provide significant contextual information.

4.4. Hybrid Retrieval and Reranking

Our retrieval architecture employs a two-stage process, consisting of an initial hybrid retrieval phase followed by a neural reranking mechanism.

In the first stage, we perform a hybrid search by combining signals from dense and sparse retrieval methods.

- For dense retrieval, query job titles and corpus entries (each comprising a job title and its corresponding description) are encoded into vector representations using our fine-tuned $E5_{large,instruct}$ model; relevance is then scored using cosine similarity.
- Concurrently, for sparse retrieval, we utilize the BM25 algorithm to compute lexical similarity scores between the query (title and description) and each corpus entry (title and description).

The relevance scores from these two retrieval components are first normalized to a common range. These normalized scores, denoted as $S_{BM25,norm}$ for BM25 and $S_{vec,norm}$ for the vector-based semantic similarity, are then integrated using a weighted linear fusion. The final fused score, S_{fused}, for each candidate document is computed as:

$$S_{fused} = (w_{BM25} \cdot S_{BM25,norm}) + (w_{vec} \cdot S_{vec,norm})$$

Based on empirical evaluation on our validation set, we determined the optimal weights to be $w_{BM25} = 0.15$ for the BM25 component and $w_{vec} = 0.85$ for the semantic vector component.

The resulting candidate list, ranked by S_{fused}, is subsequently passed to the second stage, where we used a fine-tuned version of `jina-reranker-v2-base-multilingual` cross-encoder model released by Jina AI [10] to perform a more fine-grained relevance assessment on the top-10 candidates to produce the final ranked output.

5. Experimental Setup

5.1. Baseline

We established a baseline using the `paraphrase-multilingual-MiniLM-L12-v2` model [11], which was the official baseline for this task, providing a reference for measuring improvements.

5.2. Embedding Model Selection (Zero-shot)

We evaluated several pretrained multilingual embedding models in a zero-shot setting on the validation data. The models tested included `BGE-M3` [12], `multilingual-E5-large`, `multilingual-E5-large-instruct` [9], `LaBSE` [13], and `paraphrase-multilingual-mpnet-base-v2` [11]. The `multilingual-E5-large-instruct` model demonstrated significantly superior performance, leading to its selection as our base model.

5.3. Fine-tuning Approaches (Without Full Curriculum)

The fine-tuning experiments utilized Multiple Negatives Ranking Loss (MNRL). The positive instances for this loss consisted of pairs structured as either (job title, job title) or (job title, job title + description), where the first element consistently served as the anchor and the second as the positive example. We conducted preliminary fine-tuning experiments:

- Fine-tuning multilingual $E5_{large,instruct}$ on (job title, job title) pairs resulted in overfitting.
- Fine-tuning multilingual $E5_{large,instruct}$ on (job title, job title + description) pairs showed substantial improvement over the title-only approach and zero-shot performance.

We also experimented with ESCOXLM-R [4], a multilingual transformer model pretrained on ESCO data. We fine-tuned the model on (job title, job title + description) pairs, testing mean, [CLS] token, and attention pooling strategies for deriving sentence embeddings. All pooling strategies yielded similar average mAP scores on the validation set.

5.4. Curriculum Learning Configuration

The cross-lingual curriculum learning strategy for the multilingual $E5_{large,instruct}$ model was implemented in six stages. Throughout each curriculum stage, we also employed MNRL. The data composition for each stage was as follows:

- **Stage 1:** 100% monolingual (job title, job title) pairs (en-en, es-es, de-de).
- **Stage 2:** 60% monolingual (job title, job title) pairs; 40% cross-lingual (job title, job title) pairs (en-es, en-de).
- **Stage 3:** 60% monolingual (job title, job title); 20% cross-lingual (job title, job title); 20% monolingual (job title, job title + description).
- **Stage 4:** 30% monolingual (job title, job title); 20% cross-lingual (job title, job title); 30% monolingual (job title, job title + description); 20% cross-lingual (job title, job title + description).
- **Stage 5:** 10% monolingual (job title, job title); 10% cross-lingual (job title, job title); 60% monolingual (job title, job title + description); 20% cross-lingual (job title, job title + description).
- **Stage 6:** 70% monolingual (job title, job title + description); 30% cross-lingual (job title, job title + description).

5.5. Reranker Fine-tuning

For the second-stage refinement of our retrieval pipeline, we employed the `jina-reranker-v2-base-multilingual` cross-encoder model. This model was specifically fine-tuned for the task using a Binary Cross-Entropy (BCE) loss function. The training data was formulated from (query, candidate document, label) tuples. Each candidate document consisted of a job title concatenated with its corresponding description. Positive instances were created using known relevant query-document pairs ($label = 1$), while hard negative instances ($label = 0$) were incorporated to improve the model's discriminative power. These hard negatives were mined from the corpus by utilizing the `jina-embedding-v3` model [14] to retrieve documents (job titles and their descriptions) that were highly ranked for a given query but were non-relevant.

5.6. LLM-based Reranking

We also explored using a Large Language Model for reranking. The top 10 documents retrieved by our dense retrieval (curriculum-trained $E5_{large,instruct}$) were presented to the `Gemma3-27B` model [15]. The model was prompted to reorder these candidates based on relevance to the query. The specific prompt template utilized for this task is detailed in Listing 1:

Listing 1: LLM Reranking Prompt Template

```
Below is a query and a list of 10 candidate job titles.
Rank them from most relevant (rank=1) to least relevant (rank=10).
Query: {query}
Candidates:
{candidates}
Return only the ordering as comma-separated numbers (e.g. 3,1,2,...):
```

In the prompt above, `{query}` is replaced with the actual query job title, and `{candidates}` is replaced with a numbered list of the top-10 candidate job titles and their descriptions retrieved by the dense model.

6. Results and Discussion

All results reported in this section are Mean Average Precision (mAP) scores obtained on the official validation set, unless otherwise specified for test set evaluations.

6.1. Zero-shot Embedding Model Performance

To establish a baseline and select a strong foundation model, we evaluated several pretrained multilingual embedding models in a zero-shot setting. Table 1 presents these results. The $E5_{large,instruct}$ model

Table 1
Zero-shot Performance of Multilingual Embedding Models on the Validation Set (mAP).

Model	Spanish	English	German	Average
paraphrase-multilingual-MiniLM-L12-v2 (Task Baseline)	0.3776	0.4992	0.2681	0.3816
`bge-m3`	0.3051	0.4552	0.2571	0.3391
`multilingual-e5-large`	0.3387	0.4187	0.2690	0.3421
`multilingual-e5-large-instruct`	**0.4445**	**0.5831**	**0.3908**	**0.4727**
`LaBSE`	0.3676	0.4280	0.2669	0.3542
`paraphrase-multilingual-mpnet-base-v2`	0.4176	0.5382	0.2892	0.4150

achieved the highest average mAP of 47.27%, outperforming the official task baseline by approximately 9%. This superior performance justified its selection as our base model for fine-tuning.

6.2. Performance of Fine-tuned Models (Without Full Curriculum)

We then investigated the impact of fine-tuning on performance, initially without employing the full curriculum learning strategy. Table 2 shows the results for $E5_{large,instruct}$ and ESCOXLM-R (with various pooling strategies) when fine-tuned on pairs of (job title, job title + job description). Fine-

Table 2
Fine-tuned Model Performance on the Validation Set (mAP) without Curriculum Learning.

Model	Pooling Strategy	Spanish	English	German	Average
ESCOXLM-R	Attention	0.4980	0.5971	0.4509	0.5153
ESCOXLM-R	Mean	0.5058	0.6136	0.4629	0.5274
ESCOXLM-R	CLS	0.5067	0.6165	0.4625	0.5286
$E5_{large,instruct}$	–	**0.5081**	**0.6215**	**0.4850**	**0.5382**

tuning $E5_{large,instruct}$ improved its average mAP from 47.27% (zero-shot) to 53.82%. This underscores the value of incorporating job descriptions, which provide essential context for disambiguating job titles. The $E5_{large,instruct}$ model also consistently outperformed the ESCOXLM-R variants, despite the latter being pretrained specifically on ESCO data. We hypothesize that the base architecture and the instruction-tuning of the E5 model make it more amenable for fine-tuning on this specific task structure.

6.3. Impact of Curriculum Learning and Reranking Strategies

Table 3 illustrates the performance of our system, which incorporates the full cross-lingual curriculum learning strategy for $E5_{large,instruct}$, and subsequently evaluates the impact of different reranking

Table 3
Performance of Curriculum-Trained $E5_{large,instruct}$ with Different Reranking Strategies on Validation (dev) and Test Sets (mAP).

System Configuration	Language	mAP (dev)	mAP (test)
$E5_{large,instruct}$	English	0.6497	**0.559**
	German	0.5005	**0.516**
	Spanish	0.5226	**0.527**
	Average	0.5576	**0.53**
$E5_{large,instruct}$ (LLM reranking)	English	0.6510	0.542
	German	**0.5031**	0.503
	Spanish	0.5206	0.516
	Average	0.5582	0.52
$E5_{large,instruct}$ (BM25 & cross-encoder reranking)	English	**0.6535**	0.554
	German	0.5026	0.515
	Spanish	**0.5229**	0.522
	Average	**0.5597**	**0.53**

approaches. The application of our cross-lingual curriculum learning strategy boosted the average mAP of $E5_{large,instruct}$ on the development set to 55.76%, a substantial improvement from the 53.82% achieved with direct fine-tuning (Table 2). Further enhancements were observed with reranking.

- The hybrid retrieval approach yielded the highest average mAP on the **validation** set (**55.97% mAP**).
- LLM-based reranking with `Gemma3-27B` also showed a slight improvement over the curriculum-only model on the **validation** set (**55.82% mAP**).

On the **test** set, both the curriculum-trained model without further reranking and our system (curriculum + BM25 + cross-encoder reranker) achieved an average **mAP** of **53%**. This indicates good generalization for the curriculum-trained model, though the slight gains from reranking observed on the validation set did not fully translate to the test set for all configurations.

6.4. Impact of Job Descriptions at Inference

A core component of our methodology is the enrichment of job titles with full descriptions, either sourced from ESCO or generated by an LLM. To explicitly validate the necessity of this step during inference, we conducted an ablation study on the validation set. We compared the performance of our final curriculum-trained model under two conditions: 1) the standard approach, where query titles are used to retrieve from a corpus of job title + description, and 2) a title-only approach, where query titles retrieve from a corpus containing only job titles, with descriptions omitted.

The results, presented in Table 4, confirm the significant benefit of including descriptions. The average mAP dropped from 55.76% to 53.18% when descriptions were removed from the inference corpus. This performance decrease underscores that the contextual information provided by the descriptions is crucial for the model to disambiguate similar or ambiguous job titles, justifying our data augmentation strategy for the inference phase.

Table 4
Effect of including job descriptions during inference on the validation set (mAP)

Model Configuration	Spanish	English	German	Average
Curriculum E5 (Title + Description)	0.5226	0.6497	0.5005	0.5576
Curriculum E5 (Title Only)	0.4992	0.6280	0.4683	0.5318

6.5. Computational Considerations

All experiments were conducted on a single NVIDIA A100 GPU. We analyzed the trade-off between retrieval quality and computational cost for our primary system configurations. The baseline curriculum-trained E5-large-instruct model offers the lowest latency, requiring only a single embedding pass for the query followed by a fast vector search. Our final system, which adds BM25 (negligible overhead) and a cross-encoder reranker, introduces higher latency by requiring 10 additional forward passes per query. This cost yielded a marginal mAP improvement (55.76% to 55.97% on validation), making the system suitable for applications where accuracy is paramount. The LLM-based reranking using Gemma3-27B was the most resource-intensive approach, incurring substantial latency and memory usage for a minimal performance gain on the validation set and a drop on the test set, proving impractical for this task.

6.6. Limitations

Despite the promising results, this study has certain limitations that should be acknowledged. Firstly, our data augmentation strategy relies on LLM-generated descriptions for inference, and while efforts were made to ensure contextual plausibility, the synthetic data may not fully capture the stylistic diversity or factual nuances of authentic ESCO or real-world job descriptions. Discrepancies in quality or representativeness could subtly influence model performance on job titles reliant on these synthetic contexts. Secondly, the system's evaluation is based on the English, Spanish, and German languages within the TalentCLEF dataset. Its generalization to entirely different languages or highly specialized job domains not well-represented in the training data requires further investigation; the observed lower performance for German titles across several models may hint at existing cross-lingual representation challenges. Thirdly, the current work did not incorporate specific mechanisms to assess or mitigate potential performance disparities across job titles that may be stereotypically associated with different gender groups; addressing this is an important consideration for ensuring fairness in practical applications. Finally, the optimal configuration of various components, such as curriculum learning parameters, hybrid retrieval weights, or the selected LLM prompts, was determined based on validation set performance. As is common in such iterative development, this process may have led to some degree of adaptation to the validation data's specific characteristics, as suggested by the slight variations in performance gains between the validation and test sets for some of our reranking configurations.

7. Conclusion and Future Work

Our participation in TalentCLEF 2025 Task A demonstrated the efficacy of a carefully designed curriculum learning strategy combined with data augmentation, hybrid retrieval, and cross-encoder reranking for multilingual job title matching. The phased introduction of complexity and cross-lingual signals enabled our model to achieve strong performance. For future work, we plan to explore more sophisticated hard negative mining techniques for both the embedding model and the reranker.

Further investigation and experimenting with larger, more capable LLMs for reranking (perhaps with more elaborate prompting) could also yield improvements. Additionally, we intend to improve the synthetic data generation process by ensuring that the descriptions generated by the LLM more accurately reflect the linguistic style, content scope, and structural properties inherent in authentic ESCO descriptions.

Declaration on Generative AI

During the preparation of this work, the authors used **Gemini 2.5 Pro** in order to: Grammar and spelling check. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the publication's content.

References

[1] L. Gasco, H. Fabregat, L. García-Sardiña, P. Estrella, D. Deniz, A. Rodrigo, R. Zbib, Overview of the TalentCLEF 2025 Shared Task: Skill and Job Title Intelligence for Human Capital Management, in: International Conference of the Cross-Language Evaluation Forum for European Languages, Springer, 2025.

[2] R. Zbib, L. A. Lacasa, F. Retyk, R. Poves, J. Aizpuru, H. Fabregat, V. Simkus, E. García-Casademont, Learning job titles similarity from noisy skill labels, 2023. URL: https://arxiv.org/abs/2207.00494. arXiv:2207.00494.

[3] N. Laosaengpha, T. Tativannarat, C. Piansaddhayanon, A. Rutherford, E. Chuangsuwanich, Learning job title representation from job description aggregation network, in: L.-W. Ku, A. Martins, V. Srikumar (Eds.), Findings of the Association for Computational Linguistics ACL 2024, Association for Computational Linguistics, Bangkok, Thailand and virtual meeting, 2024, pp. 1319–1329. URL: https://aclanthology.org/2024.findings-acl.77.

[4] M. Zhang, R. van der Goot, B. Plank, ESCOXLM-R: Multilingual taxonomy-driven pre-training for the job market domain, in: Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Association for Computational Linguistics, Toronto, Canada, 2023, pp. 11871–11890. URL: https://aclanthology.org/2023.acl-long.662.

[5] D. Deniz, F. Retyk, L. García-Sardiña, H. Fabregat, L. Gasco, R. Zbib, Combined unsupervised and contrastive learning for multilingual job recommendation, in: M. Kaya, T. Bogers, D. Graus, C. Johnson, J.-J. Decorte, T. D. Bie (Eds.), Proceedings of the 4th Workshop on Recommender Systems for Human Resources (RecSys in HR 2024), volume 3788 of *CEUR Workshop Proceedings*, CEUR-WS.org, Bari, Italy, 2024, pp. 1–8. URL: https://ceur-ws.org/Vol-3788/RecSysHR2024-paper_3.pdf.

[6] L. Gascó, F. M. Hermenegildo, G.-S. Laura, D. C. Daniel, P. Estrella, R. Alvaro, Z. Rabih, Talentclef 2025 corpus: Skill and job title intelligence for human capital management, 2025. URL: https://doi.org/10.5281/zenodo.15240844. doi:10.5281/zenodo.15240844.

[7] M. le Vrang, A. Papantoniou, E. Pauwels, P. Fannes, D. Vandensteen, J. De Smedt, Esco: Boosting job matching in europe with semantic interoperability, Computer 47 (2014) 57–64. doi:10.1109/MC.2014.283.

[8] A. Yang, A. Li, B. Yang, B. Zhang, B. Hui, B. Zheng, B. Yu, C. Gao, C. Huang, C. Lv, C. Zheng, D. Liu, F. Zhou, F. Huang, F. Hu, H. Ge, H. Wei, H. Lin, J. Tang, J. Yang, J. Tu, J. Zhang, J. Yang, J. Yang, J. Zhou, J. Zhou, J. Lin, K. Dang, K. Bao, K. Yang, L. Yu, L. Deng, M. Li, M. Xue, M. Li, P. Zhang, P. Wang, Q. Zhu, R. Men, R. Gao, S. Liu, S. Luo, T. Li, T. Tang, W. Yin, X. Ren, X. Wang, X. Zhang, X. Ren, Y. Fan, Y. Su, Y. Zhang, Y. Zhang, Y. Wan, Y. Liu, Z. Wang, Z. Cui, Z. Zhang, Z. Zhou, Z. Qiu, Qwen3 technical report, 2025. URL: https://arxiv.org/abs/2505.09388. arXiv:2505.09388.

[9] L. Wang, N. Yang, X. Huang, L. Yang, R. Majumder, F. Wei, Multilingual e5 text embeddings: A technical report, 2024. URL: https://arxiv.org/abs/2402.05672. arXiv:2402.05672.

[10] M. Günther, J. Ong, I. Mohr, A. Abdessalem, T. Abel, M. K. Akram, S. Guzman, G. Mastrapas, S. Sturua, B. Wang, M. Werk, N. Wang, H. Xiao, Jina embeddings 2: 8192-token general-purpose text embeddings for long documents, 2023. arXiv:2310.19923.

[11] N. Reimers, I. Gurevych, Sentence-bert: Sentence embeddings using siamese bert-networks, in: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics, 2019. URL: http://arxiv.org/abs/1908.10084.

[12] J. Chen, S. Xiao, P. Zhang, K. Luo, D. Lian, Z. Liu, M3-embedding: Multi-linguality, multi-functionality, multi-granularity text embeddings through self-knowledge distillation, in: L.-W. Ku, A. Martins, V. Srikumar (Eds.), Findings of the Association for Computational Linguistics: ACL 2024, Association for Computational Linguistics, Bangkok, Thailand, 2024, pp. 2318–2335. URL: https://aclanthology.org/2024.findings-acl.137/. doi:10.18653/v1/2024.findings-acl.137.

[13] F. Feng, Y. Yang, D. Cer, N. Arivazhagan, W. Wang, Language-agnostic BERT sentence embedding, in: S. Muresan, P. Nakov, A. Villavicencio (Eds.), Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Association for Computational

Linguistics, Dublin, Ireland, 2022, pp. 878–891. URL: https://aclanthology.org/2022.acl-long.62/. doi:10.18653/v1/2022.acl-long.62.

[14] S. Sturua, I. Mohr, M. K. Akram, M. Günther, B. Wang, M. Krimmel, F. Wang, G. Mastrapas, A. Koukounas, N. Wang, H. Xiao, jina-embeddings-v3: Multilingual embeddings with task lora, 2024. URL: https://arxiv.org/abs/2409.10173. arXiv:2409.10173.

[15] G. Team, A. Kamath, J. Ferret, S. Pathak, N. Vieillard, R. Merhej, S. Perrin, T. Matejovicova, A. Ramé, M. Rivière, L. Rouillard, T. Mesnard, G. Cideron, J. bastien Grill, S. Ramos, E. Yvinec, M. Casbon, E. Pot, I. Penchev, G. Liu, F. Visin, K. Kenealy, L. Beyer, X. Zhai, A. Tsitsulin, R. Busa-Fekete, A. Feng, N. Sachdeva, B. Coleman, Y. Gao, B. Mustafa, I. Barr, E. Parisotto, D. Tian, M. Eyal, C. Cherry, J.-T. Peter, D. Sinopalnikov, S. Bhupatiraju, R. Agarwal, M. Kazemi, D. Malkin, R. Kumar, D. Vilar, I. Brusilovsky, J. Luo, A. Steiner, A. Friesen, A. Sharma, A. Sharma, A. M. Gilady, A. Goedeckemeyer, A. Saade, A. Feng, A. Kolesnikov, A. Bendebury, A. Abdagic, A. Vadi, A. György, A. S. Pinto, A. Das, A. Bapna, A. Miech, A. Yang, A. Paterson, A. Shenoy, A. Chakrabarti, B. Piot, B. Wu, B. Shahriari, B. Petrini, C. Chen, C. L. Lan, C. A. Choquette-Choo, C. Carey, C. Brick, D. Deutsch, D. Eisenbud, D. Cattle, D. Cheng, D. Paparas, D. S. Sreepathihalli, D. Reid, D. Tran, D. Zelle, E. Noland, E. Huizenga, E. Kharitonov, F. Liu, G. Amirkhanyan, G. Cameron, H. Hashemi, H. Klimczak-Plucińska, H. Singh, H. Mehta, H. T. Lehri, H. Hazimeh, I. Ballantyne, I. Szpektor, I. Nardini, J. Pouget-Abadie, J. Chan, J. Stanton, J. Wieting, J. Lai, J. Orbay, J. Fernandez, J. Newlan, J. yeong Ji, J. Singh, K. Black, K. Yu, K. Hui, K. Vodrahalli, K. Greff, L. Qiu, M. Valentine, M. Coelho, M. Ritter, M. Hoffman, M. Watson, M. Chaturvedi, M. Moynihan, M. Ma, N. Babar, N. Noy, N. Byrd, N. Roy, N. Momchev, N. Chauhan, N. Sachdeva, O. Bunyan, P. Botarda, P. Caron, P. K. Rubenstein, P. Culliton, P. Schmid, P. G. Sessa, P. Xu, P. Stanczyk, P. Tafti, R. Shivanna, R. Wu, R. Pan, R. Rokni, R. Willoughby, R. Vallu, R. Mullins, S. Jerome, S. Smoot, S. Girgin, S. Iqbal, S. Reddy, S. Sheth, S. Põder, S. Bhatnagar, S. R. Panyam, S. Eiger, S. Zhang, T. Liu, T. Yacovone, T. Liechty, U. Kalra, U. Evci, V. Misra, V. Roseberry, V. Feinberg, V. Kolesnikov, W. Han, W. Kwon, X. Chen, Y. Chow, Y. Zhu, Z. Wei, Z. Egyed, V. Cotruta, M. Giang, P. Kirk, A. Rao, K. Black, N. Babar, J. Lo, E. Moreira, L. G. Martins, O. Sanseviero, L. Gonzalez, Z. Gleicher, T. Warkentin, V. Mirrokni, E. Senter, E. Collins, J. Barral, Z. Ghahramani, R. Hadsell, Y. Matias, D. Sculley, S. Petrov, N. Fiedel, N. Shazeer, O. Vinyals, J. Dean, D. Hassabis, K. Kavukcuoglu, C. Farabet, E. Buchatskaya, J.-B. Alayrac, R. Anil, Dmitry, Lepikhin, S. Borgeaud, O. Bachem, A. Joulin, A. Andreev, C. Hardin, R. Dadashi, L. Hussenot, Gemma 3 technical report, 2025. URL: https://arxiv.org/abs/2503.19786. arXiv:2503.19786.