SkiLLens: Recognising and Mapping Novel Skills from Millions of Job Ads Across Europe Using Language Models

Alessia De Santo[1,2] **Lorenzo Malandri**[1,2]
Fabio Mercorio[1,2] **Mario Mezzanzanica**[1,2] **Navid Nobani**[1,2]
[1]Dept of Statistics and Quantitative Methods, University of Milano-Bicocca, Milano, Italy
[2]CRISP Research Centre, University of Milano-Bicocca, Milano, Italy
{name.surname}@unimib.it

Abstract

In a rapidly evolving labor market, detecting and addressing emerging skill needs is essential for shaping responsive education and workforce policies. Online job advertisements (OJAs) provide a real-time view of changing demands, but require first retrieving skill mentions from unstructured text and then solving the entity linking problem of connecting them to standardized skill taxonomies. To harness this potential, we present a multilingual human-in-the-loop (HITL) pipeline that operates in two steps: candidate skills are extracted from national OJA corpora using country-specific word embeddings, capturing terms that reflect each country's labor market. These candidates are linked to ESCO using an encoder-based system and refined through a decoder large language models (LLMs) for accurate contextual alignment. Our approach is validated through both quantitative and qualitative evaluations, demonstrating that our method enables timely, multilingual monitoring of emerging skills, supporting agile policy-making and targeted training initiatives.

1 Introduction and Contributions

The digitalization of workforce recruitment has led to a massive surge in OJAs, offering a near real-time lens into labor market dynamics across Europe. These data provide unique opportunities to analyze shifting occupational demands and, crucially, to detect the emergence of new skills. Timely identification of such skills is vital for aligning education and training systems with evolving industry needs. To harness this potential, Eurostat[1] has collected over 450 million OJAs from 2019 to 2024 across 27 EU countries and the UK. This effort supports real-time labor market monitoring using ESCO[2], the European multilingual classification of skills, competences, qualifications, and occupations taxonomy, which serves as a reference framework for cross-country comparisons. However, ESCO's static nature—despite its coverage of over 13,000 skills—makes it difficult to keep pace with the fast-changing landscape of emerging competencies.

Motivating Example. Emerging labor market terminology illustrates the limitations of manual taxonomy maintenance. For instance, job ads related to the digitalization of several fields increasingly mention skills such as *prompt engineering*, *MLOps monitoring*, or *cloud-native deployment*, none of which were present in ESCO when they first appeared. Identifying and validating such terms currently requires experts to manually review thousands of ads—a slow and resource-intensive process that cannot keep pace with rapidly evolving skill demands. This highlights the need for automated, multilingual methods that continuously detect and position new skills within existing taxonomies.

Addressing this gap is a growing priority for the European Commission, which emphasizes the urgency of tackling skill shortages, promoting upskilling and reskilling, and preparing Europe's workforce for rapid economic and technological transformation[3]. In this paper, we present SkiLLens, a multilingual pipeline developed within the PILLARS project[4] to detect and normalize emerging skills from OJAs. The pipeline has been applied to a large dataset comprising over 18 million job advertisements from 28 European countries and 23 languages. It follows a two-step methodology: (i) extracting candidate skill terms using a combination of word embeddings and LLMs, and (ii) refining and aligning these terms with ESCO through a recommendation-based process leveraging multiple LLMs. This approach effectively captures emerging skills across languages and facilitates their integration into existing taxonomies, enabling improved labor market intelligence.

[1] https://ec.europa.eu/eurostat
[2] https://esco.ec.europa.eu/en/about-esco/what-esco
[3] https://commission.europa.eu/topics/eu-competitiveness/union-skills_en
[4] https://www.h2020-pillars.eu/

Proceedings of the 19th Conference of the European Chapter of the Association for Computational Linguistics
Volume 5: Industry Track, pages 877–885

This work makes a **threefold contribution** to the study of novel skill extraction and mapping from OJAs:

1. **Extraction of Novel Skills:** We propose a method to automatically extract *novel and emerging skills* from unstructured OJAs using embeddings and LLMs. We define *novel skills* broadly: they may reflect either (i) **temporally emerging skills**, which are new in the labor market, or (ii) **conceptually novel expressions**, which are new phrasings or specifications of existing skills. This dual perspective ensures that both genuinely new competencies and evolving articulations of existing practices are captured.Details of how this is put in practice are presented in Section 3.
2. **Skill Normalisation and ESCO Mapping:** We address the challenge of aligning extracted skills to the ESCO taxonomy, whose skills pillar covers knowledge, abilities, and attitudes across different reuse levels. Given our multilingual, unstructured dataset (27 EU countries + UK), this alignment is non-trivial and treated in depth.
3. **Large-scale Application:** We showcase the robustness and scalability of the pipeline on a large, multilingual corpus of European OJAs.

Overall, SkiLLens provides a robust foundation for tracking the evolution of skill requirements, supporting public institutions, employment services, and education providers in adapting to a rapidly changing labor market. Figure 1 depicts the overall process.

2 Background and Related Work

In this section, we discuss the core methods of the SkiLLens framework in detail, starting with the extraction of skills from OJAs and followed by their normalisation into ESCO taxonomy.

2.1 Skills Extraction

Skill Extraction (SE) can be viewed as a specialised application of Information Extraction (IE) in the labor market domain. IE is the process of identifying and extracting structured information of predefined types from unstructured natural language texts (Mooney and Bunescu, 2005).

Formally, let $\mathcal{T} = \{t_1, \ldots, t_n\}$ be the set of texts, $\mathcal{P} = \{p_1, \ldots, p_m\}$ the set of information types, and $\mathcal{A}$ the universe of atomic items. Let $\tau : \mathcal{A} \to \mathcal{P}$ assign each item its type.

$$\begin{aligned} &\mathrm{IE} : \mathcal{T} \times \mathcal{P} \to 2^{\mathcal{A}}, \\ &\mathrm{IE}(t, p) = \{\, a \in \mathcal{A} \mid \mathrm{occ}(a, t) \wedge \tau(a) = p \,\}. \end{aligned} \tag{1}$$

$$\mathrm{occ}(a, t) \overset{\text{def}}{\iff} \exists\, 1 \le i \le j \le |t| \text{ s.t. } t_i \ldots t_j = a. \tag{2}$$

For a fixed pair (t_i, p_j) the output is a finite set

$$\mathrm{IE}(t_i, p_j) = \{a_1, \ldots, a_{k_{ij}}\}. \tag{3}$$

Drawing from this formalization, we define *Skill Extraction (SE)* as the specific IE task of identifying and extracting skill entities from unstructured textual data commonly found in the labor market—such as job advertisements, resumes, or job descriptions. Let $\mathcal{T}$ be the set of unstructured text documents (e.g., job ads, résumés), $\mathcal{P}$ the set of information types, and $\mathcal{S}$ the universe of canonical skill entities. For a document $t_i \in \mathcal{T}$ and a target type fixed to skills, $p_s = \textsc{skill} \in \mathcal{P}$, the goal is to return all skill entities present in t_i. We define the skill–extraction function as

$$\mathrm{SE} : \ \mathcal{T} \times \{\textsc{skill}\} \to 2^{\mathcal{S}}, \qquad (t_i, p_s) \mapsto S(t_i, p_s), \tag{4}$$

which yields a finite set of recognised skills

$$\mathrm{SE}(t_i) = S(t_i, p_s) \ = \ \{\, s \in \mathcal{S} \mid s \preceq t_i \,\}, \tag{5}$$

where $s \preceq t_i$ denotes that at least one span in t_i realises the skill entity s. The SE process returns a structured set $\mathcal{S}$ containing all skill instances found in the input text. Early approaches to skill extraction relied on exact matching, combining manual annotation with semantic clustering (e.g., Word2Vec) and ontology construction to classify skills from job-related texts (Calanca et al., 2019; Gugnani et al., 2018; Javed et al., 2017). To handle variability in skill terminology, fuzzy matching techniques were introduced, comparing extracted phrases with controlled vocabularies like ESCO using similarity metrics such as Levenshtein Distance and Jaccard Similarity (Boselli et al., 2018). Unsupervised topic modelling, particularly Latent Dirichlet Allocation (LDA), has been used to uncover latent skill structures by applying it directly to job descriptions or using a domain-specific vocabulary (Gurcan and Cagiltay, 2019; De Mauro et al., 2018). Deep learning further advanced the field by treating skill extraction as sequence tagging or multi-label classification, using convolutional networks and ranking-based methods (Li

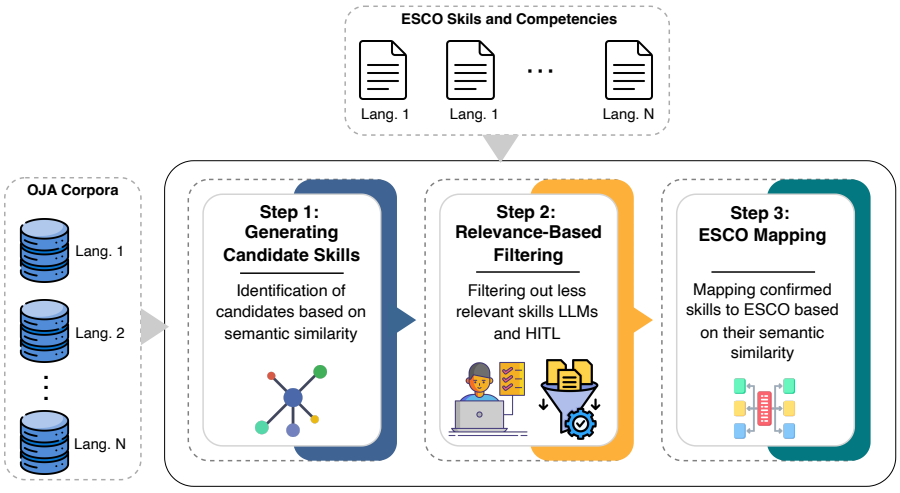

Figure 1: Overview of the SkiLLens framework for extracting and mapping novel skill expressions from OJAs across 28 countries.

et al., 2020; Jiechieu and Tsopze, 2021; Goyal et al., 2023). Recently, transformer-based models have shown promising results by leveraging contextual embeddings, typically by fine-tuning BERT or SpanBERT with classification layers (e.g., CRFs) or adapting pretrained models for domain-specific recruitment tasks (Fang et al., 2023; Zhang et al., 2022b; Bhola et al., 2020; Barducci et al., 2022). However, to our knowledge, no existing work has performed skill extraction across such a diverse set of languages, nor systematically assessed the novelty of extracted skills. To fill these gaps, we employ a rigorous expert evaluation involving human experts, specialized in different labor market contexts. This ensures extracted skills are both contextually relevant and genuinely novel.

2.2 Skills Normalisation

Skill Normalisation (SN) can be viewed as a specialized application of Semantic Retrieval (SR) in the labor market domain. SR is the process of retrieving information based on semantic similarity, typically by representing textual data as embeddings within a multidimensional vector space.

Semantic Retrieval (SR). Let $\mathcal{Q} = \{q_1, \ldots, q_n\}$ be a set of queries and $\mathcal{E} = \{e_1, \ldots, e_m\}$ a set of target elements. Assume each item is represented by an embedding, and let $\mathrm{sim}(\cdot, \cdot)$ be a similarity function (e.g., cosine). For each $q_i \in \mathcal{Q}$, SR returns the most similar target:

$$\mathrm{SR}(q_i; \mathcal{E}) = \arg\max_{e \in \mathcal{E}} \mathrm{sim}(q_i, e). \quad (6)$$

Skills Normalisation (SN). Let $\mathcal{S} = \{s_1, \ldots, s_k\}$ be the set of skill mentions extracted from OJAs, and let $\mathcal{E}_{\text{ESCO}}$ denote the set of canonical skills in ESCO (each with a preferred label), represented as embeddings. SN maps each extracted mention to its canonical ESCO entry:

$$\mathrm{SN}(s_i; \mathcal{E}_{\text{ESCO}}) = \arg\max_{e \in \mathcal{E}_{\text{ESCO}}} \mathrm{sim}(s_i, e) = \hat{s}_i \in \mathcal{E}_{\text{ESCO}}. \quad (7)$$

Aggregating over all mentions gives the normalised set

$$\mathrm{SN}(\mathcal{S}; \mathcal{E}_{\text{ESCO}}) = \{\hat{s}_1, \ldots, \hat{s}_k\} \subseteq \mathcal{E}_{\text{ESCO}}, \quad (8)$$

where each $\hat{s}_i$ corresponds to the ESCO entry whose *preferred label* is the closest in meaning to the extracted mention s_i.

Few works address mapping skills to ESCO. SkillNER (Fareri et al., 2021) extracts soft skills from texts but does not normalize them to ESCO entries. Kompetencer (Zhang et al., 2022a) categorizes skills into 23 ESCO-aligned groups without using the ESCO skills for the mapping. The same dataset is used to evaluate ESCOXLM-R (Zhang et al., 2023), a pretrained language model for skills extraction and classification. SkillGPT (Li et al., 2023) employs a language model for skill extraction and standardization but lacks empirical validation. Closest to our work, (Decorte et al., 2022) and (Clavié and Soulié, 2023) propose mapping skills to ESCO, the latter using a two-step zero-shot pipeline. In Sec.4 we compare SkiLLens against the last two cited approaches and demonstrate our superior performance.

Table 1: Comparison of Skill Normalization Approaches. (Repr.: Reproducibility, ESCO:Mapped to ESCO?, D: Data, C: Code).

Framework	Mapping Approach	Lang.	ESCO	Used?	Repr.
Kompetencer	Rule-based class. to 23 cats.	en, da	No	No	✓ D ✓ C
ESCOXLM-R	Pretrained LM	en, fr..	No	No	✓ D ✓ C
SkillGPT	LLM + vector search	en, fr, nl	Yes	No	✗ D ✓ C
Decorte et al.	Extraction + map	en	Yes	Yes	✓ D ✓ C
Clavie et al.	Zero-shot LLM	en	Yes	Yes	✓ D ✓ C

3 SkiLLens: Method and implementation

In this section we go through three steps of SkiLLens and for each step provide a description of its role and the way we implement it.

Dataset This study uses online job ads (OJAs) from the Web Intelligence Hub (WIH)[5], part of Eurostat's Trusted Smart Statistics framework. The WIH-OJA initiative, developed by Eurostat and Cedefop, aggregates OJAs from 32 countries (EU, EEA, and the UK), totalling over 450M unique postings. We rely on the curated NLP sample v3 (r20240226), which contains 69.5M ads from 2018–2023, stratified by language and seven metadata dimensions (ISCO-08 occupation, contract type, salary, working hours, education, NACE division, and experience). For this work, we extract up to 1M of the most recent ads per country (or all available), resulting in a multilingual dataset of about 18M postings covering 28 European countries[6]. Large Member States generally reach the 1M cap, while smaller ones (e.g., Malta, Cyprus, Estonia) contribute substantially fewer ads.

Step 1: Extraction of Candidate Skills

This phase identifies candidate new skills from national OJA corpora by comparing the embeddings of ESCO skills and tokens of OJAs.

Novel skills extraction via word embeddings.

Input: Raw OJA texts from 28 countries.
Output: List of candidate novel skills.

Example

> *Input*: We're seeking a **detail-oriented** Data Scientist with strong **quantitative skills**... You'll play a pivotal role in **analysing large volumes of data** using **Python, R, or Java**...
>
> *Output*: {attention to detail, mathematical aptitude, Python, R, Java, machine learning, SQL...}

[5] https://cros.ec.europa.eu/wih

[6] Access provided via the PILLARS project consortium; external access requires Eurostat approval and an NDA.

Implementation To extract candidate skills, we train FastText word embeddings (Bojanowski et al., 2017) on country-specific job ad corpora. Following (Giabelli et al., 2020), we perform a grid search across 160 model configurations. The best setup—based on the Hierarchical Semantic Similarity (HSS) metric (Malandri et al., 2020; Giabelli et al., 2022)—uses Skip-Gram with 100 dimensions, 5 epochs, and a learning rate of 0.01.Each extracted skill is compared to all ESCO skills (preferred and alternative labels, totaling ~98,000 entries) using cosine similarity, retaining the top-5 closest matches. Alternative labels can be synonyms, spelling variants, declensions, abbreviations, or other expressions commonly used by jobseekers, employers, and education institutions to refer to the concept described by the preferred ESCO term. Because these alternatives already include recombinations and lexical variations, doing the filtering with their inclusion ensures that candidate skills that are merely order recombinations of existing terms are not erroneously considered novel. To further ensure novelty and avoid near-duplicates, we filter out candidates with a fuzzy match score ≤ 70 (using `fuzz.ratio` from the `rapidfuzz` library), based on normalized Levenshtein distance, guaranteeing lexical distinction from existing ESCO entries.

Step 2: Filter Candidates by Relevance

Input: List of candidate novel skills from Step 1.
Output: List of relevant and validated candidate novel skills.

First, we filter candidates using a score that combines semantic similarity (via word embeddings) with corpus frequency (Giabelli et al., 2020). Each candidate c is scored against an ESCO skill s using:

$$S(c,s) = \alpha \cdot cos\_sim(c,s) + (1-\alpha) \cdot freq(c) \quad (9)$$

where $cos\_sim(c,s)$ is their cosine similarity in the word embedding model generated in Section 3 and $freq(c)$ is the term frequency in the corpus. We compute scores for the top-k most similar terms and retain those covering 95% of cumulative frequency, filtering out rare terms. We adopt a two-step validation process to ensure the relevance of extracted novel skill candidates before mapping them to ESCO. First, we use GPT-4 to automatically assess whether each candidate represents a valid skill, guided by a prompt with definitions, examples, and conservative filtering rules (e.g. in cases of uncertainty, the model had to flag terms

as potentially relevant rather than discard them). This step is necessary due to the high volume of candidates (5,000 per country), with the model instructed to favor recall over precision. Despite a precision of 70%, spot checks in Italy and the UK showed recall above 98%, confirming minimal loss of valid skills. Second, labor market experts from each country review the filtered list to validate the contextual relevance and novelty of each term. This human-in-the-loop step ensures that the automatically extracted skills are not only linguistically valid but also meaningful within each national labor market context. Experts assess whether each candidate term represents a genuine skill (as opposed to an occupation or experience), and whether it introduces new or emerging terminology that extends the existing skill taxonomy. Table 2 illustrates a subset of this validation process, comparing model and expert judgments for selected examples.

Table 2: Expert (Exp.) validation examples.

Skill	LLM	LLM Motivation	Exp.	Exp. Motivation
Interaction with security personnel	✔	Social-communication competence	✖	Seen as "experience," not a transferable skill.
Programming	✔	Core technical ability	✖	Too broad; should be split into specific tasks.
Prompt engineering	✔	Emerging AI-related competence	✔	Confirmed as novel and relevant.

Implementation Following suggestion of Giabelli et al. (2020), the weighting parameter in Equation 3 was set to $\alpha = 0.85$ to prioritize semantic closeness over frequency. This relevance score is applied after the syntactic filtering step described in Section 3, ensuring that retained candidates are both novel and contextually meaningful within the domain. To evaluate new skill candidates, we use an GPT4-based validation step. For classification-like tasks, carefully designed prompts improve consistency and reliability. Our approach includes:

Contextual framing: Prompts contain (i) a definition of *skill* aligned with ESCO and related work, (ii) response format instructions (starting with "yes" or "no"), and (iii) an OJA example showing the candidate term in context.

Few-shot learning: Queries are preceded by example prompts with five candidate terms and model responses. The model decides for each term ("yes"/"no") and gives a brief justification, improving accuracy and coherence. This LLM-based filtering reduces noise before expert review. Consequently, we assess the quality of extracted skills, involving labor market experts engaged through the European Network of Regional Labour Market Monitoring [7]. Experts judged the relevance and formulation of candidate skills based on original job ads in their respective languages. Thanks to the previous filtering they evaluated up to 400 candidates each. Out of 4,941 proposed novel skills, 3,552 (71.9%) were validated as relevant. However, precision varied across countries due to factors such as language quality, corpus coverage, and expert interpretation. Notably, DE, CY, and NL excluded ESCO knowledge concepts[8] from validation, which impacted their scores. Most countries exceed the global average (73%), confirming overall robustness.

Step 3: ESCO Mapping

By examining the semantic similarity among candidate skills and ESCO skills, this stage positions novel skill expressions within the most appropriate locations in the ESCO taxonomy. The process ensures standardisation and facilitates the enrichment of ESCO with relevant emerging skills.

Input: Validated list of candidate novel skills.
Output: List of n recommended mappings for each candidate novel skill.

In order to find the best embedding model and given the cross-lingual nature of our corpus, we tested both multilingual sentence embeddings and an alternative approach where all texts were translated into English and then encoded. The latter proved more effective. Embedding models were selected based on the MTEB (Massive Text Embedding Benchmark) leaderboard (Muennighoff et al., 2023), which compares over 50 models across tasks such as Semantic Textual Similarity and also considers encoding time and dimensionality. For translation, we used the DeepSeek v3. Each new skill is linked to its top 3 most similar ESCO skills according to cosine similarity. To create a multilingual benchmark, all ESCO skill labels (13,939 preferred terms) and their alternative labels are extracted across the 22 languages represented in the novel skills corpus. The benchmark task is to correctly associate each alternative label with its preferred label. Since these novel skills are not yet included in

[7] https://www.regionallabourmarketmonitoring.net
[8] https://esco.ec.europa.eu/en/classification/skill_main

ESCO, the benchmark serves as a proxy to assess our method's ability to accurately position out-of-taxonomy terms within the existing skill hierarchy. To chose the best alternative, we evaluated three language models on the english dataset, including both open and close weight, i.e, GPT4, Gemini 2.0 flash, and Mixtral-8x22B. The best performances are reached by GPT4, which we chose for the final step. Table 4 presents the empirical results for the best match selection. An example of the employed prompt is showed in listing 1.

Listing 1: Prompt Template for Best Match Selection.

```
messages=[{"role":"user","content":(
    f"Select the most similar term to: {candidate}. "+
    "Choose from these three (term + description): "+
    f"1. {alt[0]} - {desc[0]} "+
    f"2. {alt[1]} - {desc[1]} "+
    f"3. {alt[2]} - {desc[2]} "+
    "Reply with only the term. Always pick one of the three, even if unsure."
)}]
```

Note: `candidate` is the extracted skill; `alt` and `desc` contain the ESCO label options and their descriptions.

4 Evaluation

This section presents our results across two evaluations for the task of mapping to ESCO: (i) comparison with state-of-the-art methods, and (ii) benchmarking against ESCO preferred labels, highlighting the effectiveness of our approach in skill identification and categorization.

Comparison Against State-of-the-Art. We benchmarked our approach against two English-only methods: Decorte et al. (Decorte et al., 2022) and Clavié and Soulié (Clavié and Soulié, 2023), replicating their task of mapping soft skills to ESCO. Unlike their broader classification focus, our work targets novel skills only. Table 3 shows that our method outperforms both baselines on the 'tech' and 'house' datasets across all metrics (RP@1/5/10, MRR).

Table 3: Comparison of Skill Mapping Performance.

Method	RP@1	RP@5	RP@10	MRR
Tech Dataset				
Decorte et al. (2022)	n/a	0.317	0.392	0.339
Clavié & Soulié (2023)	0.465	0.615	0.689	0.537
SkiLLens (ours)	**0.633**	**0.856**	**0.901**	**0.731**
House Dataset				
Decorte et al. (2022)	n/a	0.308	0.387	0.299
Clavié & Soulié (2023)	0.630	0.567	0.610	0.507
SkiLLens (ours)	**0.639**	**0.823**	**0.904**	**0.727**

Baseline Evaluation To assess SkiLLens's ability to enrich ESCO, we created a multilingual baseline by mapping ESCO's "alternative labels" to their "preferred labels" —focusing on ESCO's most granular (4th) level. This allows us to test semantic matching while preserving taxonomy structure[9]. We randomly sampled 200 alternative labels per language, translated them into English using Deepseek v3, and encoded them via sentence embeddings, which were normalized prior to similarity search. We then retrieved the top-3 most similar preferred labels using cosine similarity, employing ChromaDB[10] for efficient vector storage and querying. We tested several top MTEB[11] models and found `thenlper/gte-large` (335M params) yielded the best results. A final selection was made using GPT-4 via ChatGPT, prompted to choose the best match among the top three. The accuracy of this LLM-enhanced match was then compared with ESCO ground truth across 22 languages. Tab. 4 shows strong performance across metrics (Acc@1, Acc@3, NDCG@3, LLM accuracy - called Refinement in Tab. 4), confirming SkiLLens's effectiveness in taxonomy enrichment.

Table 4: Performance Metrics (%) across languages.

Lang	Retrieval A@3	Retrieval N@3	Retrieval A@1	Refinement LLM
bg	78.50	77.46	68.00	**68.84**
cs	87.62	86.70	77.23	**77.39**
da	80.00	79.29	64.50	**68.84**
de	82.50	81.58	73.50	**73.50**
el	77.23	75.22	63.37	60.10
en	89.00	87.15	83.50	**84.50**
es	84.58	83.39	72.14	**73.00**
et	59.90	57.92	46.53	**50.00**
fi	78.50	77.09	64.00	**65.83**
fr	84.08	82.52	71.14	**73.00**
hr	90.50	88.67	78.00	**80.71**
hu	77.00	76.51	62.50	**71.21**
it	80.00	78.07	65.50	**70.71**
lt	75.74	75.76	61.39	**63.00**
lv	81.00	79.89	70.00	**72.50**
nl	84.00	82.25	70.00	**70.35**
pl	77.61	75.65	65.67	**67.00**
pt	85.50	84.74	72.00	**73.74**
ro	62.69	60.79	46.77	**54.50**
sk	74.00	70.90	57.50	56.78
sl	84.50	82.94	72.50	**73.37**
sv	82.00	81.82	73.00	**74.00**

Note: Bold LLM values indicate improvement over embeddings.

[9] Alternative labels include synonyms, variants, or abbreviations used in real-world job ads.

[10] https://www.trychroma.com/

[11] We employed MTEB english version 2:https://github.com/embeddings-benchmark/mteb

5 Conclusion

In this paper, we introduced SkiLLens, a multilingual pipeline for extracting and mapping novel skills from over 18 million job ads across 27+1 European countries. Our methodology combines embedding-based extraction with LLM-assisted validation and mapping into the ESCO taxonomy. The results demonstrate strong alignment with expert assessments, with over 70% of extracted skills recognized as valid by national labor market experts. Regarding mapping, our method outperforms existing baselines in all the metrics. However, performance varies across languages, highlighting areas for improvement in low-resource or morphologically rich contexts. Future work will focus on integrating feedback loops for taxonomy enrichment and improving cross-lingual alignment through domain-adaptive fine-tuning and prompt optimization.

Limitations

While SkiLLens provides a robust foundation for multilingual skill discovery and mapping, we recognize few aspects that present opportunities for further refinement and exploration.

- **Characteristics of the OJA Data Source:** Our use of Online Job Advertisements (OJAs) offers a valuable, real-time view of the labor market. We acknowledge that this data source may not capture all sectors of the economy equally. These characteristics are well-documented in recent literature, such as studies by Cedefop (Napierala and Branka, 2022) and the OECD (Tsvetkova et al., 2024). These works confirm that while biases exist, OJAs are an increasingly representative and valid source for labor market analysis, particularly when researchers account for their specific properties. In this spirit, a promising direction for our work is to enrich the framework by incorporating alternative data sources for an even more holistic perspective.
- **Nuances of Cross-Lingual Analysis:** The decision to use a pivot-translation approach was a pragmatic one that enabled effective cross-lingual comparison. We acknowledge this involves a trade-off, as some language-specific nuances may be smoothed in the process. For future iterations, we are keen to explore fine-tuning native multilingual models directly on domain-specific corpora, which could further enhance performance, especially for morphologically rich and lower-resource languages.
- **The Inherent Complexity of Skill Definition:** Integrating expert knowledge is a core strength of our methodology. At the same time, defining a "skill" and assessing its "novelty" is an inherently complex task where ambiguity can arise, particularly at the boundaries of knowledge, tasks, and experience. This reflects a broader challenge in the field. A valuable next step would be to develop a framework for measuring inter-annotator agreement across languages, which would help quantify and better understand the nuances of expert judgment.

Acknowledgments

This paper is partially supported within the research activity of a grant entitled "PILLARS — Pathways to Inclusive Labour Markets" - under the call H-2020 TRANSFORMATIONS 18-2020 "Technological transformations, skills, and globalization - future challenges for shared prosperity", grant agreement NUMBER 101004703 — PILLARS.

References

Alessandro Barducci, Simone Iannaccone, Valerio La Gatta, Vincenzo Moscato, Giancarlo Sperlì, and Sergio Zavota. 2022. An end-to-end framework for information extraction from italian resumes. *Expert Systems with Applications*, 210:118487.

Akshay Bhola, Kishaloy Halder, Animesh Prasad, and Min-Yen Kan. 2020. Retrieving skills from job descriptions: A language model based extreme multi-label classification framework. In *Proceedings of the 28th international conference on computational linguistics*, pages 5832–5842.

Piotr Bojanowski, Edouard Grave, Armand Joulin, and Tomas Mikolov. 2017. Enriching word vectors with subword information. *Transactions of the association for computational linguistics*, 5:135–146.

Roberto Boselli, Mirko Cesarini, Fabio Mercorio, and Mario Mezzanzanica. 2018. Classifying online job advertisements through machine learning. *Future Generation Computer Systems*, 86:319–328.

Federica Calanca, Luiza Sayfullina, Lara Minkus, Claudia Wagner, and Eric Malmi. 2019. Responsible team players wanted: an analysis of soft skill requirements in job advertisements. *EPJ Data Science*, 8(1):1–20.

Benjamin Clavié and Guillaume Soulié. 2023. Large language models as batteries-included zero-shot esco skills matchers.

Andrea De Mauro, Marco Greco, Michele Grimaldi, and Paavo Ritala. 2018. Human resources for big data professions: A systematic classification of job roles and required skill sets. *Information Processing & Management*, 54(5):807–817.

Jens-Joris Decorte, Jeroen Van Hautte, Johannes Deleu, Chris Develder, and Thomas Demeester. 2022. Design of negative sampling strategies for distantly supervised skill extraction. In *RecSys in HR*, volume 3218. CEUR.

Chuyu Fang, Chuan Qin, Qi Zhang, Kaichun Yao, Jingshuai Zhang, Hengshu Zhu, Fuzhen Zhuang, and Hui Xiong. 2023. Recruitpro: A pretrained language model with skill-aware prompt learning for intelligent recruitment. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 3991–4002.

Silvia Fareri, Nicola Melluso, Filippo Chiarello, and Gualtiero Fantoni. 2021. Skillner: Mining and mapping soft skills from any text. *Expert Systems with Applications*, 184:115544.

Anna Giabelli, Lorenzo Malandri, Fabio Mercorio, Mario Mezzanzanica, and Navid Nobani. 2022. Embeddings evaluation using a novel measure of semantic similarity. *Cognitive Computation*, 14(2):749–763.

Anna Giabelli, Lorenzo Malandri, Fabio Mercorio, Mario Mezzanzanica, and Andrea Seveso. 2020. Neo: A tool for taxonomy enrichment with new emerging occupations. In *International Semantic Web Conference*, pages 568–584. Springer.

Nidhi Goyal, Jushaan Kalra, Charu Sharma, Raghava Mutharaju, Niharika Sachdeva, and Ponnurangam Kumaraguru. 2023. Jobxmlc: Extreme multi-label classification of job skills with graph neural networks. In *Findings of the Association for Computational Linguistics: EACL 2023*, pages 2181–2191.

Akshay Gugnani, Vinay Kumar Reddy Kasireddy, and Karthikeyan Ponnalagu. 2018. Generating unified candidate skill graph for career path recommendation. In *2018 IEEE International Conference on Data Mining Workshops (ICDMW)*, pages 328–333. IEEE.

Fatih Gurcan and Nergiz Ercil Cagiltay. 2019. Big data software engineering: Analysis of knowledge domains and skill sets using lda-based topic modeling. *IEEE access*, 7:82541–82552.

Faizan Javed, Phuong Hoang, Thomas Mahoney, and Matt McNair. 2017. Large-scale occupational skills normalization for online recruitment. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 31, pages 4627–4634.

Kameni Florentin Flambeau Jiechieu and Norbert Tsopze. 2021. Skills prediction based on multi-label resume classification using cnn with model predictions explanation. *Neural Computing and Applications*, 33(10):5069–5087.

Nan Li, Bo Kang, and Tijl De Bie. 2023. Skillgpt: a restful api service for skill extraction and standardization using a large language model. *arXiv preprint arXiv:2304.11060*.

Shan Li, Baoxu Shi, Jaewon Yang, Ji Yan, Shuai Wang, Fei Chen, and Qi He. 2020. Deep job understanding at linkedin. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 2145–2148.

Lorenzo Malandri, Fabio Mercorio, Mario Mezzanzanica, and Navid Nobani. 2020. Meet: A method for embeddings evaluation for taxonomic data. In *2020 International Conference on Data Mining Workshops (ICDMW)*, pages 31–38. IEEE.

Raymond J Mooney and Razvan Bunescu. 2005. Mining knowledge from text using information extraction. *ACM SIGKDD explorations newsletter*, 7(1):3–10.

Niklas Muennighoff, Nouamane Tazi, Loic Magne, and Nils Reimers. 2023. Mteb: Massive text embedding benchmark. In *Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics*, pages 2014–2037.

V. Napierala, J.; Kvetan and J. Branka. 2022. *Assessing the representativeness of online job advertisements*. Publications Office of the European Union.

A. Tsvetkova and 1 others. 2024. How well do online job postings match national sources in large english speaking countries?: Benchmarking lightcast data against statistical sources across regions, sectors and occupations. OECD Local Economic and Employment Development (LEED) Papers 2024/01, OECD Publishing, Paris.

Mike Zhang, Kristian Nørgaard Jensen, and Barbara Plank. 2022a. Kompetencer: Fine-grained skill classification in danish job postings via distant supervision and transfer learning. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pages 436–447.

Mike Zhang, Kristian Nørgaard Jensen, Sif Dam Sonniks, and Barbara Plank. 2022b. Skillspan: Hard and soft skill extraction from english job postings. *arXiv preprint arXiv:2204.12811*.

Mike Zhang, Rob van der Goot, and Barbara Plank. 2023. Escoxlm-r: Multilingual taxonomy-driven pretraining for the job market domain. In *The 61st Annual Meeting of the Association for Computational Linguistics*, pages 11871–11890. Association for Computational Linguistics.