DEPARTMENT OF INTERNATIONAL AND EUROPEAN ECONOMIC STUDIES

ATHENS UNIVERSITY OF ECONOMICS AND BUSINESS

Skill Mapping with AI: A Scalable Approach to Policy Analysis and Decision-Making

Phoebe Koundouri

Conrad Landis

Georgios Feretzakis

Working Paper Series

25-20

February 2025

Skill Mapping with AI: A Scalable Approach to Policy Analysis and Decision-Making

Phoebe Koundouri [1] *, Conrad Landis [2], Georgios Feretzakis [2]

1) School of Economics and ReSEES Research Laboratory, Athens University of Economics and Business; Department of Technology, Management and Economics, Denmark Technical University (DTU); Sustainable Development Unit, Athena Research Centre; UN SDSN (Global Climate Hub, European Hub, Greek Hub).

2) ReSEES Research Laboratory, Athens University of Economics and Business; Sustainable Development Unit, Athena Research Centre; School of Science and Technology, Hellenic Open University, Athens, Greece.

** Correspondence: pkoundouri@aueb.gr*

Abstract

In an era of rapidly evolving labor market demands, aligning policy documents with the skills required for emerging and existing occupations has become increasingly critical. This paper presents a scalable, AI-driven framework for skill mapping that integrates advanced sentence-embedding models, FAISS for high-speed similarity searches, and the European Skills/Competences, Qualifications and Occupations (ESCO) classification. By automatically extracting and analyzing skill references in policy texts, the framework helps policymakers and analysts identify recurring competencies, detect emerging themes (e.g., sustainability or digital literacy), and pinpoint potential workforce gaps. Additionally, it introduces a systematic method for assessing occupation-level relevance—calculating the overlap between policy-cited skills and ESCO-defined occupations to guide targeted upskilling and reskilling efforts. Empirical results suggest that this AI-enabled approach can markedly enhance both the speed and accuracy of policy analysis compared to traditional manual reviews, ultimately supporting data-driven decision-making at scale.

1. Introduction

Global labor markets are changing at a rapid and unpredictable pace, fueled by technological innovation, globalization, and changing economic priorities. Recent studies confirm an increased focus on digital competency, sustainability, and flexible skill sets with adaptability in sectoral disruption (Kraus, 2022; World Economic Forum, 2023). In response, policymakers, educational leaders, and private-sector representatives have accelerated efforts to seek systemic approaches for balancing public programs and guidance with emerging workforce requirements (OECD, 2023;

European Commission, 2022a). Reconciliation is becoming increasingly important in sectors such as renewable energy, data science, and e-governance, in which skills must respond to rapid marketplace transformations (Rikala, 2024).

Traditional policy analysis—which often relies on manual document reviews, expert panels, and time-intensive interviews—faces distinct challenges in this new environment. Manual processes not only risk introducing subjective biases but also struggle to handle the volume and variety of policy documents produced across multiple domains (European Commission, 2022b). In response, researchers and institutional bodies have begun to explore data-driven tools, including artificial intelligence (AI) and natural language processing (NLP) techniques, to streamline policy assessment and skill forecasting (Rashid, 2024). By leveraging machine learning to identify relevant skills automatically, AI-based solutions can offer more accurate insights into workforce readiness and highlight critical skill gaps that must be addressed to remain competitive in the global economy (OECD, 2023; World Economic Forum, 2023).

The current paper introduces a reproducible, AI-driven methodology that extracts relevant skills from large policy text corpora using vector embedding models, FAISS (a fast similarity search library), and the European Skills/Competences, Qualifications and Occupations (ESCO) classification framework. Specifically, we harness the Sentence-BERT family of models to transform both policy text and ESCO-defined skills into vector representations, enabling rapid similarity-based matching. Through this approach, policymakers can obtain actionable feedback on the skills most frequently cited or implied in policy documents, facilitating timely adjustments to educational curricula, workforce development strategies, and regulatory interventions (European Commission, 2022a).

Beyond identifying skill sets relevant to an issue, this article sets out a mechanism for gauging occupational pertinence. By estimating skill intersection between occupations defined in terms of ESCO and policy documents, policymakers can calculate jobs most affected by, or most in tune with, a specific policy (OECD, 2023). By adding an additional level of analysis, such areas can then become apparent for focused upskilling and reskilling interventions, and for a more effective use of resources and a deeper coordination between legislative interventions and labour market realities (World Economic Forum, 2023). In pilot-testing out such an AI-facilitated skill mapping mechanism for scalability, accuracy, and ease of integration, the article both sets out a blue-print for future method and an empirical basis for future educational, workforce, and social planning-related policy decision-making options.

2. Background and Related Work

Recent years have seen a boom in studies dealing with skill extraction automation of textual information, in reaction to increased demand for flexible workforce development. Several studies have emphasized AI-facilitated approaches for efficiency and accuracy improvement in emerging competency determination, for recruitment, curricula planning, and policy evaluation (Liu, 2024; Babashahi et al., 2024; OECD, 2023). Most recently, work identifies AI-facilitated approaches can bridge between high-policy programs, such as European Green Deal, and Sustainable Development Goals (SDGs), providing a coherent framework for impact evaluation (Koundouri et al., 2022). Other studies apply similar machine learning approaches for discovering SDG information in human security policies (Koundouri et al., 2024a) and suggesting AI integration in energy infrastructure for attaining global climate objectives (Koundouri et al., 2024b). In this section, relevant work concerning skill extraction methodologies, analysis frameworks for policies, and European Commission's role in developing an ESCO taxonomy for harmonization of skill definitions is discussed.

2.1 AI in Skill Identification

Skill extraction from textual documents—ranging from job advertisements to policy briefs—has become a vibrant area of inquiry as organizations look to automate segments of the workforce planning process (World Economic Forum, 2023). Early techniques often relied on manually curated keyword lists or rudimentary rule-based systems, which proved limited when dealing with complex language or domain-specific nuances. In contrast, more recent research has emphasized embedding-based NLP models that capture semantic relationships in a contextual manner. These models transform text into vector representations—commonly referred to as embeddings—making it possible to measure similarity between phrases, sentences, or entire paragraphs in a more robust way than simple keyword matching. Such techniques have been shown to be particularly effective for multi-lingual or highly specialized policy texts, where vocabulary shifts and contextual subtleties can diminish the effectiveness of traditional rule-based methods. Recent advances in transformer-based architectures have further refined the ability to detect latent skill references, improving precision and recall (Supriyono, 2024).

2.2 Policy Analysis and Decision-Making

Policy documents typically address a wide array of social, economic, and technological issues, making it challenging for analysts to isolate the sections that directly pertain to workforce development. Historically, manual reviews and expert interviews dominated this space, but such methods tend to be labor-intensive and susceptible to reviewer bias. Driven by mounting evidence that data-oriented methods yield more consistent and quicker policy insights, AI-driven text analytics is increasingly adopted to augment, or in some cases replace, traditional qualitative approaches (OECD, 2023). Automated content analysis not only helps in timely policy revision but also in identifying gaps or overlaps in regulations. This is especially relevant in sectors like green energy, digital healthcare, and Industry 4.0, where emerging technologies and new forms of labor can rapidly become cornerstones of national and transnational policy (European Commission, 2022b). By systematically extracting skills from large policy corpora, governments and multinational bodies can conduct more precise and comparative evaluations, thereby enhancing the accountability and strategic coherence of policy interventions (Naeem, 2024; OECD, 2023).

2.3 The ESCO Classification

Given the complexity and heterogeneity of skill definitions across different regions and industries, standardized taxonomies are vital to maintain consistency in both research and practice. The European Skills/Competences, Qualifications and Occupations (ESCO) classification, developed by the European Commission, functions as a multi-lingual and continuously updated repository of skills, competences, and occupations (European Commission, 2022a). By providing an exhaustive list of preferred labels, alternative labels, and concise descriptions for each skill, ESCO reduces ambiguity and enables cross-country comparisons—an essential feature given the varying terminologies in EU Member States (European Commission, 2022b). Embedding ESCO into AI-driven skill extraction workflows confers immediate advantages, such as ensuring that identified skills adhere to a recognized standard and that subsequent analysis—whether linking skills to occupations or evaluating policy impact—remains transparent and interoperable. This standardization facilitates alignment with labor market data, supports the design of upskilling and reskilling initiatives, and fosters a shared framework for dialogue among policymakers, educators, and industry leaders (OECD, 2023).

3. Methodology

The methodology employed in this study integrates large-scale text processing, semantic similarity techniques, and standardized skill taxonomies to extract relevant competencies from policy documents. The process begins by collecting policy texts as HTML files, which are converted into raw text through parsing with BeautifulSoup (Richardson, 2007). Each document is split into smaller, coherent segments (e.g., paragraphs) to enable more granular analysis. Following segmentation, the system applies a combination of TF-IDF filtering and part-of-speech analysis, conducted via spaCy (Honnibal & Montani, 2017), to identify domain-relevant nouns and proper nouns. This token-level filtering helps reduce extraneous text and narrows the focus to segments with potentially significant skill-related information.

Next, the European Skills/Competences, Qualifications and Occupations (ESCO) taxonomy is integrated to provide a consistent, multilingual framework for identifying and comparing skills. Developed by the European Commission, ESCO categorizes occupations, competencies, and qualifications in a structured format that facilitates clear cross-country alignment. Each skill entry in the ESCO database—encompassing the official label, any alternative labels, and a brief description—is encoded into dense vector representations using a Sentence-BERT model such as multi-qa-MiniLM-L6-cos-v1. This specific model is selected for its ability to handle multilingual text effectively, thus improving the accuracy of similarity matches even when policy documents or skill definitions use varied terminology. Once created, these skill embeddings undergo L2 normalization, a process that constrains each vector to a unit length. By normalizing the vectors, the pipeline can rely on cosine similarity approximations that are computationally efficient, especially during large-scale similarity searches conducted with libraries like FAISS. This systematic approach ensures that both the policy text segments and the ESCO-defined skills are positioned in a semantically coherent space, making it easier to identify meaningful overlaps between policy language and the standardized skill set.

To manage high volumes of embeddings efficiently, the system creates a FAISS index specifically for the vectors associated with ESCO skills (Johnson et al., 2019). FAISS—short for Facebook AI Similarity Search—was developed by Facebook AI Research to speed up the process of finding nearest neighbors in high-dimensional vector spaces, which is often a computationally expensive task. By indexing the ESCO skill embeddings in FAISS, the approach can leverage GPU acceleration to further improve performance, making it feasible to handle large-scale datasets without excessive response times.

Once the ESCO skills have been indexed, each segment (or paragraph) of the policy text goes through an identical embedding process using the same Sentence-BERT model (Reimers & Gurevych, 2019). These paragraph embeddings are then queried against the FAISS index to retrieve the top K closest matching skills. During this step, the system applies a similarity threshold to screen out weaker matches. Any paragraph-skill pair failing to meet or exceed this threshold is discarded, ensuring that the final matches reflect genuinely meaningful semantic overlaps. This mechanism helps maintain both precision and efficiency, allowing analysts to concentrate on the paragraph-skill pairs that truly indicate relevant competencies within the policy text.

Skill occurrences in individual paragraphs are then summed to expose recurring competencies most in demand, shedding lights in trends in policies targeting specific competencies, such as competency in terms of digital competency or competency in terms of sustainability. These recurring competencies can then be analyzed in relation to mappings at an occupations level derived out of ESCO. By comparing overlaps between referenced skills in policy documents and skills for a specific occupations, one can make an estimation of the level at which a policy is relevant to specific segments in workforce (OECD, 2023; World Economic Forum, 2023). High overlaps in skills in demand in occupations arise as candidates for specific interventions in terms of upskilling, reskilling, or additional intervention in terms of policy.

Finally, the analysis output is consolidated into both raw data formats (JSON, CSV) and user-friendly reports. Interactive dashboards and pie charts, created with visualization libraries such as Plotly, facilitate rapid review of skill distributions (Plotly Technologies Inc., 2015), and hyperlinked lists of skill identifiers allow users to directly consult ESCO descriptions. This approach increases transparency and reproducibility, as each step—from text extraction and embedding to similarity scoring—can be systematically retraced. Overall, the methodology combines natural language processing, vector similarity search, and standardized classification systems to deliver a scalable, data-driven framework for skill-oriented policy analysis.

The diagram (Figure 1) below shows a multi-stage pipeline that begins with Policy Documents. The text is first extracted (via HTML Parsing) and split into manageable segments (Text Segmentation). Subsequent steps—such as TF-IDF Filtering and Part-of-Speech Analysis—refine the text data. In parallel, relevant skills are pulled from the ESCO Skills Database and converted to Embeddings for efficient similarity computations using a FAISS Index. Once policy text embeddings and ESCO skill embeddings are generated, a Similarity Search step pinpoints the most relevant skills within each text segment. After setting a Similarity Threshold and Counting Skill Occurrences, the workflow proceeds to Analysis & Visualization, where frequency analyses, visualizations, and occupation mappings are generated. Finally, Final Outputs such as JSON/CSV data, interactive dashboards, and skill-occupation reports are produced to inform policy insights and decision-making.

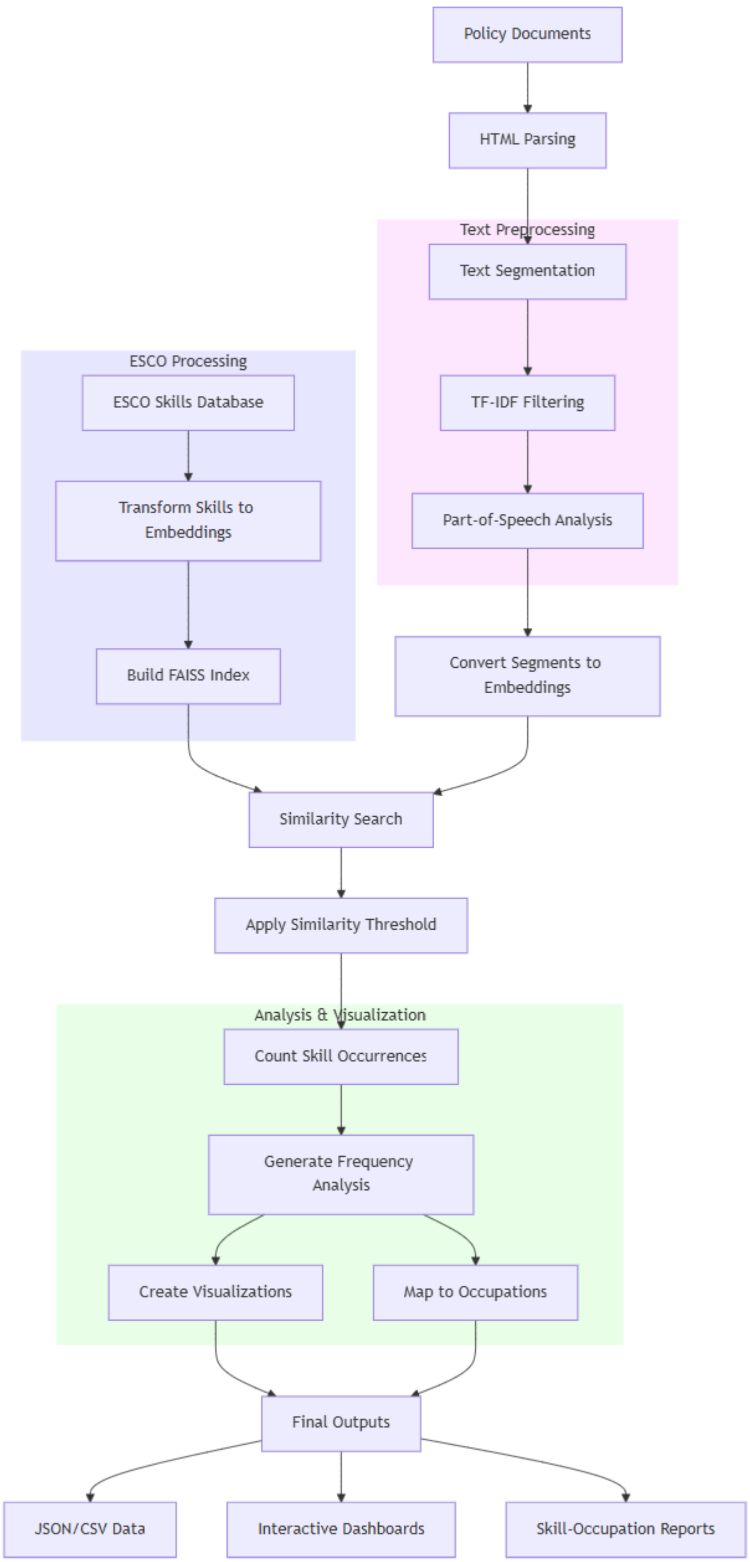

Figure 1. A high-level workflow illustrating how policy documents are parsed, preprocessed, and analyzed using ESCO-based skill embeddings and similarity search.

4. Results

The results presented here illustrate how the AI-driven skill extraction and occupation relevance pipeline can yield actionable insights from policy documents of varying scope and complexity. Although the following subsections reference a single case example for demonstration, the methodology applies broadly to any policy text where identifying workforce skill requirements is paramount.

4.1 Paragraph-Level Skill Matching

After an initial parsing phase, the text is split into chunks to enable more precise contextual analysis. Each chunk is encoded into a dense vector representation using a Sentence-BERT model and then compared to an indexed ESCO skill database via FAISS. Applying a predefined similarity threshold ensures that only high-relevance matches are retained.

In practice, many chunks may produce no substantial skill matches, particularly if they are short or contain broad, generic statements. However, where specific terminology or domain-related references appear, the model retrieves relevant ESCO skill IDs at the similarity predefined scores. Across a single policy text, this procedure can reveal anywhere from a handful to dozens of unique skill IDs, forming the core dataset for subsequent analyses.

4.2 Most Frequently Matched Skills

Once all paragraphs are processed, a frequency analysis highlights which skill IDs appear most often above the similarity threshold. This step surfaces recurring themes and domain focus areas within the policy. Skills linked to logistics management, regulatory compliance, digital competencies, or sustainability commonly emerge in transport-related policies, whereas other domains (e.g., healthcare or cybersecurity) may emphasize different competencies altogether. This frequency profile offers a quick snapshot of the policy's main focal points and can guide deeper, more specialized reviews.

4.3 Skill Proportions Visualization

To provide an intuitive overview of which skills predominate in a given policy text, the pipeline generates a pie chart showing each ESCO skill ID alongside its relative frequency. An accompanying HTML file optionally includes clickable links that lead to the standardized ESCO definitions of each skill, enabling stakeholders to explore detailed descriptions, scope, and interrelated competencies. This visualization helps decision-makers rapidly identify the most prominent skill areas emphasized in a document.

In figure 2 below we can see the chart is composed of multiple colored slices, each representing a specific ESCO skill URL. Percentages next to each slice range roughly from 1.69% to around 6.78%, indicating how often each skill appears relative to the total set of matched skills. A legend beneath the chart shows color-coded skill URLs, enabling viewers to cross-reference each slice with its corresponding ESCO skill definition

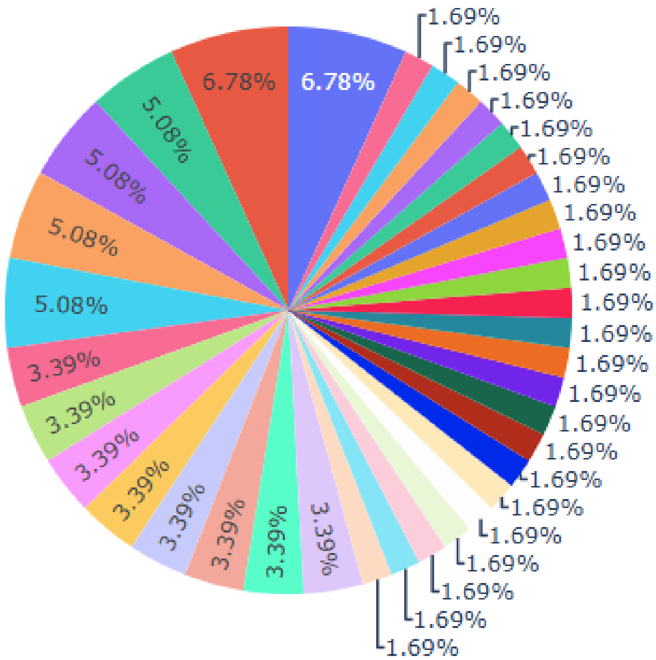

Figure 2: Pie chart illustrating the relative frequency of ESCO skills identified in a policy document.

In figure 3 is the list that accompany the aforementioned piechart which is the first part of a list of ESCO skill URLs displayed alongside small, colored blocks. Each block's color matches one of the pie slices in the main chart, so users can click the URL or refer to the text to learn more about that particular skill's definition and context within the ESCO taxonomy

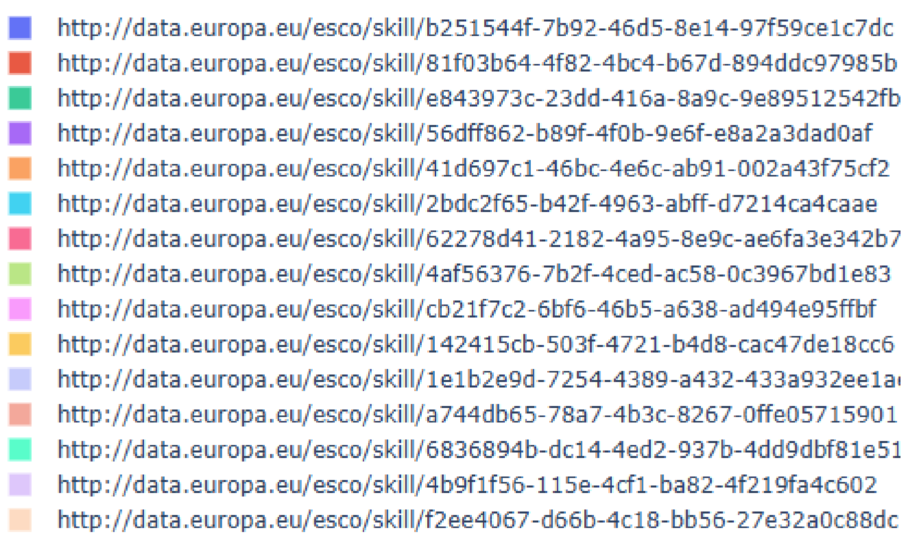

Figure 3: Legend associating each skill URL with a color in the corresponding pie chart.

4.4 Occupation Relevance Analysis

The system not only measures the frequency of specific skills but also cross-references those skills with ESCO's occupation data, in which each occupation is associated with a set of essential or recommended competencies. By evaluating the overlap between the set of skills identified in the policy and the skill requirements defined for each occupation, the pipeline produces a relevance score (often expressed as a percentage). Roles with high scores are typically those that the policy explicitly addresses or significantly affects. When a substantial portion of an occupation's skill set matches the policy-driven skills, that occupation is deemed highly relevant. In contrast, occupations with moderate or niche relevance indicate only partial overlaps, suggesting that the policy plays a smaller or more specialized role for those roles. Even with a lower overall alignment, these findings may pinpoint areas where targeted upskilling or retraining efforts could be beneficial. This occupation mapping proves especially helpful for developing focused workforce strategies, highlighting where policy objectives strongly align with current labor-market demands, as well as revealing skill gaps that may require additional attention.

4.5 Efficiency and Scalability

The workflow leverages FAISS for rapid vector similarity search and performs batch embeddings for segments of policy text. This design choice allows the approach to process documents swiftly—each paragraph-level encoding and lookup can execute in fractions of a second on modern hardware. Consequently, the methodology scales well to large corpora, including multiple policy documents or extensive legislative archives, without substantial performance constraints—particularly when GPU acceleration is available.

4.6 Summary of Key Findings

Paragraph-level matching identifies most recurring competencies, offering a fact-based picture of a key skill theme for a policy, and then, when such skills have been translated into occupations under an ESCO-classified skill, whose jobs will most gain impact will become apparent, and planning and anticipation can then follow for training, regulative, and budget-allocation requirements. The resulting outputs—such as pie charts, CSV files, and HTML reports—enable swift, evidence-based decision-making and foster a shared understanding of policy priorities across various stakeholders.

Overall, this AI-powered system converts unstructured policy text into a well-organized set of relevant skills and occupations. By aligning policy language with established taxonomies, decision-makers gain a clear, data-backed basis for strategic planning, targeted interventions, and more effective policy implementation.

5. Discussion

This AI-driven methodology supports faster, more objective analyses of policy texts, potentially transforming how governments and organizations plan for future skill needs. By mapping skill demands to occupations, policymakers can proactively identify gaps in training programs, develop targeted upskilling initiatives, and adjust regulatory frameworks to match evolving labor market requirements (World Economic Forum, 2023).

Beyond accelerating policy review, the proposed framework creates a more transparent and data-oriented decision-making environment. In traditional expert-driven processes, subjective biases or a limited pool of expertise can distort which skills are highlighted. By contrast, AI-based skill extraction offers a repeatable procedure for scanning large corpora, thus amplifying the range of inputs considered. Additionally, embedding-based techniques enable analysts to capture contextual nuances in language, potentially uncovering emerging skills or thematic clusters that might go unnoticed in purely keyword-based approaches. This can be especially valuable for interdisciplinary or future-facing sectors (e.g., digital sustainability, AI ethics) where definitions remain fluid.

However, a few considerations must be addressed:

1. Model and Data Bias: The underlying NLP models may inherit biases from their training data. Validating the extracted skills with domain experts remains critical. Certain underrepresented or rapidly evolving domains may also be inadequately captured in existing models. Regular retraining on broader or updated corpora, along with human-in-the-loop validation, could mitigate these risks (Chen, 2023).

2. Granularity of Skills: ESCO's skill granularity could either be too high-level or too detailed, depending on policy contexts. Further customization or fine-tuning might be necessary for domain-specific analyses. In rapidly transforming fields (e.g., AI, biotechnology), new or niche skills might not yet appear in official taxonomies. Mechanisms for continuously updating skill libraries—and reconciling them with ESCO—may be needed to keep pace with change (European Commission, 2022b).

3. Language and Domain Coverage: While the system currently leverages spaCy for English text, additional language models or translations would be required for truly multi-lingual policies. This requirement becomes even more pressing in multinational contexts, such as the European Union, where policies are often issued in multiple languages. Future research could explore employing multilingual embedding models, thereby broadening the framework's applicability to diverse linguistic landscapes.

4. Interpretability and Thresholding: Although embedding-based matching offers robust semantic similarity, the rationale behind model inferences may not always be immediately clear to policymakers. Selecting an appropriate similarity threshold can also be subjective and may vary by domain, potentially impacting which skills are flagged or omitted. Enhanced interpretability—through attention heatmaps or examples of matched phrases—could build trust in AI-driven policy analysis (OECD, 2023b).

5. Implementation and Operational Costs: Scaling the solution to very large text corpora may require specialized hardware (e.g., GPUs) and a stable data infrastructure to manage embeddings, indexes, and outputs. Smaller governmental bodies or institutions with limited resources may face barriers in adopting such systems without additional support or collaboration.

Moreover, as language and AI become increasingly dominant in policy analysis, information safeguard techniques become a high-priority issue for researchers (Feretzakis et al., 2024a) and for creating trust in AI (Feretzakis et al., 2024). Having such controls in place ensures compliance with legislation for data and creates increased confidence in automated processes for policy.

Regardless of such concerns, the model is a strong evidence-based tool for reconciling workforce realities and policy aims. By combining AI-facilitated language analysis with normalized taxonomies (ESCO), it establishes a coherent path towards ongoing tracking of skills, adaptable development of policy, and focused workforce development interventions.

6. Conclusion

This study has constructed a repeatable, AI-enabled skill extraction and analysis for big policy text corpora, leveraging FAISS-based search for similar items, Sentence-BERT embeddings, and ESCO as a normalized classification scheme. By mapping passages of text onto vector representations and comparing them with ESCO-established skills, policymakers and analysts gain fact-based information about competencies most stressed or hinted at in policy documents. That, in its turn, helps in developing evidence-based curricula adaptations in education, workforce development programs, and regulative actions.

Scalability emerges as a key benefit: high-dimensional indexing through FAISS allows the framework to be extended to large or multi-lingual policy collections with minimal computational bottlenecks. In addition, the methodology's modular design—encompassing text parsing, segmentation, vector embeddings, similarity search, and outcome reporting—makes it straightforward to integrate into existing policy analysis pipelines or adapt for new domains such as education, corporate HR, and research.

Despite these advantages, several future directions and open challenges remain:

- Multilingual Expansion: While demonstrated primarily on English texts, expanding to additional languages would address the needs of international organizations, especially the European Union, where policies frequently span multiple national languages.
- Evolving Skill Taxonomies: As emerging fields (e.g., AI ethics, hydrogen technologies) introduce new skill sets, ESCO and similar frameworks require periodic updates or domain-specific extensions. Tools for automated taxonomy augmentation could keep the methodology current.
- Interpretability and Trust: Policymakers and practitioners often need to understand why certain skills are highlighted. Integrating explainable AI features, such as attention heatmaps or model-justified snippets, could build confidence and reduce adoption barriers.
- Advanced NLP Techniques: The rapid evolution of large language models (LLMs) suggests new avenues for fine-grained skill extraction, contextual disambiguation, and cross-lingual policy alignment. Further research might explore how generative LLMs or domain-adapted embeddings could enhance both accuracy and domain coverage.

In sum, the framework outlined here demonstrates how AI can streamline policy analysis by transforming unstructured text into a structured map of relevant skills and potential occupational impacts. This not only accelerates the policy review process but also increases its objectivity, supporting more timely, targeted, and transparent decision-making. As AI technologies continue to advance and the global workforce faces ever more dynamic skill demands, such scalable, data-driven solutions will be indispensable for bridging the gap between policy intent and labor market realities.

References

Association for Computing Machinery. (2024). *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '24)*. ACM. https://doi.org/10.1145/3626772

Babashahi, L., Barbosa, C. E., Lima, Y., Lyra, A., Salazar, H., Argôlo, M., Almeida, M. A. d., & Souza, J. M. d. (2024). AI in the Workplace: A Systematic Review of Skill Transformation in the Industry. Administrative Sciences, 14(6), 127. https://doi.org/10.3390/admsci14060127

Chen, P., Wu, L., & Wang, L. (2023). AI Fairness in Data Management and Analytics: A Review on Challenges, Methodologies and Applications. *Applied Sciences*, *13*(18), 10258. https://doi.org/10.3390/app131810258

European Commission. (2022a). ESCO classification – European skills, competences, qualifications and occupations. Retrieved from https://esco.ec.europa.eu

European Commission. (2022b). The European Skills Agenda for sustainable competitiveness, social fairness, and resilience. Official Journal of the European Union.

Feretzakis, G., Papaspyridis, K., Gkoulalas-Divanis, A., & Verykios, V. S. (2024a). Privacy-Preserving Techniques in Generative AI and Large Language Models: A Narrative Review. Information, 15(11), 697. https://doi.org/10.3390/info15110697

Feretzakis, G., & Verykios, V. S. (2024). Trustworthy AI: Securing Sensitive Data in Large Language Models. AI, 5(4), 2773–2800. https://doi.org/10.3390/ai5040134

Honnibal, M., & Montani, I. (2017). spaCy 2: Natural language understanding with Bloom embeddings. To appear. https://spacy.io

Johnson, J., Douze, M., & Jégou, H. (2019). Billion-scale similarity search with GPUs. IEEE Transactions on Big Data, 7(3), 535–547.

Koundouri, P., Theodossiou, N., Stavridis, C., Devves, S., & Plataniotis, A. (2022). A methodology for linking the energy-related policies of the European Green Deal to the 17 SDGs using machine learning (DEOS Working Papers 2202). Athens University of Economics and Business. https://mpra.ub.uni-muenchen.de/122118/1/MPRA_paper_122118.pdf

Koundouri, P., Aslanidis, P. S., Dellis, K., Feretzakis, G., & Plataniotis, A. (2024a). Uncovering the SDG content of Human Security Policies through a Machine Learning web application. DEOS Working Papers, 2406, Athens University of Economics and Business. https://mpra.ub.uni-muenchen.de/121972/

Koundouri, P., Feretzakis, G., & Alamanos, A. (2024b). Integrating AI into Energy Systems: The approach of the Global Climate Hub. https://wpa.deos.aueb.gr/docs/2025.Global.Climate.Hub.AI.pdf

Kraus, S., Durst, S., Ferreira, J. J., Veiga, P., Kailer, N., & Weinmann, A. (2022). Digital transformation in business and management research: An overview of the current status quo. International Journal of Information Management, 63, 102466. https://doi.org/10.1016/j.ijinfomgt.2021.102466

Liu, J., Chen, K. & Lyu, W. Embracing artificial intelligence in the labour market: the case of statistics. Humanit Soc Sci Commun 11, 1112 (2024). https://doi.org/10.1057/s41599-024-03557-6

Naeem, G., Asif, M., & Khalid, M. (2024). Industry 4.0 digital technologies for the advancement of renewable energy: Functions, applications, potential and challenges. *Energy Conversion and Management: X, 24*, 100779. https://doi.org/10.1016/j.ecmx.2024.100779

Organisation for Economic Co-operation and Development (OECD). (2023a). AI and the labor market: The potential and the pitfalls. OECD Policy Brief. Retrieved from https://www.oecd.org/employment/AI-labour-market-policy-brief.pdf

Organisation for Economic Co-operation and Development (OECD). (2023b). OECD Skills Outlook 2023: Skills for the digital transition. OECD Publishing. https://doi.org/10.1787/c61e8147-en

Plotly Technologies Inc. (2015). Collaborative data science. https://plotly.com

Rashid, A. B., & Kausik, M. A. K. (2024). AI revolutionizing industries worldwide: A comprehensive overview of its diverse applications. Hybrid Advances, 7, 100277. https://doi.org/10.1016/j.hybadv.2024.100277

Reimers, N., & Gurevych, I. (2019). Sentence-BERT: Sentence embeddings using Siamese BERT-networks. Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing (EMNLP), 3982–3992.

Richardson, L. (2007). Beautiful Soup documentation. https://www.crummy.com/software/BeautifulSoup/

Rikala, P., Braun, G., Järvinen, M., Stahre, J., & Hämäläinen, R. (2024). Understanding and measuring skill gaps in Industry 4.0 — A review. Technological Forecasting and Social Change, 201, 123206. https://doi.org/10.1016/j.techfore.2024.123206

Supriyono, Wibawa, A. P., Suyono, & Kurniawan, F. (2024). Advancements in natural language processing: Implications, challenges, and future directions. Telematics and Informatics Reports, 16, 100173. https://doi.org/10.1016/j.teler.2024.100173

World Economic Forum. (2023). The Future of Jobs Report 2023. World Economic Forum. Retrieved from https://www.weforum.org/reports/the-future-of-jobs-report-2023