Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Towards explicit soft skills labelling in ESCO through semantic NLP analysis

Ciprian Panzaru

ciprian.panzaru@e-uvt.ro

West University of Timisoara https://orcid.org/0000-0001-7652-4159

Anamaria Grama

West University of Timisoara

Research Article

Keywords: softs skills, ESCO, NLP, semantic similarity, labour market

Posted Date: November 6th, 2024

DOI: https://doi.org/10.21203/rs.3.rs-5393557/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: The authors declare no competing interests.

Abstract

The classification and identification of soft skills within the European Skills, Competences, Qualifications, and Occupations (ESCO) framework presents a significant challenge, as ESCO does not explicitly indicate which of its listed skills are considered “soft”. This lack of clear labelling complicates the strategic alignment of educational programs and workforce development initiatives with the evolving demands of today’s labour market, where soft skills are increasingly recognised as essential for professional success. Our study seeks to address this challenge by conducting a systematic literature review to identify and define a comprehensive set of soft skills, and by employing Natural Language Processing (NLP) techniques to map these skills to the ESCO taxonomy. Through semantic analysis and comparison of soft skills identified from scientific literature against the ESCO skillset, we establish a refined classification that explicitly labels soft skills within the ESCO framework. Our findings provide a necessary clarification of soft skills within the standardised European context, offering valuable insights for policymakers, educators, and HR professionals focused on fostering soft skills development. By clarifying the soft skills embedded within the ESCO framework, this research contributes to the advancement of skill taxonomies and promotes a more integrated approach to the development and assessment of soft skills.

Introduction

In the rapidly evolving landscape of the 21st-century labour market, soft skills - encompassing personal attributes, personality traits, inherent social cues, and communication abilities - have become indispensable components of employment, education, and personal development. Recent advancements in the business environment, driven by technological progress and the increasing competitiveness and dynamism of the global economy, have contributed to the growing demand for new skill sets (Dolce et al. 2020). Unlike hard skills, which are often measurable and technical in nature, soft skills involve a broad spectrum of interpersonal and cognitive abilities, including teamwork, leadership, creativity, and emotional intelligence. These capabilities are increasingly recognised for their critical role in determining professional success, facilitating smooth integration into diverse work environments, and enhancing individual adaptability in an era characterised by continuing technological advancements and rapidly changing occupational demands (Clarke 2018; Succi 2019; Toner 2011; Tsirkas et al. 2020).

The classification and identification of soft skills within frameworks, such as the European Skills, Competences, Qualifications, and Occupations (ESCO), encounter significant challenges due to the absence of a distinct labelling system explicitly dedicated to soft skills. In ESCO, these competencies are subsumed under the broader category of “Transversal Knowledge, Skills, and Competences”, which encompasses a diverse array of abilities extending beyond soft skills alone. ESCO recognises the foundational role of these skills in personal development and their critical importance in acquiring technical skills necessary for labour market integration and success. Their broad applicability across multiple occupational sectors and industries underscores their value within professional environments (European Commission 2024). However, the lack of precise and consistent classification complicates

the alignment of educational curricula and labour development initiatives with the specific demands of the modern labour market, where the role of soft skills is increasingly acknowledged as vital to professional achievement and advancement.

To address these challenges, our research aims to identify which ESCO skills can be classified as soft skills by applying advanced Natural Language Processing (NLP) techniques. Through the extraction of soft skill examples from literature and their systematic alignment with the skills listed in the ESCO framework, this study establishes a refined classification that explicitly identifies and categorises specific ESCO skills as soft. This classification enhances the representation of soft skills within ESCO, differentiating them from other skill types and providing essential insights for aligning educational programs and workforce development with current labour market needs.

The structure of this article is organised as follows: Section 2 reviews the literature on the identification and classification of soft skills, examining the various definitions and terminologies used across disciplines. Section 3 outlines the methodological approach used to match soft skills with ESCO-listed skills. In Section 4, the results are presented, highlighting both exact and semantic matches between soft skills identified in the literature and those within the ESCO framework. This section also provides an analysis of the distribution of soft skills within ESCO, with a focus on cross-sector and sector-specific applications. Finally, Section 5 concludes with recommendations for refining soft skill classifications within ESCO, alongside potential implications for policy, education, and workforce development.

Literature Review

The concept of soft skills encompasses a wide range of non-technical abilities that are crucial for personal development, educational achievement, and professional success. Despite their recognised importance, there remains a notable lack of consensus regarding the precise definition of soft skills. Current practices for identifying and cataloging soft skills vary significantly, resulting in notable discrepancies in how these skills are understood, taught, and assessed across contexts. Often contrasted with hard skills, soft skills are more associated with emotional intelligence (EQ) rather than intellectual intelligence (IQ), with Goleman (1996) being one of the earliest to highlight this distinction. Siekmann and Fowler (2017) and Succi (2019) further observe that, unlike hard skills, soft skills are inherently challenging to categorise due to their subjective nature and the contextual variations in how they are perceived. Over the years, soft skills have been referred to by various terms such as life skills (World Health Organization 1994), social skills, generic competencies, interpersonal skills, leadership skills, transversal skills, and even key competences for lifelong learning (European Union 2006). Furthermore, the discussion around soft skills has increasingly intersected with concepts such as 21st-century skills and skills for the future of work, where digital competencies are also often included (Ananiadou and Claro 2009). This plurality of terminologies has only added to the complexity of establishing a clear and unified understanding of soft skills in both academic and professional contexts. The challenge of defining soft skills has been a persistent issue in academic literature, with scholars presenting diverse and often divergent perspectives on what constitutes these skills. Definitions vary

widely, reflecting differing interpretations and emphases across studies. The term “soft” itself is often substituted by words like “generic”, “fundamental”, “key”, “personal”, “transferable”, or “emotional” depending on the context. Similarly, the term “skills” can be ambiguously replaced with “abilities”, “competencies”, or even “qualities”, contributing to further confusion (Bennett et al. 1999; Chamorro-Premuzic et al. 2010; Matteson et al. 2016).

Davis-Blake et al. (2002) offered an early definition of soft skills as “skills and traits pertaining to personality, attitude, and behaviour”, distinguishing these from formal knowledge or technical expertise and broadening the concept beyond technical competencies. Later, Heckman and Kautz (2012) expanded on this foundation by describing soft skills as personality traits, goals, motivations, and preferences valued in the labour market, educational settings, and beyond. They highlighted the “Big Five” personality traits - openness, conscientiousness, extraversion, agreeableness, and emotional stability - as core components, underscoring the role of personality in defining soft skills. Similarly, Yorke (2006) characterises soft skills as a “combination of dispositions, understandings, attributes, and practices”. Adding to this, Knight and Page (2007) refer to them as “tricky competences” due to their elusive and context-dependent nature, highlighting their continuous development throughout one’s life. Haselberger et al. (2012) take a more comprehensive view, defining soft skills as the dynamic combination of cognitive, metacognitive, interpersonal, intellectual, and practical skills. These perspectives underscore the complex nature of soft skills and their critical role in enabling individuals to navigate both professional and personal challenges effectively. As highlighted by Bennett et al. (1999), Chamorro-Premuzic et al. (2010), and Dolce et al. (2020), the definition of soft skills remains elusive, primarily due to semantic differences in how these skills are described across various sectors and languages.

The fluid nature of soft skills also leads to debates about the distinction between soft and hard skills, with some scholars arguing that the line between the two is often blurred. For instance, Papoutsoglou et al. (2017) define soft competences as the “ability to use knowledge, personal, social, and/or methodological skills” in work or study contexts, while hard competences are associated with the practical application of knowledge and problem-solving abilities. González-Eras and Aguilar (2019) suggest that competences can be seen as a demonstration of the application of generic skills in specific contexts. In some cases, what is considered a soft skill in one field, such as singing, may be classified as a hard skill in another, as in the case of professional performers. This interchangeable nature of skills, depending on context, underscores the challenge of establishing universally applicable definitions and classifications (Khaouja et al. 2019).

The complexity of soft skills definition extends to their classification, with various scholars proposing diverse taxonomies to clarify their scope. Cinque and Ciappei (2014) for example developed a taxonomy that divides soft skills into four categories based on the dimension of actions: political, strategic, organisational, and ethical. These categories are further highlighted across personal, social, and methodological dimensions, illustrating the multi-layered nature of soft skills.

As definitions of soft skills have evolved, the demand for rigorous methodologies to effectively identify, classify, and analyse these complex competencies has similarly intensified. Traditional qualitative approaches, including interviews and surveys, initially provided valuable insights into the role and perception of soft skills. However, the growing complexity of these skills, compounded by their varied interpretations across disciplines and industries, has highlighted the limitations of these methods in capturing the full range of soft skills. In response, recent advancements in technology, particularly Natural Language Processing (NLP), have emerged as promising tools to address these challenges, allowing for more systematic and scalable approaches to soft skills analysis and integration into frameworks such as ESCO.

Chiarello et al. (2021) illustrates the potential of advanced NLP techniques through a text mining analysis aimed at evaluating the adaptability of ESCO framework in addressing skill demands driven by Industry 4.0. Through this analysis, they underscore that utilising methods like Named Entity Recognition (NER) and lexical searches can effectively facilitate the integration of emergent technological competencies into ESCO, ensuring its responsiveness to rapid advancements. Further expanding on the application of NER techniques, Fareri et al. (2021) developed SkillNER, a system trained to extract soft skills from unstructured text sources such as scientific literature and job descriptions within the ESCO framework. Continuing this exploration, Khaouja et al. (2021) reviewed 108 research articles on the topic of identifying skills from online job postings, emphasising that NLP and information extraction techniques have been critical in building customised databases of skills from textual knowledge bases. Machine learning techniques, including neural networks and word embeddings, are also used to capture skills from text-based resources with increasing precision. A notable study by Zhang et al. (2022) applied machine learning techniques using noisily labeled data to train models to classify skills extracted from job advertisements. The researchers demonstrated that their model's performance in identifying skills closely matched human extraction, showing that automated systems could be nearly as effective as human experts in identifying and classifying job-related skills, including soft skills. However, the study also highlighted some persistent challenges, particularly with NLP's ability to fully capture the subtleties of soft skills descriptions. Moreover, many studies still rely on human judgment for coding and categorising soft skills, introducing potential biases and subjectivity into the research process.

While significant progress has been made in understanding and applying technologies like NLP and machine learning to classify and analyse soft skills, substantial gaps remain in the standardisation of definitions, categorisation, and evaluation methodologies. This lack of uniformity complicates the alignment of educational frameworks with labor market needs, impeding the development of cohesive strategies for soft skills education and assessment (Bak and Jordan 2018; Balcar 2016). The absence of a standardised methodology across frameworks such as developed by the OECD, CEDEFOP or the US O*NET system further challenges policymakers, educators, and industry specialists in building a workforce equipped not only with technical proficiency but also essential interpersonal and cognitive abilities (Cinque and Ciappei 2014; Siekmann and Fowler 2017). Initiatives such as ESCO's recent labeling of "green" skills (European Commission 2022) underscore the need for similar efforts to explicitly categorise soft skills, ensuring that these competencies are effectively integrated into

educational curricula and workforce development strategies to meet labour market requirements more effectively.

Methodology

Building on the challenges identified in defining and classifying soft skills, this research takes a novel approach by combining a structured literature review with advanced Natural Language Processing (NLP) techniques to identify and categorise soft skills within the ESCO framework. This research offers a methodological contribution by integrating a structured literature review with advanced Natural Language Processing (NLP) techniques to systematically identify and categorise soft skills within the ESCO framework, bridging empirical literature insights with the semantic precision of NLP to enhance the alignment and comprehensiveness of soft skills classification.

In the initial stage, a targeted literature review was conducted to systematically extract and analyse specific soft skills as defined and conceptualised by various scholars. An analysis of the literature, particularly of scientific articles focused on soft skills, is analogous to consulting subject matter experts. However, in this context, scientific articles draw upon a more extensive and diverse pool of expert knowledge, thereby providing a more comprehensive and reliable perspective. Our focus was exclusively on identifying soft skills without synthesising or reclassifying existing definitions. The literature review was conducted using the Web of Science database, selected for its broad coverage of high-impact, peer-reviewed scholarly publications. This selection ensured that the soft skills identified were drawn from credible and relevant sources. The following keywords were employed: “soft skill”, “soft skills”, “soft-skills”, and “soft-skill”. These variations were included to account for different terminologies and hyphenations across publications. The search was restricted to the “Topic” field, which encompasses the article title, abstract, and keywords, to ensure that the articles explicitly referenced or focused on soft skills as a central aspect of their discussion or conceptual framework. Subsequently, a filter was applied to include only review articles, as they typically provide a synthesis of existing research, offering a more complete and consolidated understanding of soft skills concept. This approach enabled us to capture diverse conceptualisations of soft skills, thereby enhancing the robustness of the dataset by incorporating insights from multiple sources rather than relying on isolated empirical studies. After downloading the initial list, we conducted a preliminary analysis of the titles and abstracts to refine the selection. Articles mentioning soft skills only in passing or in unrelated contexts were removed, ensuring that the final list was closely aligned with the research focus. In cases where the relevance of an article was ambiguous, a second, more detailed review of the introduction or conclusion was performed to confirm its suitability. This filtering process ensured that the final list consisted of articles making substantive contributions to the understanding of soft skills, thereby enhancing the precision and relevance of the study. Following the refinement process, 134 articles that met the inclusion criteria and directly addressed soft skills were identified. An additional 62 papers, identified during the documentation process as particularly relevant to our research objectives, were manually included. Although these papers were not retrieved through the initial Web of Science search, they were incorporated due to their significant contributions to the subject, as indicated by citation networks,

references in key review articles, or their recognised prominence within the soft skills literature. This supplementary inclusion step ensured that the list of papers was as exhaustive as possible, encompassing all pertinent literature. Consequently, the final list comprised 196 articles. From this collection, 683 skills were initially extracted, and after removing duplicates, 202 unique soft skills were retained. These skills constituted the dataset used for comparison with the ESCO classification. The complete list of soft skills extracted from literature is presented in Annex 1.

In the second stage, we sought to identify which skills within the ESCO framework could be categorised as soft skills by matching them with the soft skills extracted from the literature. We utilised the ESCO framework, version 1.2.0., released on 13 May 2024. To construct our dataset of ESCO skills, we extracted both the preferred labels, and the alternative labels provided for each concept in ESCO1. These labels formed the ESCO skills dataset, which we then systematically compared with the dataset of soft skills derived from the literature. The ESCO dataset consisted of 13,939 preferred labels and 86,886 alternative labels.

With embeddings generated for both the soft skills extracted from scientific literature and the skills listed in the ESCO framework, the final stage of the analysis involved matching these two sets of skills.

In the initial phase of our matching process, we employed exact string matching to identify identical skills between the soft skills dataset, extracted from the literature and the ESCO dataset. This direct comparison was conducted in a case-insensitive manner and excluded punctuation, focusing solely on word-for-word matches without any transformations or semantic interpretations. The matching process involved systematically comparing each skill from the literature-derived dataset with those in the ESCO dataset using straightforward lookup functions to identify perfect matches. While exact matching is highly reliable it has limitations, such as the inability to capture synonyms, variations in word forms, or different word orders.

To overcome the limitations inherent in exact matching and to identify a broader range of correspondences between soft skills identified in the literature and those catalogued within the ESCO framework, the subsequent phase involved the application of Natural Language Processing (NLP) techniques. This computational approach facilitated a systematic alignment and cross-referencing of the two datasets beyond identical lexical matches. Semantic similarity analysis was employed to detect nuanced relationships between skill descriptions, enabling the identification of semantically related skills despite variations in exact wording. Semantic similarity measures the degree of likeness between two pieces of text based on the meaning they convey. Unlike syntactic similarity, which focuses on the structural aspects of language (e.g., grammar, syntax), semantic similarity explores contextual and conceptual parallels. This process is crucial for various applications, including information retrieval, text summarisation, sentiment analysis, and machine translation. Semantic similarity is rooted in the idea that words, phrases, sentences, and even larger text segments can have varying degrees of relatedness based on their meanings. For instance, “car” and “automobile” are semantically very similar, whereas

“car” and “tree” are not. The challenge lies in quantifying this similarity in a way that computers can process.

Prior to conducting the semantic analysis, extensive data preprocessing was carried out to standardise and clean both the literature-derived dataset and the ESCO dataset. This process ensured data consistency and improved the accuracy of the matching procedure. Key preprocessing steps included converting all text to lowercase, removing punctuation, special characters, and excess whitespace, while preserving stop words in multi-word expressions to maintain semantic integrity. Duplicate entries were eliminated to prevent bias in the matching results.

To quantify the semantic similarity between the skills in the ESCO dataset and the predefined soft skills from the literature, we employed Sentence-BERT (SBERT) 2, a pre-trained model optimised for generating dense vector embeddings of sentences and phrases. The following mathematical representation outlines the computational methodology employed to match soft skills from literature with ESCO skills.

Let $T_{lit} = \{t_1,\ t_2,\ t_3 \ldots\ \ t_n\}$ represent the set of skill terms from the literature dataset and $T_{esco} = \{s_1,\ s_2,\ s_3 \ldots\ \ s_m\}$ represent the set of skills from the ESCO dataset. In both datasets terms were represented by a single word or a multi word expression (e.g., “communication” or “manage other people’s feelings”). Each skill description t_i from the literature dataset and s_j from the ESCO dataset was encoded as a vector embedding using the SBERT model M:

$$v_i^{lit} = M\left(t_i\right),\ \ v_j^{esco} = M\left(s_j\right)$$

1

where $v_i^{lit} \in\ \mathbb{R}^d$ and $v_j^{esco} \in\ \mathbb{R}^d$ and d is the dimensionality of the embedding space. This process generated two sets of embeddings: one for the literature dataset and one for the ESCO dataset.

The semantic similarity between each skill in the ESCO dataset and all skills in the literature dataset was measured using cosine similarity3. Cosine similarity quantifies how closely two vectors align in a high-dimensional space, and is defined as follows:

$$S\left(v_j^{esco},\ v_i^{lit}\right) = \frac{v_j^{esco} \bullet\ v_i^{lit}}{||\ v_j^{esco}\ ||\ \ ||\ v_i^{lit}\ ||}$$

2

where $v_j^{esco} \bullet\ v_i^{lit}$ denotes the dot product of vectors v_j^{esco} (the embedding of an ESCO skill) and v_i^{lit} (the embedding of a soft skill from the literature), and $||\ v_j^{esco}\ ||$ and $||\ v_i^{lit}\ ||$ are the Euclidean norms (magnitudes) of the of the vectors. The cosine similarity score ranges from − 1 to 1, where 1 indicates identical directionality (maximum semantic similarity), 0 indicates orthogonality (no similarity), and − 1 indicates opposite directionality.

For each ESCO skill s_j, cosine similarity scores were computed against all soft skills in the literature dataset. The maximum cosine similarity score $S_{max}\left(v_j^{esco}\right)$ was then selected for each ESCO skill, representing the strongest semantic match:

$$S_{max}\left(v_j^{esco}\right) = max\left\{S\left(v_j^{esco} \bullet \ v_1^{lit}\right),\ S\left(v_j^{esco} \bullet \ v_2^{lit}\right),\ \ldots\ , S\left(v_j^{esco} \bullet \ v_n^{lit}\right)\right\}$$

3

Thus, for each ESCO skill, the highest similarity score was identified as the best match to the set of predefined soft skills.

A threshold-based approach was employed to classify which ESCO skills could be considered soft skills. This threshold was selected to distinguish relevant soft skills from other skill types, with higher similarity scores indicating a stronger semantic relationship to soft skills. If $S_{max}\left(v_j^{esco}\right) \geq \ \theta$, then s_j is classified as a soft skill.

To ensure the robustness of the threshold selection, a manual review of skills near the threshold (both above and below) was conducted. This validation step ensured that the threshold provided an optimal balance between capturing relevant soft skills and excluding irrelevant ones. Skills classified as soft skills were reviewed for false positives (incorrect inclusions), and those excluded were checked for false negatives (incorrect exclusions).

This approach ensured that the model captured relevant soft skills with accuracy and minimised misclassification.

[1] ESCO functions as a comprehensive lexicon organised into two primary pillars: occupations and skills/competences. The skills pillar is subdivided into four hierarchical classifications: Knowledge, Language Skills and Knowledge, Skills, and Transversal Skills. Each concept within the skills pillar is accompanied by detailed metadata that includes “preferred terms” or preferred labels, which represent the primary descriptors for each concept, “non-preferred terms” or alternative labels, which cover synonyms, spelling variants, declensions, and abbreviations, and “hidden terms”, which encompass outdated, misspelled, or politically incorrect terms. Each concept is also described with additional metadata, such as scope notes, reusability levels, and the interrelationships among skills and between skills and occupations. There is no distinction between skills and competences.

[2] The SBERT model was implemented using the “sentence-transformers” library in Python. Skills from both datasets were embedded using the “paraphrase-MiniLM-L6-v2” model. SBERT was selected due to its proven efficiency in tasks involving semantic similarity, particularly for handling both single words and multi-word expressions.

[3] Cosine similarity scores were computed using the “pytorch_cos_sim” function from the “sentence-transformers” library in Python.

Results and Discussion

The research led to the development of a collection of soft skills within the ESCO framework, as presented in Annex 2. This collection was derived through a matching process between the soft skills identified in the literature and those listed in ESCO, yielding several key insights. The initial phase of the study demonstrated that, out of 202 soft skills identified in the literature review, 25 skills were found to have an exact match within the ESCO framework. Of these, 11 were identified as preferred label, while 14 appeared among the alternative labels. This represents approximately 0.078% of preferred labels and 0.016% of alternative labels being identified as soft skills through exact matching, see Table 1. Although the percentage of exact matches is relatively low, these matched skills represent core competencies widely recognised in both academic research and professional frameworks as essential soft skills.

Table 1
Exact matches between literature soft skills and ESCO skills

Preferred label	Alternative labels
systems thinking	system thinking, integrated thinking, integrate thinking, systems thought, integration of thinking
meet commitments	meet commitment, show conscientiousness, keep promises, meet targets, show reliability, demonstrate commitment, show responsibility, show commitment, achieve goals, demonstrate steady effort, demonstrate conscientiousness, meeting commitments, demonstrate punctuality, demonstrate reliability, work to deadlines, demonstrate loyalty
customer service	customer services, **customer care**, customer support, customer servicing, customer well being, customer satisfaction
communication	conversation, interaction, correspondence, communicating
assume responsibility	demonstrate a conscientious attitude, show responsible behaviour, be ready to take on responsibility, be responsible, showing responsibility, act responsibly, prove sense of responsibility, be ready to take responsibilities, demonstrate responsibility
cope with uncertainty	handle ambiguity, dealing with uncertainty, deal with uncertainties, handle uncertainty, manage uncertainty
entrepreneurship	business initiative, entrepreneurialism
risk management	risk assessment, risk administration, risk governance, hazard management
conflict management	conflicts management, conflict resolution, conflict supervision, conflict administration, management of conflicts
ethics	moral studies, the study of moral principles
make decisions	take decisions, choose, make decision, making decisions, decide, make choices
apply change management	alter management, administer change management, applying change management, change management application, change management, revise management, implement change management, rework management, apply changing management, utilise **change management**, apply change management, employ change management
demonstrate curiosity	display curiosity, demonstrating **curiosity**, show curiosity, demonstrate interest, show interest, be curious, curiosity, exhibit curiosity
approach challenges positively	show sense of humour, keep a positive attitude, adopt a constructive approach, **maintain a positive attitude**
exercise self-control	**self-control**, apply self-control, exercising self-control, exercise self-control
reflexion	reflexion on own behaviour, mindfulness, reflexion on own behaviour, **self-**

Note: Skills that are exact matches for soft skills are marked in "italic bold".

Preferred label	Alternative labels
	awareness, self-reflection
show empathy	demonstrate consideration, exhibit consideration, show concern for others, show courtesy, communicate an empathic response, display consideration, demonstrate concern for others, treat people fairly, show good manners, **concern for others**, showing empathy, be empathic
cope with stress	deal with pressure, perform under pressure, accept challenges, cope with pressures, **work under pressure**, manage pressure, handle pressure, show mental resilience, stay cool under pressure, coping with pressure
develop strategy to solve problems	devise strategy, come up with strategy to solve problems, set goals, set targets, design strategy to solve problems, developing strategy to solve problems, identify goals, develop strategies to solve problems, develop strategy to solve problem, create implementation strategy, adapt implementation strategy, **strategic thinking**, create strategy to solve problems, devise action plan, develop objectives, develop problem solving strategy, identify tasks, decide priorities
morality	ethics, decency, honour, **integrity**, **morals**, moral
listen actively	maintain active listening, apply active listening techniques, use active listening techniques, listen actively, **active listening**, listen with attentive participation
conflict management	conflicts management, **conflict resolution**, conflict supervision, conflict administration, management of conflicts
tolerate stress	permit stress, tolerating stress, stress toleration, endure stress, abide stress, **stress tolerance**
Note: Skills that are exact matches for soft skills are marked in "italic bold".	

The matched skills encompass a broad spectrum of essential competencies critical for both personal and professional growth. These include key communication skills, such as "active listening" and "communication", which are essential for effective interpersonal interactions. Additionally, skills like "conflict management", "conflict resolution", and "customer service" highlight the importance of human interactions, both in internal teams and client-facing roles. Leadership and decision-making abilities, such as "entrepreneurship", "strategic thinking", "make decisions", and "change management", highlights the important role of leadership in dynamic environments. Personal attributes such as "self-awareness", "self-control", "integrity", and "concern for others" indicates the growing recognition of emotional intelligence as a critical professional competency. These traits, along with "stress tolerance" and the ability to "work under pressure", underscore the need for resilience in modern workplaces. The ability to "cope with uncertainty" and maintain a "positive attitude" reflects the increasing importance of adaptability and a growth mindset. Furthermore, the inclusion of skills such as "systems thinking", "risk management", "ethics", and "morals" highlights the growing focus on holistic, ethical decision-making and responsibility in the professional sphere. The presence of "curiosity" and "maintain a positive attitude" indicates the significance of continuous learning and an optimistic approach to problem-solving and workplace challenges.

Although the exact matches between the literature-based soft skills and the ESCO framework were limited, these matches helped identify several key soft skills. However, further investigation was required to fully explore the alignment between the two sources. By examining the alignment through semantic similarity, conceptually related skills that were not captured through exact matching alone were identified. This approach provided a more comprehensive understanding of which skills within the ESCO framework could be considered soft skills, thereby expanding the scope beyond exact matches. The analysis involved examining both the similarity scores of the preferred labels and the associated alternative labels within the ESCO dataset. To enhance the robustness of the semantic analysis, exact-matched skills were retained within the NLP analysis as control elements, enabling the benchmarking of similarity scores and validation of the model's performance. Only skills that scored above a similarity threshold of $\theta = 0.70$ were classified as soft skills. Specifically, ESCO skills were categorised as soft skills if either their preferred label or at least one of their alternative labels had a similarity score equal to or greater than 0.70. This approach ensured that conceptually related skills with high semantic similarity were not excluded due to minor labelling differences between preferred and alternative labels. The decision to use this threshold was informed by a detailed manual review of skills with scores near the threshold, both above and below. Particular attention was given to the differences in similarity scores between preferred labels and their associated alternative labels, especially in cases where there were significant variations in scores. Skills initially classified as soft skills based on their similarity scores were further evaluated semantically and contextually. Those found not to align with the conceptual framework of soft skills were excluded from the final collection.

As a result, 235 concepts (preferred labels along with a total of 1,462 associated alternative labels) from the ESCO framework were categorised as soft skills. The remaining 13,704 skills, either with similarity scores below the 0.70 threshold or excluded upon manual review, were classified as non-soft skills and omitted from the final categorisation.

The distribution of soft skills within the ESCO framework indicates that a significant portion of identified skills are broadly applicable across sectors, with a concentration in cross-sector and sector-specific categories, see Fig. 1.

Figure 1: Distribution of soft skills among ESCO skill by reuse level

The analysis identified that 112 cross-sector skills, representing 47.66% of all ESCO skills labeled as soft, embody competencies with broad applicability across diverse professional environments. Skills such as "apply diplomatic principles", "create solutions to problems", and "manage feedback" illustrate essential qualities such as adaptability, effective communication, and interpersonal engagement, attributes foundational to employability and valued across sectors(Ananiadou & Claro, 2009; Goleman, 1996) This substantial proportion reflects the high transferability of cross-sector skills, underscoring the labour market's emphasis on versatile competencies that enable individuals to navigate varied workplace environments successfully.

Conversely, sector-specific skills accounted for 33.19% of the matched ESCO skills, with 78 competencies identified as vital to particular industries. Sector-specific competencies include “enterprise risk management”, “customer insight”, and “social alliances”. These findings resonate with the literature, which suggests that certain soft skills hold specialised significance within particular sectors, providing targeted interpersonal and cognitive abilities that enhance performance within industry-specific contexts (Succi 2019).

The study also identified transversal skills, which comprise 12.34% of the matched skills and represent 29 competencies universally valued across sectors. Examples include “show initiative”, “exercise self-control”, “accept criticism and guidance”, and “meet commitments”, all of which foster resilience, collaboration, and adaptability, rendering them indispensable in diverse professional settings. Certain transversal skills, such as “show initiative” and “exercise self-control”, are particularly pertinent in leadership roles, while competencies like “accept criticism and guidance” and “work in teams” facilitate growth through feedback and cooperation. Furthermore, the analytical competency “think critically” emerged as essential for decision-making and problem-solving across varied contexts, reinforcing the literature’s emphasis on critical thinking as a core professional asset (Goleman 1996). These transversal skills are consistent with the findings of Chamorro-Premuzic et al. (2010), who highlight the role of emotional intelligence in managing interpersonal dynamics effectively within workplace settings.

In a smaller proportion, occupation-specific skills constituted 6.81% of the total matched ESCO soft skills, with 16 competencies identified as essential for specialised roles but less transferable across broader contexts. Examples of these highly specialised competencies include “systems thinking”, “patient autonomy”, and “collaborate with coaching team”, each vital within niche roles requiring advanced technical knowledge. This limited representation aligns with Clarke (2018) findings on the need for specialised soft skills in roles where technical expertise and domain-specific knowledge take precedence over general adaptability.

Across categories, communication and problem-solving emerged as foundational skills that consistently appeared in various ESCO categories. Communication facilitates effective interpersonal interactions, conflict resolution, and team dynamics, critical for both internal collaboration and client-facing functions. Similarly, problem-solving fosters critical thinking and innovation, equipping individuals to address complex challenges, a crucial competency in fast-paced, evolving professional environments. The high prevalence of these competencies across categories underscores their centrality in fostering productivity and cohesion, aligning with studies by Davis-Blake et al. (2002) and Robles (2012) that emphasise these competencies as integral to both personal and professional success. Further supporting this alignment, transversal competencies such as “adapt to change” and “show initiative” highlight the importance of behavioral adaptability in dynamic work environments, echoing Heckman and Kautz (2012) findings on dispositional traits as core to soft skills. Skills associated to emotional intelligence highlighted by Goleman (1996), were also identified as soft skills, with labels such as “show empathy” and “exercise self-control”. These findings reinforce Chamorro-Premuzic et al. (2010) assertion

regarding the critical role of emotional intelligence in addressing both individual and relational challenges in the workplace,

The categorisation of soft skills within the ESCO framework demonstrates a balanced distribution across cross-sector, sector-specific, and occupation-specific competencies, with a focus on both universally applicable and context-specific skills. This structured classification strengthens ESCO's alignment with labor market requirements, supporting a blend of foundational and specialised soft skills essential for employability and professional development across a wide array of sectors.

Conclusions

This research represents a substantial advancement in standardising soft skills terminology and integrating them into established frameworks like ESCO, with significant implications for labour markets, education, and policymaking. It contributes to the ongoing dialogue highlighted by scholars such as Heckman and Kautz (2012) and Cinque and Ciappei (2014) on the necessity of consistent soft skill classification. A key outcome of this study is the development of a comprehensive collection of soft skills based on the ESCO taxonomy, providing a structured foundation for soft skills categorisation. The findings confirm the widespread relevance of soft skills, both cross-sectoral and sector-specific, underscoring their critical role in today's dynamic workforce. This corroborates earlier research on the transferability and adaptability of soft skills in various contexts (Clarke 2018; Succi 2019). As automation, artificial intelligence, and other advanced technologies reshape occupational landscapes, a distinct categorisation of soft skills within ESCO highlights these competencies alongside technical skills, which have traditionally dominated labor market frameworks. Similar to ESCO's existing "green" and "digital" skills labels, a distinct "soft skills" classification enhances their visibility and focus. This enhancement supports more targeted skill development initiatives, streamlines recruitment processes, and provides a standardised lexicon beneficial for both employers and job seekers. A clearer identification of transversal and sector-specific soft skills could also facilitate labour mobility, support career adaptability, and contribute to global standards for soft skills recognition.

Several recommendations emerge from this study to further refine the ESCO framework. First, expanding the emotional intelligence (EI) skill within ESCO would align with the growing importance of EI in professional contexts. While ESCO includes competencies like "show empathy" and "exercise self-control", the literature highlights a broader range of EI-related skills, such as managing interpersonal relationships and accurately perceiving others' emotions. Adding detailed labels for these competencies would enrich ESCO's EI representation and better meet workforce demands. As proposed by Fareri et al. (2021), automated identification of such nuanced skills can support HR professionals and educators in developing more accurate soft skill assessments, ensuring that vital interpersonal competencies are both identifiable and trainable across sectors. Another area for improvement involves the explicit contextualisation of soft skills across different sectors. ESCO could benefit from a clearer differentiation of how soft skills, such as communication and teamwork, are operationalised in distinct industries. For example, effective communication in healthcare requires empathy and patient-centered clarity, while in

customer service, it involves responsiveness and problem resolution. Refining ESCO to reflect these contextual variations would enhance its applicability and precision. Furthermore, while ESCO include references to creativity, the increasing emphasis on innovation suggests a more nuanced approach. The literature underscores creativity's importance not only in traditional arts but also in problem-solving and strategic roles across diverse sectors. ESCO could expand labels related to creative thinking, particularly within transversal and cross-sector skills, to capture the shift toward innovation-driven roles. Adaptability and the willingness to learn have also emerged as essential transversal competencies due to rapid industry evolution. As employees are required to continuously update their knowledge and skills, ESCO could incorporate more specific soft skills related to lifelong learning, self-assessment, and proactive self-development. This would ensure that the framework addresses the growing need for individuals to adapt to ongoing professional growth. Finally, while the study confirmed a semantic alignment between the soft skills identified in the literature and skills within ESCO, more refinement in terminology is needed. Terms like "problem-solving" and "critical thinking" have broad applications but vary in sectoral nuances. ESCO could consider adopting more precise language that captures both the overarching relevance of these skills and their specific sectoral manifestations.

While this study contributes to the clarity of soft skills within the ESCO framework, certain limitations should be acknowledged. Difficult access to relevant data constrained the literature review, as many articles were not open-access. Additionally, while NLP methods provided valuable insights, they encountered challenges in fully capturing contextual nuances due to the reliance on pre-trained models rather than a custom-trained model specific to soft skill analysis. This reliance on generalised data sources limited the model's precision, particularly in achieving high contextual relevance. Although pre-trained SBERT embeddings performed well, a custom-trained SBERT model tailored specifically for soft skills would likely enhance precision and alignment. Another methodological challenge involved establishing an optimal semantic similarity threshold to classify skills accurately; while the selected threshold yielded meaningful classifications, it introduced some subjectivity, potentially impacting the consistency of soft skill identification across diverse contexts. Moreover, the inherent subjectivity in defining soft skills across various disciplines, industries, and cultural contexts adds complexity to efforts toward standardised classification. Future research could refine this classification process by integrating more comprehensive soft skills datasets and by further exploring advanced customised NLP models that can capture contextual nuances with higher accuracy.

Despite these limitations, the study's findings offer valuable insights with practical applications across multiple sectors. The implications of these findings are broad. For academia, they contribute to an enriched understanding of soft skills, supporting curriculum design and skill development initiatives. HR professionals may leverage these insights to refine recruitment and evaluation practices, while entrepreneurs and managers can apply them to foster high-performing teams. Policymakers, in turn, can utilise these standardised definitions to better align educational and labor policies with market demands, ultimately facilitating more effective public policy and enhancing workforce resilience.

Declarations

Competing interests

The authors declare that they have no competing interests

Ethical approval

This article does not contain any studies with human participants performed by any of the authors.

Funding

No funding was received.

Authors' contributions

A.G. participated in data collection, conducted the literature review, drafted the main manuscript text, assisted in interpreting the results, and contributed to the revision process. C.P. participated in data processing, performed the analysis, interpreted the results, and contributed to writing and revising the document.

Availability of data and materials

The dataset of soft skills extracted from the literature (Annex 1) and the soft skills collection generated through the matching process (Annex 2) are provided as electronic supplementary files.

References

1. Ananiadou K, Claro M (2009) 21st Century Skills and Competences for New Millennium Learners in OECD Countries
2. Bak O, Jordan C (2018) The adoption of soft skills in supply chain and understanding its current role in supply chain management skills agenda: A UK perspective. Benchmarking Int J 26. https://doi.org/10.1108/BIJ-05-2018-0118
3. Balcar J (2016) Is it better to invest in hard or soft skills? Economic Labour Relations Rev 27:453–470. https://doi.org/10.1177/1035304616674613
4. Bennett N, Dunne E, Carré C (1999) Patterns of core and generic skill provision in higher education. High Educ (Dordr) 37:71–93. https://doi.org/10.1023/A:1003451727126
5. Chamorro-Premuzic T, Arteche A, Bremner AJ, Greven C, Furnham A (2010) Soft skills in higher education: importance and improvement ratings as a function of individual differences and

academic performance. Educ Psychol (Lond) 30:221–241. https://doi.org/10.1080/01443410903560278

6. Chiarello F, Fantoni G, Hogarth T, Giordano V, Baltina L, Spada I (2021) Towards ESCO 4.0–Is the European classification of skills in line with Industry 4.0? A text mining approach. Technol Forecast Soc Change 173:121177
7. Cinque M, Ciappei C (2014) Soft Skills per il governo dell'agire: la saggezza e le competenze prassico-pragmatiche. Franco Angeli, Milano
8. Clarke M (2018) Rethinking graduate employability: the role of capital, individual attributes and context. Stud High Educ 43:1923–1937. https://doi.org/10.1080/03075079.2017.1294152
9. Davis-Blake A, Moss P, Tilly C (2002) Stories Employers Tell: Race, Skill, and Hiring in America. Adm Sci Q 47:375. https://doi.org/10.2307/3094813
10. Dolce V, Emanuel F, Cisi M, Ghislieri C (2020) The soft skills of accounting graduates: perceptions versus expectations. Acc Educ 29:57–76. https://doi.org/10.1080/09639284.2019.1697937
11. European Commission: Green Skills and Knowledge Concepts: Labelling the ESCO classification (2022)
12. European Commission (2024) ESCO: European Skills, Competences, Qualifications and Occupations, https://esco.ec.europa.eu/en
13. European Union (2006) Recommendation of the European Parliament and of the Council of 18 December 2006 on key competences for lifelong learning. Official J Eur Union 394:10–18
14. Fareri S, Melluso N, Chiarello F, Fantoni G (2021) SkillNER: Mining and mapping soft skills from any text. Expert Syst Appl 184:115544
15. Goleman D (1996) Emotional intelligence. Why it can matter more than IQ. Learning 24:49–50
16. González-Eras A, Aguilar J (2019) Determination of Professional Competencies Using an Alignment Algorithm of Academic Profiles and Job Advertisements, Based on Competence Thesauri and Similarity Measures. Int J Artif Intell Educ 29:536–567. https://doi.org/10.1007/s40593-019-00185-z
17. Haselberger D, Oberhuemer P, Perez E, Cinque M, Capasso F (2012) Mediating Soft Skills at Higher Education Institutions: Guidelines for the design of learning situations supporting soft skills achievement
18. Heckman JJ, Kautz T (2012) Hard evidence on soft skills. Labour Econ 19:451–464. https://doi.org/10.1016/J.LABECO.2012.05.014
19. Khaouja I, Kassou I, Ghogho M (2021) A Survey on Skill Identification From Online Job Ads. IEEE Access 9:118134–118153. https://doi.org/10.1109/ACCESS.2021.3106120
20. Khaouja I, Mezzour G, Carley KM, Kassou I (2019) Building a soft skill taxonomy from job openings. Soc Netw Anal Min 9:43. https://doi.org/10.1007/s13278-019-0583-9
21. Knight P, Page A (2007) The assessment of 'wicked'competences
22. Matteson ML, Anderson L, Boyden C (2016) Soft skills': A phrase in search of meaning. portal: Libr Acad 16:71–88

23. Papoutsoglou M, Mittas N, Angelis L (2017) Mining People Analytics from StackOverflow Job Advertisements. In: 2017 43rd Euromicro Conference on Software Engineering and Advanced Applications (SEAA). pp. 108–115
24. Robles MM (2012) Executive perceptions of the top 10 soft skills needed in today's workplace. Bus communication Q 75:453–465
25. Siekmann G, Fowler C (2017) Identifying Work Skills: International Approaches. Discussion Paper. National Centre for Vocational Education Research (NCVER)
26. Succi C (2019) Are you ready to find a job? Ranking of a list of soft skills to enhance graduates' employability. Int J Hum Resour Dev Manage 19:281–297
27. Toner P (2011) Workforce skills and innovation: An overview of major themes in the literature
28. Tsirkas K, Chytiri A-P, Bouranta N (2020) The gap in soft skills perceptions: A dyadic analysis. Educ + Train 62:357–377
29. World Health Organization (1994) Life skills education for children and adolescents. schools. Pt. 1, Introduction to life skills for psychosocial competence. Pt. 2, Guidelines to facilitate the development and implementation of life skills programmes. World Health Organization
30. Yorke M (2006) Employability in Higher Education: What It Is, What It Is Not
31. Zhang M, Jensen KN, van der Goot R, Plank B (2022) Skill extraction from job postings using weak supervision. arXiv preprint arXiv:2209.08071

Annex 2

Annex 2 is not available with this version.

Figures

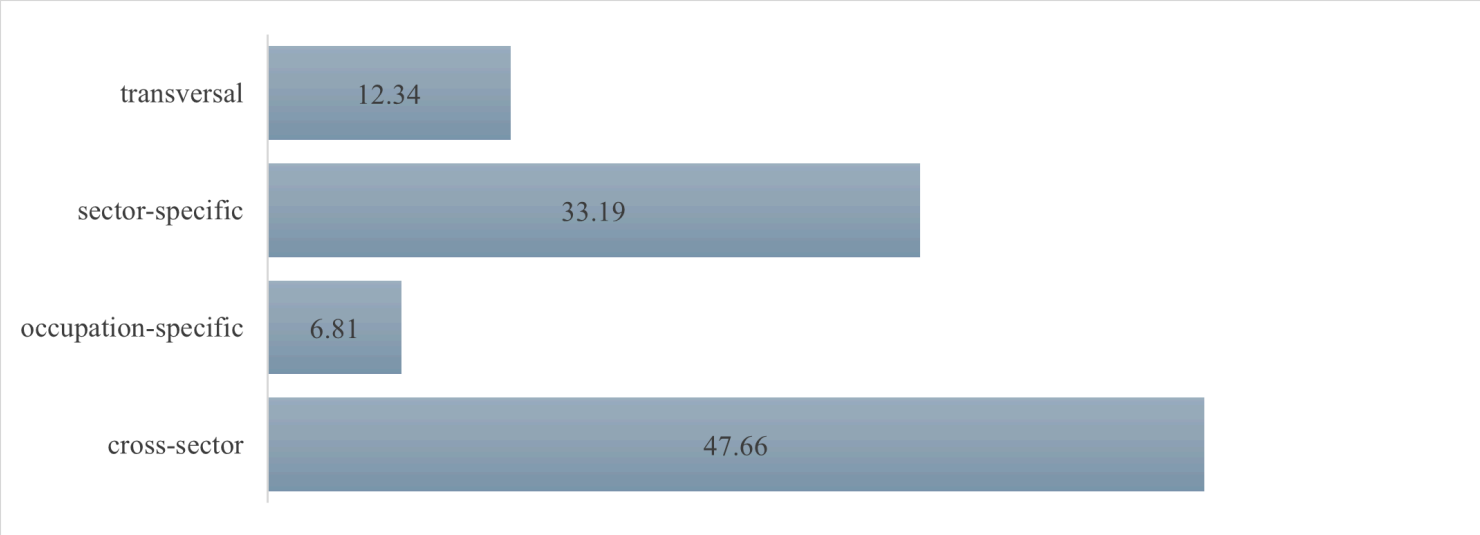

Figure 1

Distribution of soft skills among ESCO skill by reuse level

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Annex1Listofsoftskillsfromliterature.xlsx
- Annex3Generatedsoftskillscollection.xlsx