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.

Mapping Skill Demand and Occupational Interdependencies in Online Job Postings: A Sentence-BERT and Vector Space Approach

Karolina Bolesta

kboles@sgh.waw.pl

SGH Warsaw School of Economics

Research Article

Keywords: Technological forecasting, Skill demand, Sentence-BERT, ESCO taxonomy, Online job vacancies, Natural Language Processing

Posted Date: November 9th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11112475/v1

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

Additional Declarations: No competing interests reported.

Mapping Skill Demand and Occupational Interdependencies in Online Job Postings: A Sentence-BERT and Vector Space Approach

Abstract

The accelerating pace of technological change, automation and artificial intelligence is profoundly restructuring global labour markets, shifting the focus of demand-side analyses from static occupational titles toward dynamic and multifaceted skill portfolios. Traditional macroeconomic monitoring instruments, such as employer surveys and administrative registries, frequently suffer from significant reporting latency, aggregate oversimplification and rigid taxonomic classifications that fail to capture emerging competencies in real time. This study investigates the structural interdependencies and statistical dispersion of demanded skills across nine major occupational groups by analyzing a massive repository of web-scraped job postings from leading commercial recruitment platforms in Poland (SOJO dataset, 2019-2023). Leveraging Natural Language Processing, raw vacancy descriptions were embedded using a deep contextual Sentence-BERT architecture and mapped into the European Skills, Competences, Qualifications and Occupations (ESCO) taxonomy via Facebook AI Similarity Search (FAISS) dense vector indexing at a strict similarity threshold. In addition, the distribution, concentration and dispersion of 74 medium-level skill concepts were assessed through box-plot parameters, interquartile ranges and outlier analyses. The empirical findings confirm that skill demand varies significantly across occupational categories in both type and frequency, while revealing strong cross-occupational penetration of digital literacy, operational coordination and transversal capabilities into traditionally manual occupations. These results provide actionable insights for technological forecasting, curriculum modernization and data-driven active labour market policies.

Keywords: Technological forecasting; Skill demand; Sentence-BERT; ESCO taxonomy; Online job vacancies; Natural Language Processing.

1. Introduction

The accelerating diffusion of digital technologies, artificial intelligence, machine learning algorithms and automated systems is profoundly transforming the global socioeconomic landscape, fundamentally altering how productive tasks are coordinated and redefining employer expectations across the entire occupational spectrum (Acemoglu & Restrepo, 2018a, 2020; D. Autor & Salomons, 2018; Brynjolfsson & McAfee, 2014). As economic production transitions toward Industry 4.0, smart manufacturing and data-driven organizational models, the operational composition of jobs becomes increasingly dynamic (Gonzalez Ehlinger & Stephany, 2024; Schedlbauer et al., 2021). Rather than eliminating entire occupational categories in a single stroke, modern technological change substitutes human labour in routine, codifiable activities while complementing non-routine cognitive, analytical and interpersonal tasks (D. Autor et al., 2003; Bessen, 2016, 2019). Consequently, the analytical perspective in labour economics and technological forecasting has shifted away from rigid, static occupational titles toward granular, flexible portfolios of human capabilities and competencies (Acemoglu & Autor, 2010; Gonzalez Ehlinger & Stephany, 2024).

In this rapidly evolving environment, conventional labour market monitoring instruments, such as standardized macroeconomic surveys, census registries and periodic administrative reporting, frequently face critical structural constraints (Beręsewicz et al., 2021; Handel, 2003). These traditional tools are characterized by substantial administrative publication lags, high collection costs, aggregate oversimplification and rigid categorization systems that cannot capture newly emerging capabilities in real time (Beręsewicz et al., 2021). By the time a new technological specialization is formally integrated into an official occupational classification, recruitment practices in the private sector have already transformed significantly (Data et al., 2024; Senger et al., 2024). Therefore, capturing the real-time dynamics of unmet labour demand requires leveraging non-traditional, granular data sources generated continuously by the digital recruitment ecosystem (Khaouja et al., 2021; Romanko & O’Mahony, 2022).

The rapid digitalization of the recruitment landscape has established online job boards and corporate career portals as the primary transmission channel for labour market signals, yielding massive repositories of web-scraped online job advertisements (SOJO) that directly articulate immediate operational demands (Tsvetkova et al., 2024). Employers utilize online vacancies to specify the precise mix of professional knowledge, tool proficiencies, certifications and soft skills required for open positions (Rios et al., 2020). However, utilizing online job postings for systematic socioeconomic forecasting presents formidable analytical hurdles. Vacancies are published in unstructured, heterogeneous natural language formats characterized by variable phrasing, contextual nuances, company-specific jargon and a lack of standardized descriptive taxonomies (Guetterman et al., 2018). Classical keyword matching and basic bag-of-words text mining algorithms are fundamentally limited in their capacity to disambiguate semantic intent, resolve polysemy or capture complex interdependencies among demanded competencies (Khaouja et al., 2021; Varelas et al., 2022).

To address these empirical limitations, recent advancements in computational linguistics and Natural Language Processing (NLP), specifically deep contextual sentence transformers and dense vector indexing, provide robust methodologies for mapping unstructured natural language text into standardized international ontologies, such as the European Skills, Competences, Qualifications and Occupations (ESCO) framework(Chiarello et al., 2021; Data et al., 2024; Senger et al., 2024). Transformer architectures, such as Sentence-BERT (SBERT), project text sequences into high-dimensional vector spaces where geometric proximity corresponds to semantic equivalence, enabling the automated extraction and categorization of competencies regardless of lexical variability (SBERT, n.d.).

The primary purpose of this research is to investigate the structural composition, co-occurrence patterns and statistical dispersion of demanded skills across major occupational groups by analyzing large-scale unstructured vacancy data from commercial recruitment platforms in Poland. This study addresses the central question of how skill requirements are distributed and interconnected within the modern digital recruitment landscape, testing the hypothesis that the type and frequency of demanded skills vary significantly across major occupational categories while

demonstrating distinct cross-occupational penetration of transversal and digital capabilities.

To fulfill this objective, the research executes a multi-stage empirical pipeline. First, an asynchronous web-scraping architecture retrieves unstructured job advertisements from leading commercial recruitment platforms in Poland (SOJO dataset), paired with the granular hierarchical taxonomy of the ESCO Skills pillar (S-pillar) (ESCO, n.d.). Second, a semantic embedding framework utilizing Sentence-BERT (all-mpnet-base-v2) in conjunction with Facebook AI Similarity Search (FAISS) vector indexing is deployed to extract and standardize 296 discrete low-level skill concepts from raw vacancy texts at a strict similarity threshold (Meta AI, n.d.; SBERT, n.d.). Third, the distribution, concentration and dispersion of 74 medium-level skill groups are evaluated using box-plot metrics, interquartile ranges and outlier identification to reveal emergent specialized requirements and hybridization within formal job roles. Finally, the empirical outcomes are synthesized to provide strategic insights into technological forecasting, workforce adaptability and data-driven educational policy design.

2. Theoretical Foundation and Literature Review

2.1. The Nature of Technological Change, Productivity and Automation

Technological change represents one of the most fundamental socioeconomic challenges and opportunities of the modern era (Van Vliet et al., 2021). In classical and neoclassical economics, technological progress is conceptualized as an outward shift in the production function, allowing organizations to achieve higher output levels using identical or fewer inputs, thereby expanding total factor productivity (TFP) and driving long-term economic growth (Baumol, 1967; Brynjolfsson & McAfee, 2014; Francis et al., 2020; Stoneman & Battisti, 2010). Technical progress serves as the primary catalyst for the historical evolution from agrarian and industrial models toward the modern knowledge economy (Van Doorn & van Vliet, 2024). Depending on the theoretical framework, technological advancements are treated either as exogenous shocks, occurring independently of market mechanisms, or as endogenous responses driven by economic incentives, institutional frameworks and capital accumulation (Acemoglu, 2002; Cole, 2023).

Historically, major technological breakthroughs have fundamentally altered economic structures, beginning with the steam engine during the First Industrial Revolution, electricity and assembly line manufacturing during the Second, and the development of computing and the Internet during the Third (Greenwood, 1997; Jevons, 1931; Mohajan, 2019; Taalbi, 2019). The Fourth Industrial Revolution extends beyond the mere continuation of digital computing, introducing autonomous physical-cyber systems, advanced robotics and cognitive artificial intelligence that actively participate in decision-making and problem-solving (Acemoglu & Restrepo, 2018b; Badet, 2021). While early automation tools were entirely managed and directed by human operators, modern AI systems demonstrate increasing degrees of autonomy, expanding their operational scope into complex cognitive domains previously deemed uniquely human (Grace et al., 2018; Leslie & Perini, 2024; Sarker, 2021).

The diffusion and economic realization of technological advancements depend heavily on the interaction between internal technical capabilities and external institutional environments (Evenson & Westphal, 1995; Lai et al., 2020; Swargiary, 2024). The transition from an internal invention developed within research laboratories to a widespread external standard requires substantial time and capital investment, often leading to a productivity lag or retardation period between the initial introduction of a technology and its measurable impact on macroeconomic productivity (David, 1990; Mokyr et al., 2015). In economic history, the introduction of labor-saving machinery has frequently generated societal anxiety and fears of technological unemployment, dating back to the Luddite protests of the early nineteenth century (Mayer & Oosthuizen, 2021; Mokyr et al., 2015; Solomon, 1987). John Maynard Keynes popularized the concept of technological unemployment in the 1930s, characterizing it as a temporary frictional phase wherein the economization of labor outpaces the discovery of new uses for human effort (McGaughey, 2022). Similarly, Wassily Leontief drew historical analogies between the replacement of horses by automobiles and the potential displacement of human labor by computers, formalizing input-output models to illustrate how technical improvements in one industry ripple across the wider economy (Leontief, 1983, 1986).

In modern economic literature, the drivers of technological change are traditionally analyzed through two complementary hypotheses: technology-push and demand-pull

(Hötte, 2023; Mowery & Rosenberg, 1979). The technology-push perspective emphasizes scientific discoveries, engineering breakthroughs and autonomous technological opportunities as the primary drivers of innovation (Klevorick et al., 1995). Conversely, the demand-pull hypothesis posits that market demand, consumer preferences and emerging societal needs direct the focus and velocity of inventive activities (Acemoglu, 2002; Mowery & Rosenberg, 1979). Price mechanisms and market size play decisive roles in determining whether organizations adopt innovations, as firms evaluate the marginal profitability and return on investment associated with substituting human labor with automated systems (Acemoglu, 2002; Arrow, 1972).

The macroeconomic impact of technological change on the workforce can be categorized into factor-neutral and factor-biased productivity shifts (Gao et al., 2023; Violante, 2008). Neutral factor productivity enhances the efficiency of all inputs equally, whereas biased factor productivity disproportionately augments or substitutes specific production factors, fundamentally altering the relative value of different skill sets and labor categories (Acemoglu, 1998). Technology operates along a continuum between labor-augmenting (complementing human workers and elevating their productivity) and labor-saving (directly substituting human tasks with machines) (Sears, 2017; Ziesemer, 2023). Although automation eliminates specific routine tasks, the reduction in production costs and subsequent price declines frequently expand aggregate market demand, creating a reinstatement effect that generates entirely new tasks and job roles across the economic system (Acemoglu & Restrepo, 2018a; Hötte, 2023).

2.2. The Task-Based Model, Polarization and Routinization

To understand the precise mechanisms through which technological change restructures the labor market, modern economic theory relies on the task-based framework formalized by David Autor and Daron Acemoglu (Acemoglu & Autor, 2010; D. Autor, 2013). The task-based model uncouples the concept of a job or occupation from the specific capabilities of an individual worker, defining a job as a bundle of distinct productive activities or tasks, while human capital is defined as a set of skills applied to execute those tasks (Acemoglu & Autor, 2010; Poletaev & Robinson, 2008).

Tasks are traditionally classified into a two-by-two matrix: routine versus non-routine, and cognitive versus manual (D. Autor et al., 2003). Routine tasks are defined as activities that follow precise, codifiable programmatic steps and standardized rules that can be fully specified in computer software or executed by automated machinery (D. Autor et al., 2003; Bessen, 2016). Examples include repetitive bookkeeping, administrative filing, standardized calculations and assembly line material handling. In contrast, non-routine tasks require contextual problem-solving, abstract reasoning, creativity, pattern recognition and complex interpersonal communication (non-routine cognitive), or sensory-motor adaptability, visual perception and physical flexibility in unstructured environments (non-routine manual) (D. Autor et al., 2003; Spitz-Oener, 2006).

The evolution of labor market demand under technological change has been analyzed through two successive theoretical frameworks: Skill-Biased Technical Change (SBTC) and Routine-Biased Technical Change (RBTC). The SBTC framework, prominent during the late twentieth century, argued that computerization raised the relative productivity of highly skilled, tertiary-educated workers, thereby widening wage differentials between college and non-college workers (Acemoglu, 1998; Katz & Murphy, 1992; Machin, 2004). Under SBTC, technological adoption was viewed as monotonically skill-favoring, increasing the demand for skilled labor across all sectors (Battisti et al., 2022; Guo et al., 2024).

However, empirical developments in the late 1990s and 2000s revealed that employment growth was not strictly linear, leading to the establishment of the RBTC hypothesis (D. H. Autor et al., 2006; Goos & Manning, 2007). RBTC posits that computerization and automation directly substitute middle-skill occupations that are heavily routine-intensive, such as clerical office support, precision craft trades and routine manufacturing operations, while complementing high-skill non-routine cognitive roles (e.g., managerial, professional and engineering occupations) and leaving low-skill non-routine manual roles relatively unaffected (e.g., personal care, food services and cleaning) (Fernández-Macías & Hurley, 2017; Goos et al., 2014, 2021).

This dynamic produces job polarization, characterized by a U-shaped pattern of employment growth across the wage and skill distribution, wherein employment shares expand at the top and bottom tiers while hollowing out middle-tier positions (D. H. Autor

et al., 2006; Goos & Manning, 2007; Rutledge & Guan, 2015; Sebastian, 2018). As routine middle-skill positions disappear, displaced workers face a bifurcation: individuals with adaptable human capital transition into higher-paying non-routine cognitive occupations, whereas those lacking advanced capabilities are pushed into low-wage non-routine manual service positions, exacerbating overall earnings inequality (Acemoglu & Restrepo, 2018a; D. H. Autor, 2019).

Job polarization has been empirically documented across numerous developed economies, including the United States, Western Europe and Nordic countries (Acemoglu, 2002; Asplund et al., 2011; Dabla-Norris et al., 2019; Goos et al., 2014; Siena & Zago, 2024). In Central and Eastern European (CEE) economies, including Poland, labor market restructuring has displayed unique transitional features (Ł. Arendt & Grabowski, 2019; Hardy et al., 2016a, 2016b). During the post-socialist transformation and subsequent integration into the European Union, Poland experienced a structural shift characterized by a rapid contraction of agricultural employment and an expansion of routine cognitive positions in shared service centers, administrative hubs and modern manufacturing facilities (Hardy et al., 2016b; Kucharski & Kukulak-Dolata, 2022). Analyses by (Ł. Arendt & Grabowski, 2019) and (L. Arendt et al., 2023) indicate that between 2004 and 2016, Poland exhibited a positive wage premium for non-routine cognitive jobs and wage penalties for routine roles, reflecting the broader European polarization trajectory while maintaining high persistent demand for specific technical and manual competencies linked to industrial subcontracting and foreign direct investment.

2.3. Competence Taxonomies, Employability and the Digital Divide

As technological transformation accelerates, traditional proxies for human capital, such as years of formal schooling or broad occupational titles, become insufficient for measuring true workplace capability (Allen & Van der Velden, 2001; Cappelli, 2015; Handel, 2003). Modern labor market research increasingly focuses on the multidimensional definition of skills, conceptualized as the capacity to apply knowledge, know-how and social proficiencies to execute complex work activities (Attewell, 1990; Council of the European Union, 2018; Green, 2011).

To standardize the representation of human capital, international bodies and research institutions have developed comprehensive competence frameworks, such as the Occupational Information Network (O*NET) in the United States, the European Skills, Competences, Qualifications and Occupations (ESCO) framework in the European Union, and the Polish Human Capital Balance (BKL) (Gruza, 2018; National Centre for O*NET Development, n.d.; PARP, 2015).* The ESCO taxonomy organizes the labor market into multi-layered pillars, categorizing skills into high-level, medium-level and low-level groups that encompass specialized domain knowledge, practical abilities and transversal competencies (ESCO, n.d.; Gruza, 2018).

A fundamental distinction in competence research is drawn between job-specific skills and transversal (transferable) skills (Hortigüela-Alcalá et al., 2019; Nägele & Stalder, 2017). Job-specific competencies represent narrow, specialized technical abilities tied to particular operational processes, tools or legal standards (e.g., operating specific CNC machinery or configuring specialized network protocols) (Grugulis & Stoyanova, 2011; Panos et al., 2014). Transversal skills, by contrast, represent cross-cutting capabilities, such as critical thinking, communication, problem-solving, digital literacy, self-organization and teamwork, that are applicable across a broad spectrum of occupations and industries (Andrews & Higson, 2008; Van Laar et al., 2020). Transversal skills enhance an individual's overall employability, defined as the ongoing capacity to obtain, maintain and transition between employment opportunities throughout a career (De Cuyper et al., 2010; Harvey, 2001; Peeters et al., 2019; Van der Heijden et al., 2009).

In the context of the digital transformation, digital competencies have evolved from niche technical requirements into foundational transversal capabilities necessary for participation in the modern knowledge society (Anderson, 2008; Bejaković & Mrnjavac, 2020; Carlisle et al., 2023; Ozkan-Ozen & Kazancoglu, 2022). The literature on the digital divide illustrates an evolution across multiple levels: from the first level (basic physical access to hardware and the Internet), through the second level (skills and digital literacy required to utilize software effectively), to the third level (the capacity to translate digital engagement into tangible socioeconomic benefits and productivity outcomes) (Rogers, 2001; Scheerder et al., 2017; Vasilescu et al., 2020). With the emergence of generative AI and automated decision systems, a fourth digital divide is

developing around the ability to effectively collaborate with, prompt and evaluate artificial intelligence systems in everyday workplace environments (Fulton et al., 2022; Grace et al., 2018; Leslie & Perini, 2024). In Poland, empirical studies utilizing PIAAC and national survey data have demonstrated significant regional and demographic disparities in digital skills, highlighting that educational modernization and lifelong learning are essential for preventing digital exclusion (Drabowicz, 2014; Jakubowski, 2015; Palczyńska & Świst, 2018; Śledziewska & Włoch, 2015).

2.4. Big Data, Online Job Portals and NLP in Demand-Side Analysis

Understanding real-time shifts in skill demand requires empirical methodologies capable of processing large-scale, demand-side information directly from the hiring market (Doogan, 2005; Romanko & O'Mahony, 2022). While administrative registries and statistical surveys (such as the Polish Demand for Labour survey) provide representative benchmarks, they often suffer from extensive publication delays and aggregate job requirements into broad occupational codes that obscure specific technical expectations (Beręsewicz et al., 2021, 2024).

The rapid growth of commercial online recruitment portals has created extensive opportunities for Big Data analytics in labour economics (Brenčič & Norris, 2010; Schedlbauer et al., 2021; Stevenson, 2008). Online job advertisements capture real-time employer demand, articulating explicit requirements for competencies, certifications, software tools and behavioral traits (Beręsewicz et al., 2024; Khaouja et al., 2021; Rios et al., 2020). Research utilizing online vacancy corpora across international labor markets has highlighted the accelerating demand for hybrid skill sets that combine domain expertise with digital fluency and interpersonal coordination (Acemoglu et al., 2022; Hershbein & Kahn, 2018; Saussay et al., 2022; Verma et al., 2022).

To address these challenges, computational linguistics has advanced toward transformer-based neural language models, such as BERT (Bidirectional Encoder Representations from Transformers) and its specialized variant Sentence-BERT (SBERT) (SBERT, n.d.). SBERT utilizes siamese and triplet network structures to generate semantically rich, fixed-dimensional dense vector embeddings of sentences and paragraphs. In vector space, semantic similarity between text fragments

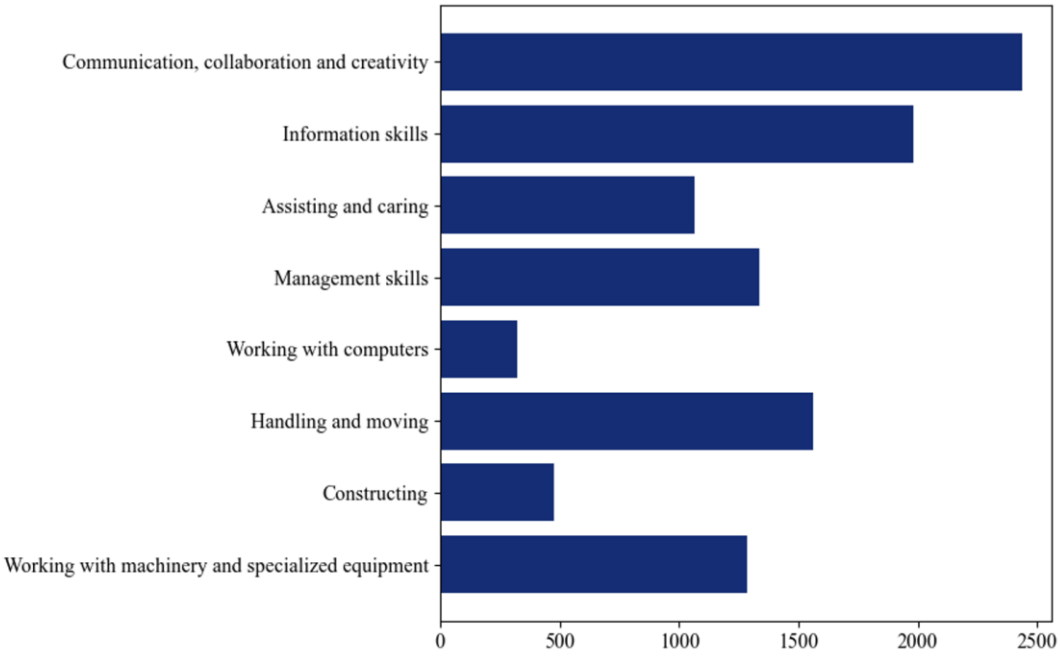
corresponds directly to geometric proximity, such as cosine similarity or Euclidean distance (Bishop & Bishop, 2023). To deploy these dense neural embeddings across millions of vacancy descriptions, high-performance vector indexing engines are required (Meta AI, n.d.). Facebook AI Similarity Search (FAISS) provides an open-source library optimized for fast similarity search and clustering of dense high-dimensional vectors, leveraging advanced indexing architectures and GPU acceleration to perform nearest-neighbor queries across billions of vector comparisons in milliseconds. Combining SBERT contextual embeddings with FAISS vector indexing establishes a robust, scalable architecture for mapping unstructured recruitment text onto formal classification ontologies, such as ESCO, providing a highly reliable foundation for technological forecasting and empirical skill demand analysis (Data et al., 2024; Senger et al., 2024).

3. Data and Empirical Methodology

3.1. Datasets and Processing Architecture

The empirical architecture of this study integrates two rich data sources: the official ESCO taxonomy and the System of Online Job Offers (SOJO) database.

The ESCO framework, established by the European Commission, provides a standardized, multi-lingual ontology of the European labor market, comprising 13,939 skills categorized within four overarching pillars (ESCO, n.d.; European Commission, 2021). This research focuses exclusively on the "S - Skills" pillar, which captures position-related skills, knowledge and competences. The S-pillar is organized into three hierarchical levels: 8 high-level skill domains, 74 medium-level skill groups and 296 low-level discrete skill concepts. Each low-level concept is mapped to multiple granular narrower abilities that provide detailed textual descriptions of practical capabilities, averaging 34.2 narrower abilities per concept. Figure 1 illustrates the distribution of narrower abilities across the eight high-level ESCO skill domains, showing strong depth in communication, information processing, handling and management.



[Figure 1: The number of narrower abilities in the high-level ESCO skills classification]
(Source: Own elaboration based on ESCO S-pillar data).

The job vacancy dataset is derived from the SOJO repository developed by the Institute of Labour and Social Studies in Warsaw. SOJO aggregates job postings scraped continuously from the leading online recruitment portals in Poland, including pracuj.pl, olx.pl, praca.pl, gratka.pl, careerjet.pl and gazetapraca.pl. The extracted records cover the observation window from January 2019 to October 2023. Each raw record contains metadata, portal identifiers, occupational codes and raw advertisement URLs. Records lacking a valid assigned major occupational group under the Polish Classification of Occupations and Specializations (KZiS), which is harmonized directly with the International Standard Classification of Occupations (ISCO-08), were excluded from the analysis (representing 12.3% of the initial dataset). The final analytical dataset comprises 4,836,024 unique online vacancy records spanning nine major occupational groups (excluding military personnel).

To retrieve the full textual content of advertisements from the stored URLs, an asynchronous web scraping engine was constructed in Python using aiohttp, asyncio, BeautifulSoup and regular expressions. The script implemented concurrency throttling using semaphores to manage server loads and prevent rate-limiting exceptions. Raw HTML pages were retrieved, parsed and processed through a text-cleaning pipeline

that removed non-ASCII characters, HTML tags, special formatting characters, punctuation and Polish stop-words, followed by lowercase conversion and whitespace normalization.

3.2. Semantic Embedding and FAISS Vector Search Architecture

To extract skill requirements from unstructured text without relying on rigid keyword lookups, a semantic embedding and vector indexing framework was deployed using Sentence-BERT (all-mpnet-base-v2) and FAISS (Meta AI, n.d.; SBERT, n.d.). The all-mpnet-base-v2 model is a transformer architecture fine-tuned on over one billion sentence pairs to optimize semantic representations, achieving top-tier performance on standard semantic textual similarity benchmarks (SBERT, n.d.).

The semantic matching pipeline operated in four sequential stages:

First, all 296 low-level ESCO skill concepts and their associated narrower abilities were transformed into dense contextual vector embeddings. Each skill text s_k was projected into a 768-dimensional continuous vector space:

$$E(s_k) = \text{SBERT}(s_k) \in \mathbb{R}^{768} \qquad (1)$$

Second, the complete set of skill embeddings $\mathcal{S} = \{E(s_1), E(s_2), \dots, E(s_K)\}$ was structured into an optimized FAISS vector database $\mathcal{I}$ designed for accelerated exact and approximate nearest-neighbor retrieval (Meta AI, n.d.):

$$\mathcal{I} = \text{FAISS_Index}(\mathcal{S}) \qquad (2)$$

Third, for each cleaned job offer text J_i, the dense text embedding $E(J_i) \in \mathbb{R}^{768}$ was computed using the identical SBERT model. The FAISS index executed vector distance queries to compute the Euclidean distance between the vacancy embedding and all indexed skill embeddings:

$$d(s_k, J_i) = \|E(s_k) - E(J_i)\|_2 = \sqrt{\sum_{m=1}^{768} (E(s_k)_m - E(J_i)_m)^2} \qquad (3)$$

Fourth, Euclidean distances were normalized into confidence percentage similarity scores:

$$\text{Score}(s_k, J_i) = 100 - d(s_k, J_i) \qquad (4)$$

To ensure high semantic precision and eliminate spurious matches, a strict similarity threshold of 99% was applied. All skill matches meeting or exceeding this cutoff were retained and transformed into binary presence indicators ($1 =$ present, $0 =$ absent). These identified concepts were subsequently mapped upward into the 74 medium-level and 8 high-level ESCO skill domains for statistical analysis.

3.3. Statistical and Visualization Methodology

The empirical distribution and structural interdependencies of extracted skills were evaluated using two statistical methodology across the nine major occupational groups. To evaluate the concentration, spread and specialized requirements within occupational categories, box-plot distributions were generated for the 74 medium-level ESCO skill groups across all nine major occupational groups. Each box plot characterizes the median (Q_2), the 25th percentile (Q_1), the 75th percentile (Q_3), the interquartile range (IQR $= Q_3 - Q_1$), upper and lower whiskers (defined as $1.5 \times$ IQR from the quartiles), and individual outlier points falling beyond the whiskers. This granular visualization identifies core operational baselines, high-variability competencies and extreme outlier job postings that reflect role hybridization and advanced technical specialization.

4. Empirical Results

4.1. Baseline Skill Demand in Polish Commercial Portals: The BKL Framework

Before examining deep ESCO semantic matches, the baseline frequency of skill requirements was assessed using the 11 skill categories defined under the Polish Human Capital Balance (BKL) framework across the nine major occupational groups.

In Group 1 (Managers), recruitment postings are heavily dominated by self-organizational skills (approaching 300,000 occurrences), interpersonal skills (nearly 295,000 occurrences), technical proficiencies (exceeding 300,000 occurrences) and artistic/creative capabilities (over 260,000 occurrences).

In Group 2 (Professionals), technical skills represent the single largest category (exceeding 930,000 occurrences), followed closely by interpersonal skills (over 870,000 occurrences), self-organization (over 860,000 occurrences) and artistic/creative capabilities (over 810,000 occurrences).

In Group 3 (Technicians and Associate Professionals), interpersonal competencies lead with over 740,000 occurrences, closely accompanied by technical skills (over 710,000 occurrences), self-organization (over 660,000 occurrences) and artistic/creative abilities (nearly 600,000 occurrences).

In Group 4 (Clerical Support Workers), interpersonal skills (nearly 490,000 occurrences), technical proficiencies (over 450,000 occurrences) and self-organization (over 440,000 occurrences) dominate job requirements. Artistic/creative skills (over 350,000 occurrences) and office competencies (over 200,000 occurrences) are also prominent.

In Group 5 (Service and Sales Workers), interpersonal communication represents the dominant requirement (exceeding 420,000 occurrences), accompanied by technical skills (over 330,000 occurrences), self-organization (over 320,000 occurrences) and creative abilities (nearly 290,000 occurrences).

In Group 6 (Skilled Agricultural, Forestry and Fishery Workers), while overall online vacancy volume is lower due to recruitment channels operating offline, interpersonal skills (nearly 5,000 occurrences), self-organization (over 2,700 occurrences), technical skills (over 2,600 occurrences) and office competencies (over 2,400 occurrences) constitute the most demanded capabilities.

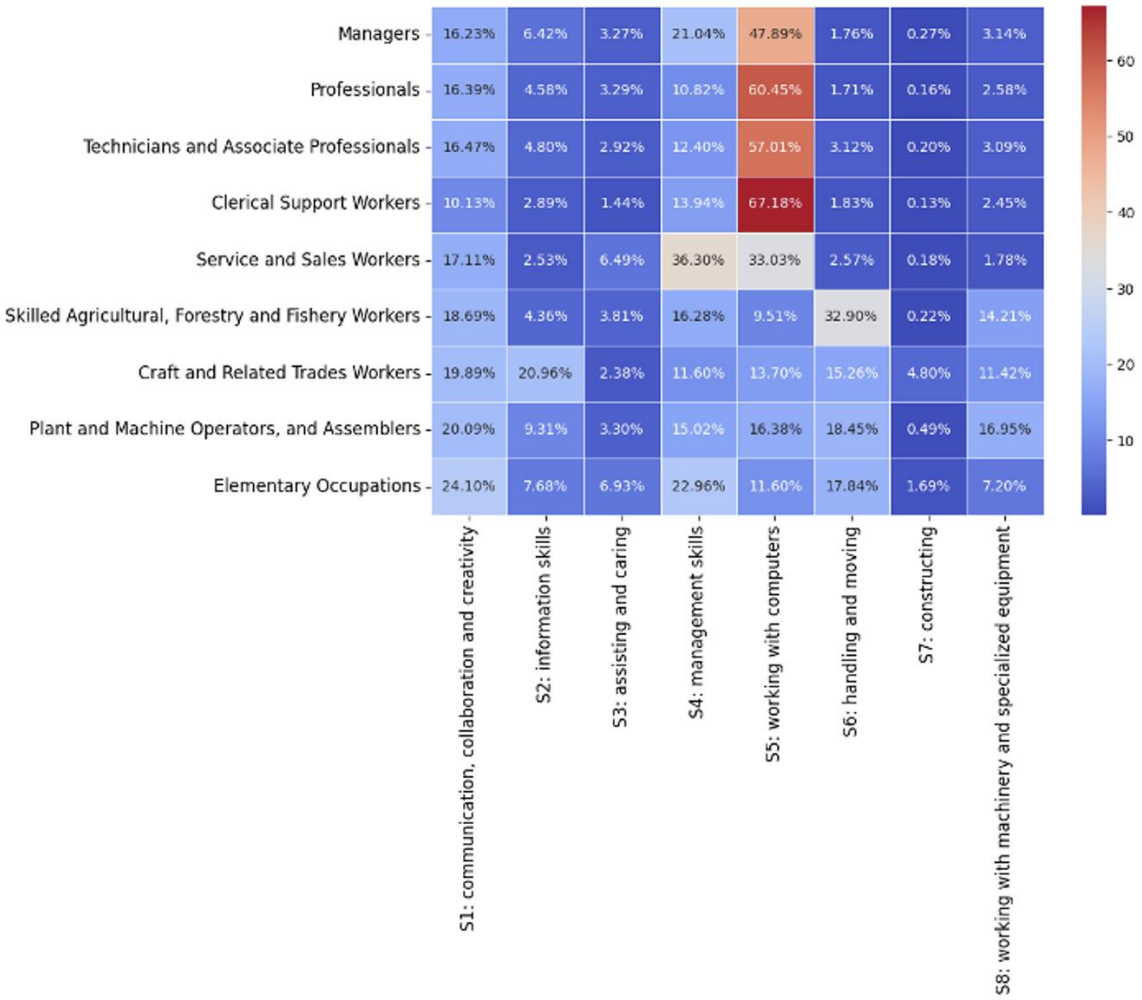
In Group 7 (Craft and Related Trades Workers), interpersonal skills lead with over 370,000 occurrences, followed by technical proficiencies (nearly 340,000 occurrences), self-organization (over 300,000 occurrences) and artistic abilities (over 250,000 occurrences).

In Group 8 (Plant and Machine Operators and Assemblers), technical capabilities represent the dominant skill category (over 230,000 occurrences), closely accompanied by interpersonal skills (nearly 200,000 occurrences), self-organization (over 160,000 occurrences) and artistic/creative execution (over 140,000 occurrences).

In Group 9 (Elementary Occupations), interpersonal skills are the most frequently recorded category (nearly 390,000 occurrences), alongside technical proficiencies (over 230,000 occurrences), self-organization (over 230,000 occurrences) and creative execution (over 170,000 occurrences). Office skills appear in approximately 150,000 postings, while mathematical skills account for over 85,000 occurrences.

4.2. Aggregate ESCO Skill Allocation across Major Occupational Groups

Figure 2 presents the aggregate percentage shares of the eight high-level ESCO skill domains across the nine major occupational groups in the SOJO dataset, illustrating the distribution of requirements identified through the semantic matching pipeline.



[Figure 2: The share of high-level ESCO skills among the major occupational group]
(Source: Own elaboration based on SOJO dataset).

The empirical distribution illustrates distinct patterns of competence concentration and diffusion across the occupational hierarchy. Working with computers (S_5) represents the single most dominant skill requirement across white-collar occupations, accounting for 67.18% of all identified skills among Clerical Support Workers, 60.45% among Professionals, 57.01% among Technicians and Associate Professionals, and 47.89% among Managers. Digital literacy is also heavily required in Service and Sales positions (33.03%) and maintains substantial double-digit presence in traditionally manual groups, including Plant and Machine Operators (16.38%), Craft Trades (13.70%) and Elementary Occupations (11.60%).

Management skills (S_4) represent the primary category in Service and Sales roles (36.30%), while accounting for 22.96% in Elementary Occupations, 21.04% in Managers, 16.28% in Agricultural Workers, 15.02% in Plant Operators, and 13.94% in

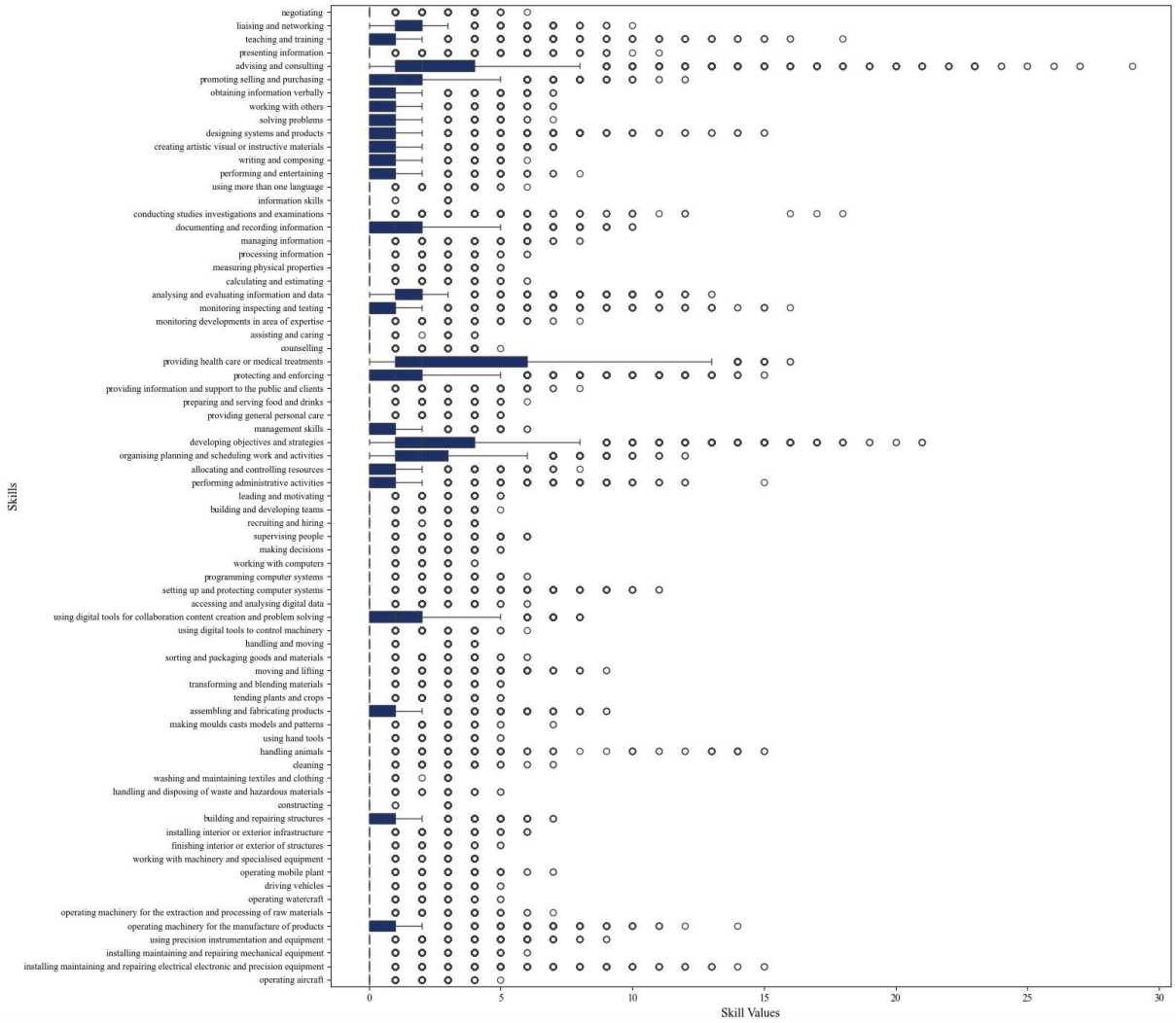
Clerical Support roles. This high prevalence reflects employer demand for workflow management, resource allocation, team supervision and customer-oriented coordination across both white-collar and operational positions.

Physical handling and machinery operations (S_6 and S_8) concentrate primarily within manual and industrial occupations. Handling and moving (S_6) accounts for 32.90% of requirements in Skilled Agricultural Workers, 18.45% in Plant and Machine Operators, 17.84% in Elementary Occupations, and 15.26% in Craft and Related Trades. Working with machinery and specialized equipment (S_8) accounts for 16.95% in Plant Operators, 14.21% in Agricultural Workers, and 11.42% in Craft Trades. Information skills (S_2) reach their highest concentration in Craft and Related Trades (20.96%), reflecting technical reading, blueprint interpretation and material specification requirements. Constructing skills (S_7) remain heavily specialized, peaking at 4.80% in Craft Trades.

Communication, collaboration and creativity (S_1) maintains a remarkably consistent and substantial share across the entire occupational spectrum, ranging from 10.13% in Clerical Support Workers to 24.10% in Elementary Occupations, averaging approximately 18% across all groups. This steady cross-occupational presence confirms that transversal communicative and collaborative abilities operate as universal baseline requirements throughout the modern recruitment market.

4.3. Granular Skill Dispersion and Outlier Dynamics of 74 Medium-Level ESCO Skills (Box Plots)

To assess intra-group variation, competency spread and role specialization, the distribution of the 74 medium-level ESCO skill groups was analyzed using box plots across all nine major occupational groups. Figures 3 through 11 illustrate the box plots for Groups 1 to 9.

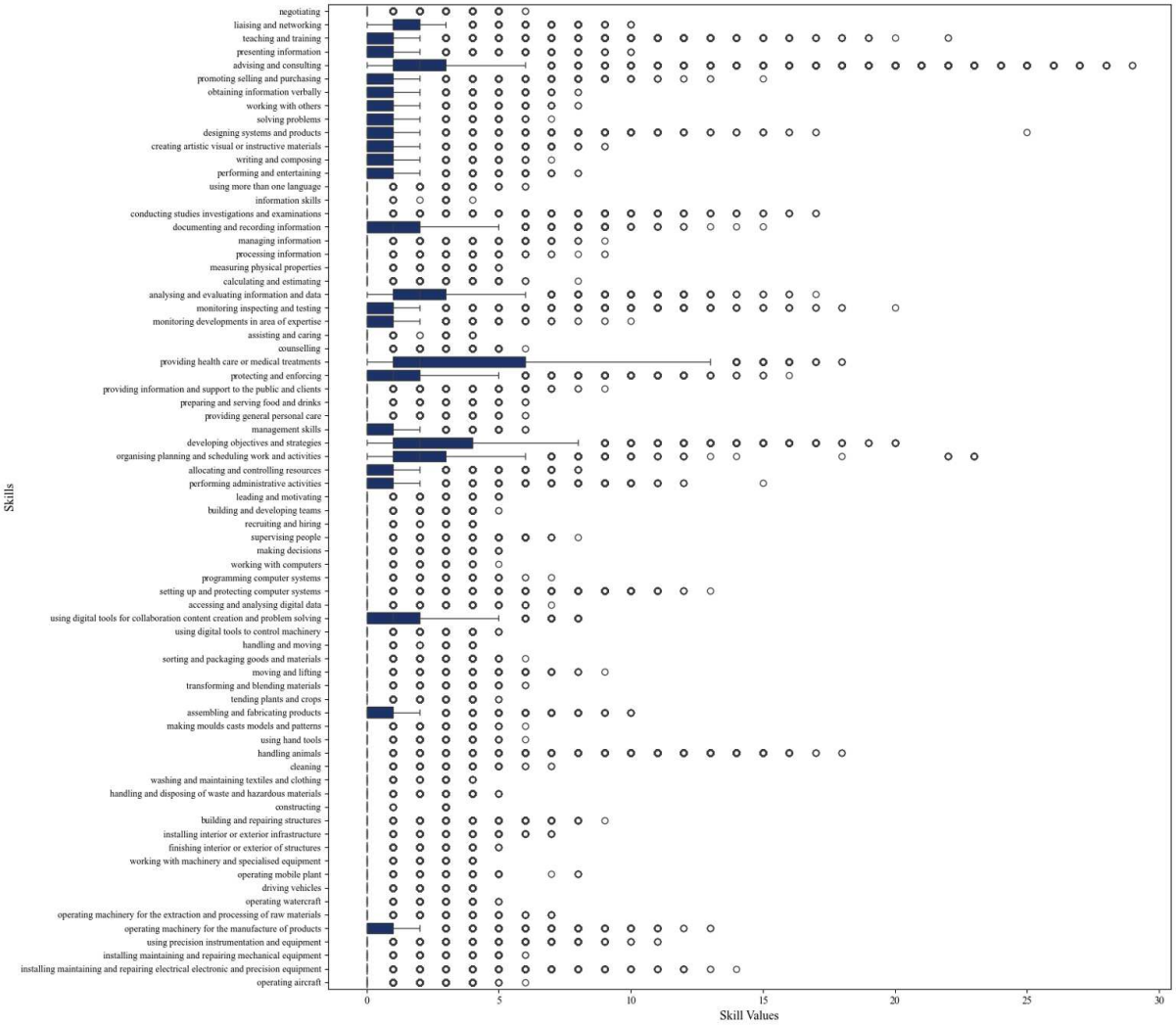


[Figure 3: Box-plot distribution analysis of 74 medium-level ESCO skills across 1st major occupational group – Managers]

(Source: Own elaboration).

In Group 1 (Managers), the skill distribution displays substantial variation across key strategic, advisory and healthcare-related competencies. The largest interquartile range occurs in *providing health care or medical treatments*, spanning from $Q_1 = 1$ to $Q_3 = 6$ (IQR $= 5$), with whiskers reaching 13.5 and extreme outliers at 14 and 15, reflecting specialized medical directorships and healthcare management roles. *Developing objectives and strategies* and *advising and consulting* both exhibit an IQR of 3 ($Q_1 = 1, Q_3 = 4$), with upper whiskers extending to 8.5 and maximum values reaching 21 and 29, respectively, indicating that while strategic planning is central to management, certain advisory roles require extraordinary depth. *Organizing, planning and scheduling work activities* has an IQR of 2 ($Q_1 = 1, Q_3 = 3$), with whiskers reaching 6 and outliers at 7, 8 and 12. *Protecting and enforcing* exhibits an IQR from 0 to 2, with

whiskers reaching 5 and outliers up to 15, capturing compliance and regulatory leadership roles.

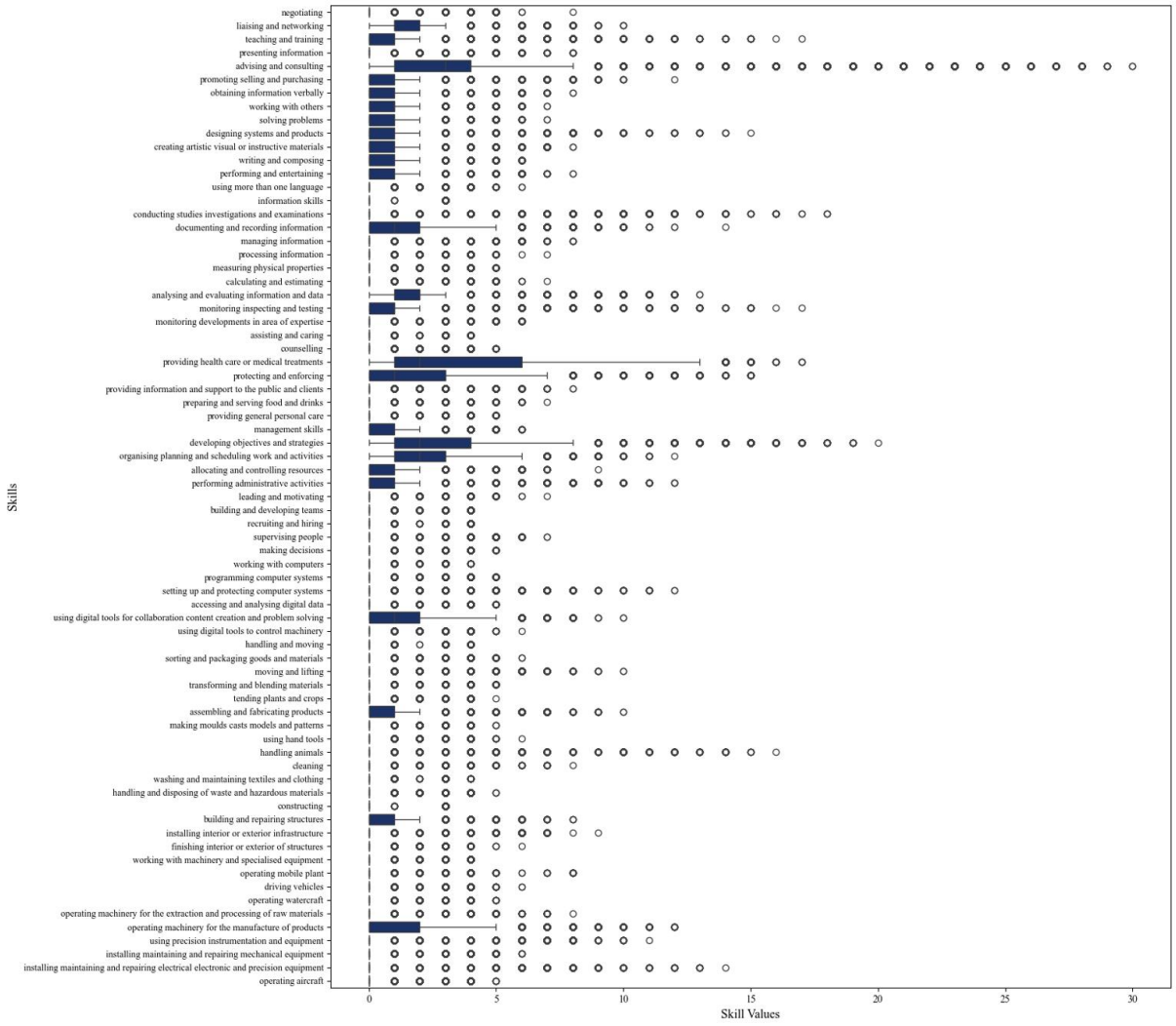


[Figure 4: Box-plot distribution analysis of 74 medium-level ESCO skills across 2nd major occupational group – Professionals]

(Source: Own elaboration).

In Group 2 (Professionals), *providing health care or medical treatments* exhibits the widest spread, with an IQR of 5 ($Q_1 = 1, Q_3 = 6$), an upper whisker at 13.5 and outliers reaching 18, identifying specialized medical and clinical professionals. *Developing objectives and strategies* has an IQR of 3 ($Q_1 = 1, Q_3 = 4$), whiskers extending to 8.5 and outliers clustering around 9 to 20. *Advising and consulting* displays an IQR of 2 ($Q_1 = 1, Q_3 = 3$), an upper whisker at 6 and extreme outliers reaching 29, highlighting high-level consulting and domain expert positions. *Organizing, planning and scheduling work activities* has an IQR of 2, whiskers extending to 6 and maximum

values reaching 23. *Protecting and enforcing* demonstrates an IQR from 0 to 2, whiskers to 5 and outliers up to 16, reflecting legal, inspection and compliance professions.

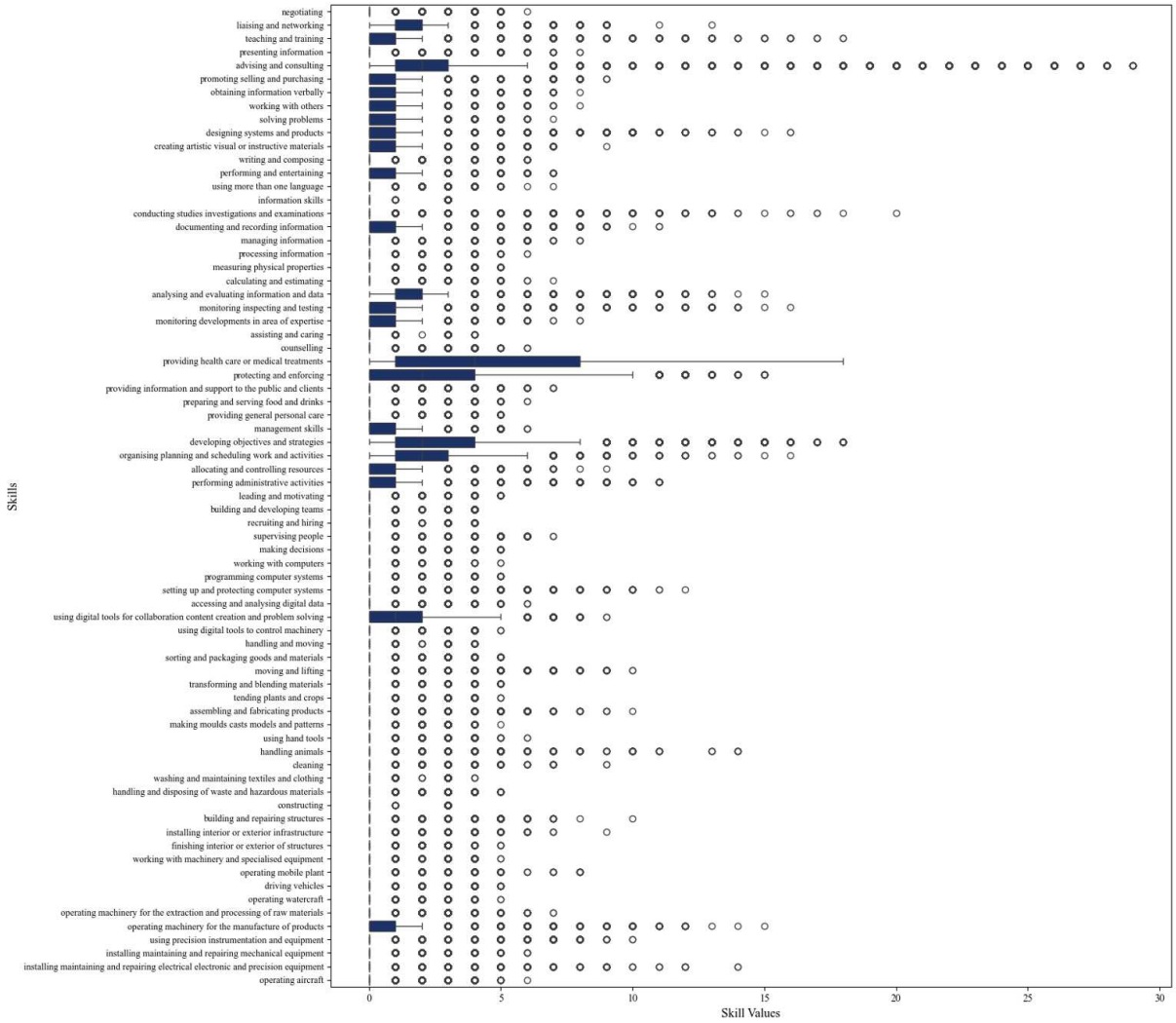


[Figure 5: Box-plot distribution analysis of 74 medium-level ESCO skills across 3rd major occupational group – Technicians and Associate Professionals]

(Source: Own elaboration).

In Group 3 (Technicians and Associate Professionals), *providing health care or medical treatments* spans from $Q_1 = 1$ to $Q_3 = 6$ (IQR = 5), with whiskers reaching 13.5 and outliers at 14, 15 and 17, reflecting emergency medical technicians and clinical laboratory personnel. *Developing objectives and strategies* shows an IQR of 3 ($Q_1 = 1, Q_3 = 4$), whiskers to 8.5 and outliers reaching 20. *Protecting and enforcing* displays an IQR of 3 ($Q_1 = 0, Q_3 = 3$), whiskers extending to 7.5 and outliers up to 15, capturing technical inspectors and occupational safety technicians. *Advising and*

consulting exhibits an IQR of 3 ($Q_1 = 1, Q_3 = 4$), whiskers to 8.5 and extreme outliers reaching 30, reflecting technical sales advisors and specialized systems consultants. *Operating machinery for the manufacture of products* exhibits an IQR of 2 ($Q_1 = 0, Q_3 = 2$), whiskers to 5 and outliers at 6, 7, 8 and 12.

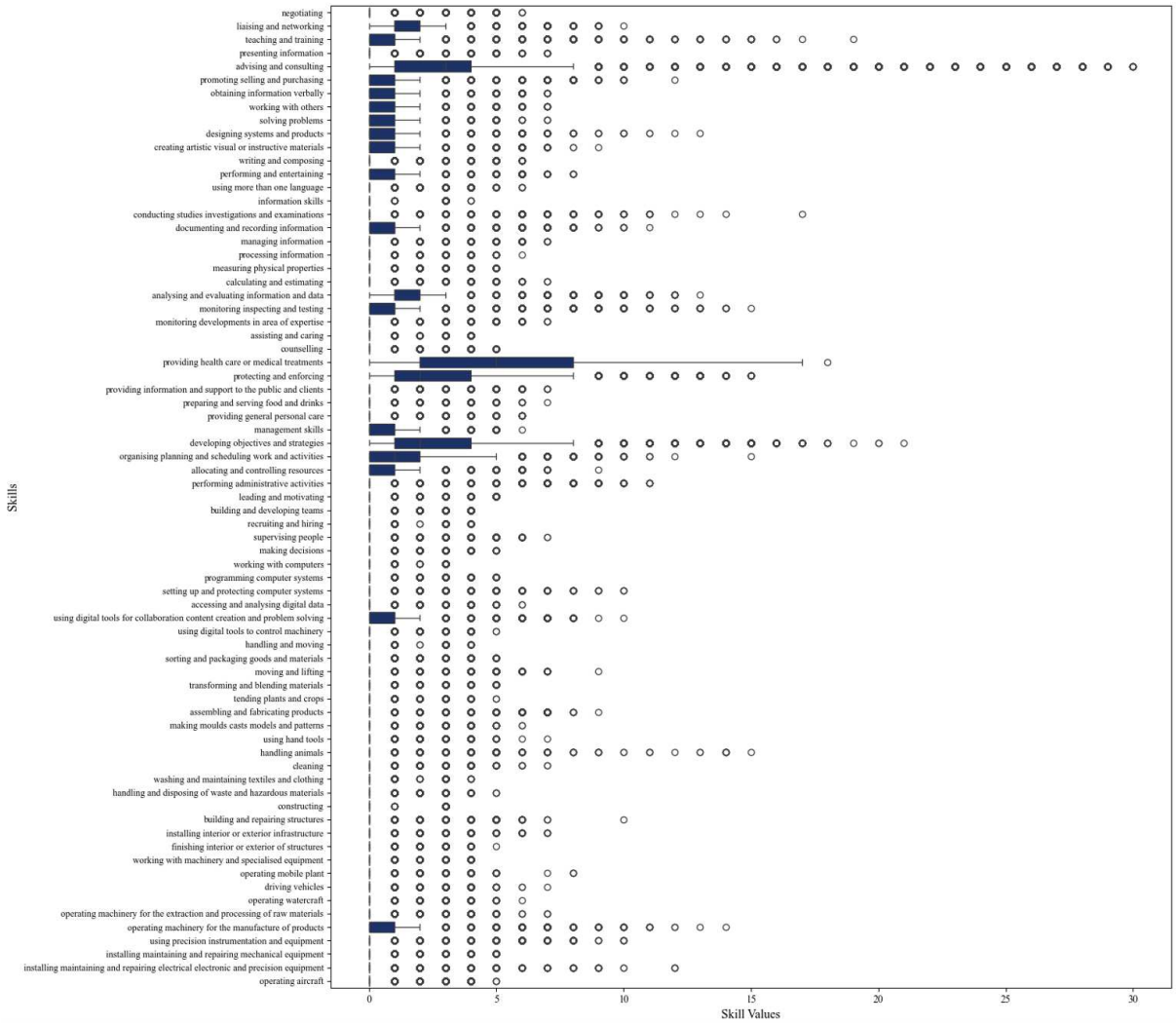


[Figure 6: Box-plot distribution analysis of 74 medium-level ESCO skills across 4th major occupational group – Clerical Support Workers]

(Source: Own elaboration).

In Group 4 (Clerical Support Workers), *providing health care or medical treatments* records the largest IQR of 7 ($Q_1 = 1, Q_3 = 8$), with whiskers extending from 0 to 18, capturing medical registrars and hospital administrative staff. *Protecting and enforcing* displays an IQR of 4 ($Q_1 = 0, Q_3 = 4$), whiskers extending to 10 and outliers at 11, 14 and 15, reflecting legal documentation clerks and compliance administrators. *Developing objectives and strategies* has an IQR of 3 ($Q_1 = 1, Q_3 = 4$), whiskers to 8.5

and outliers up to 18. *Organizing, planning and scheduling work activities* has an IQR of 2 ($Q_1 = 1, Q_3 = 3$), whiskers to 6 and outliers at 7, 8, 9 and 16. *Using digital tools for collaboration* presents an IQR of 2 ($Q_1 = 0, Q_3 = 2$), whiskers to 5 and outliers at 6, reflecting specialized data management and remote office coordination roles.

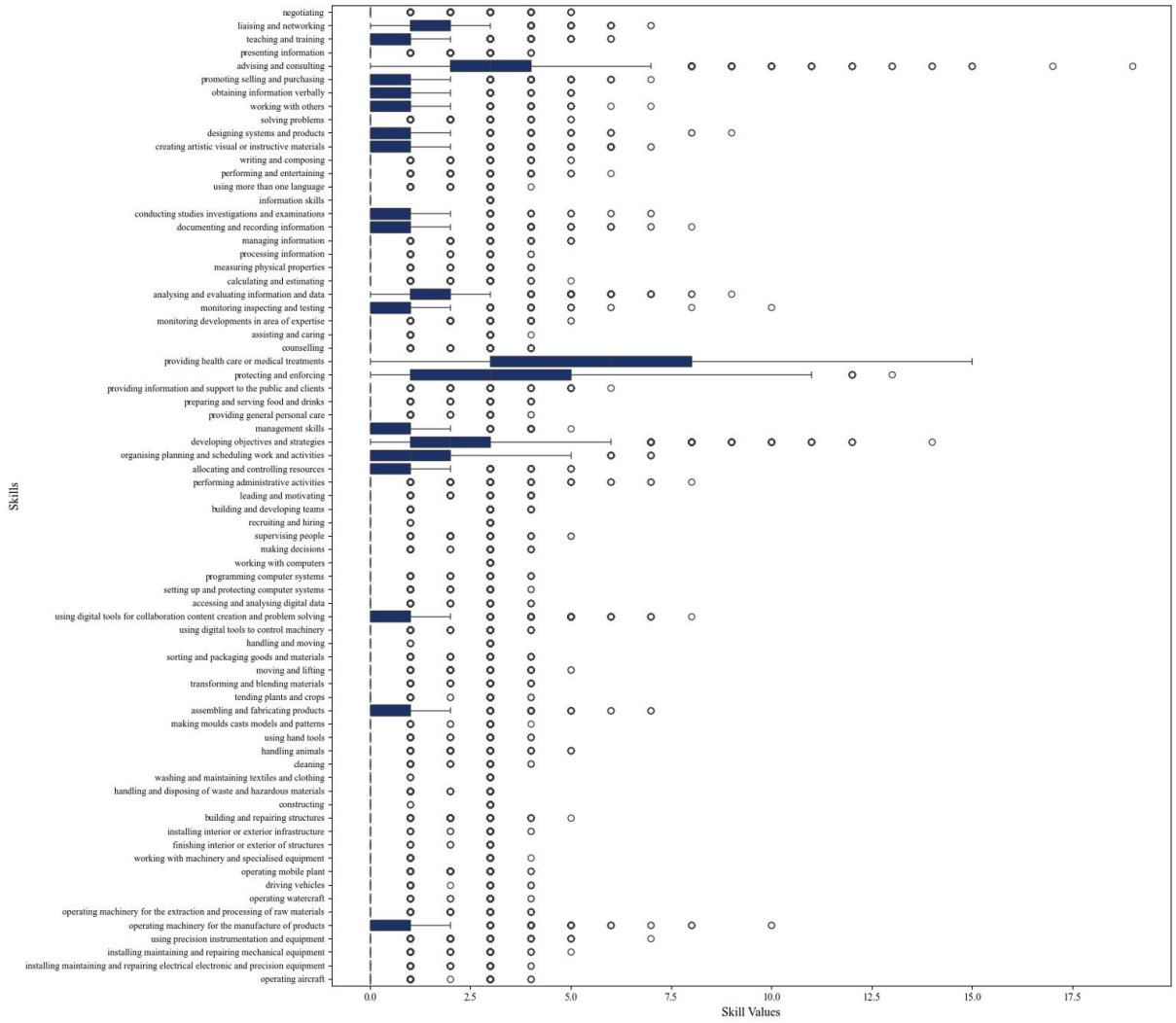


[Figure 7: Box-plot distribution analysis of 74 medium-level ESCO skills across 5th major occupational group – Service and Sales Workers]

(Source: Own elaboration).

In Group 5 (Service and Sales Workers), *providing health care or medical treatments* displays an IQR of 6 ($Q_1 = 2, Q_3 = 8$), whiskers extending to 17 and an outlier at 18, capturing pharmacy assistants, care attendants and medical sales representatives. *Developing objectives and strategies*, *protecting and enforcing*, and *advising and consulting* all display identical IQRs of 3 ($Q_1 = 1, Q_3 = 4$), with whiskers reaching 8.5. Outliers in *advising and consulting* span from 9 to 21 with an extreme maximum at 30,

identifying high-end client advisors and technical sales consultants. *Protecting and enforcing* displays outliers from 9 to 15, reflecting security services and consumer compliance staff. *Organizing, planning and scheduling work activities* has an IQR of 2 ($Q_1 = 0, Q_3 = 2$), whiskers to 5 and outliers at 6, 7, 8 and 15, reflecting retail and event coordination roles.

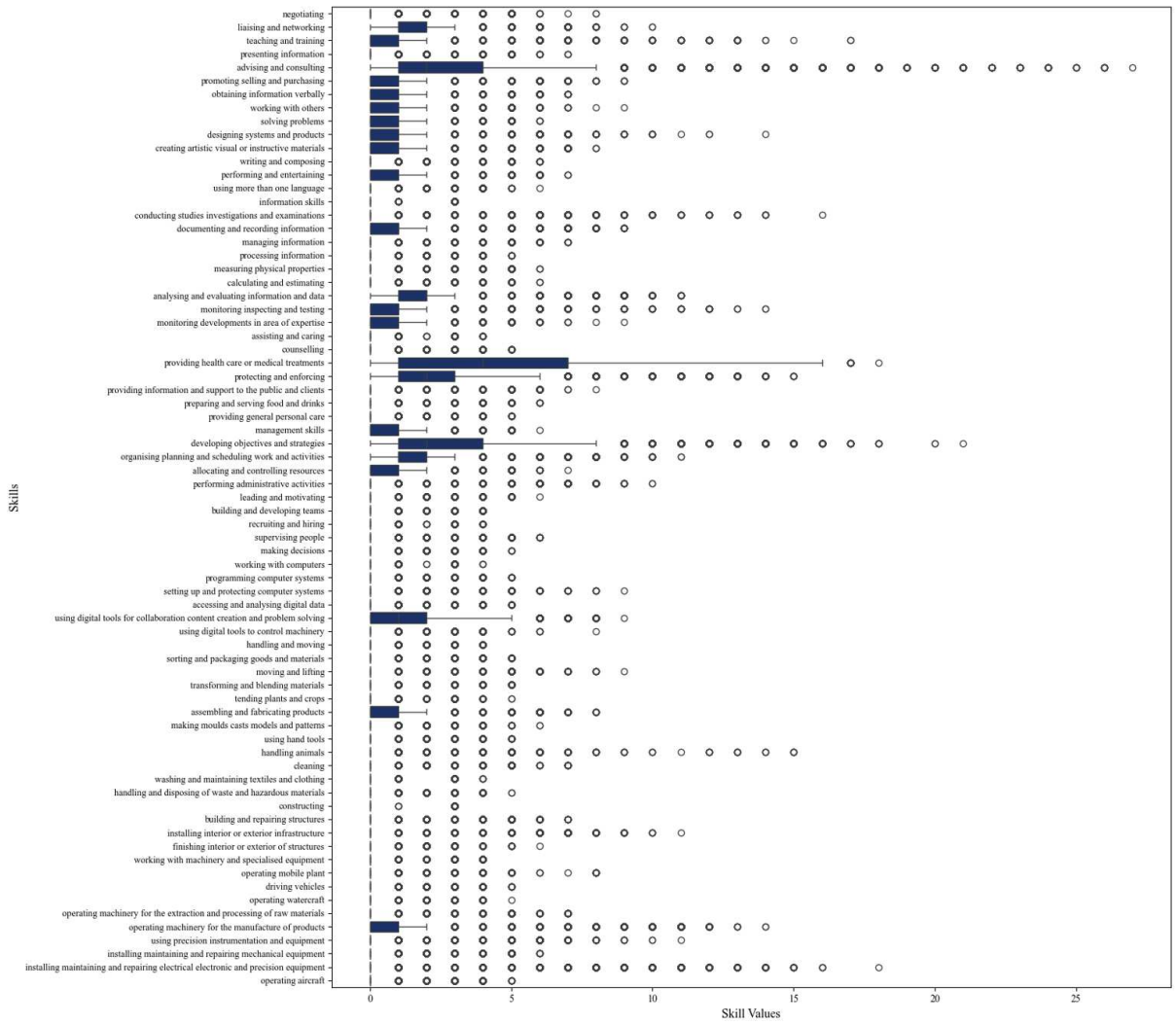


[Figure 8: Box-plot distribution analysis of 74 medium-level ESCO skills across 6th major occupational group – Skilled Agricultural, Forestry and Fishery Workers]

(Source: Own elaboration).

In Group 6 (Skilled Agricultural, Forestry and Fishery Workers), *providing health care or medical treatments* has an IQR of 5 ($Q_1 = 3, Q_3 = 8$), with whiskers extending to 15, reflecting veterinary care, animal husbandry and agricultural first aid. *Protecting and enforcing* displays an IQR of 4 ($Q_1 = 1, Q_3 = 5$), whiskers to 11 and outliers at 12 and 13, capturing forestry wardens, environmental inspectors and agricultural regulation

compliance officers. *Developing objectives and strategies* and *advising and consulting* both have an IQR of 2 ($Q_1 = 1, Q_3 = 3$ and $Q_1 = 2, Q_3 = 4$), with whiskers reaching 6 and 7, and outliers reaching 14 and 19, respectively, identifying farm managers and agricultural consultants. *Organizing, planning and scheduling work activities* has an IQR of 2, whiskers to 5 and outliers at 6 and 7.

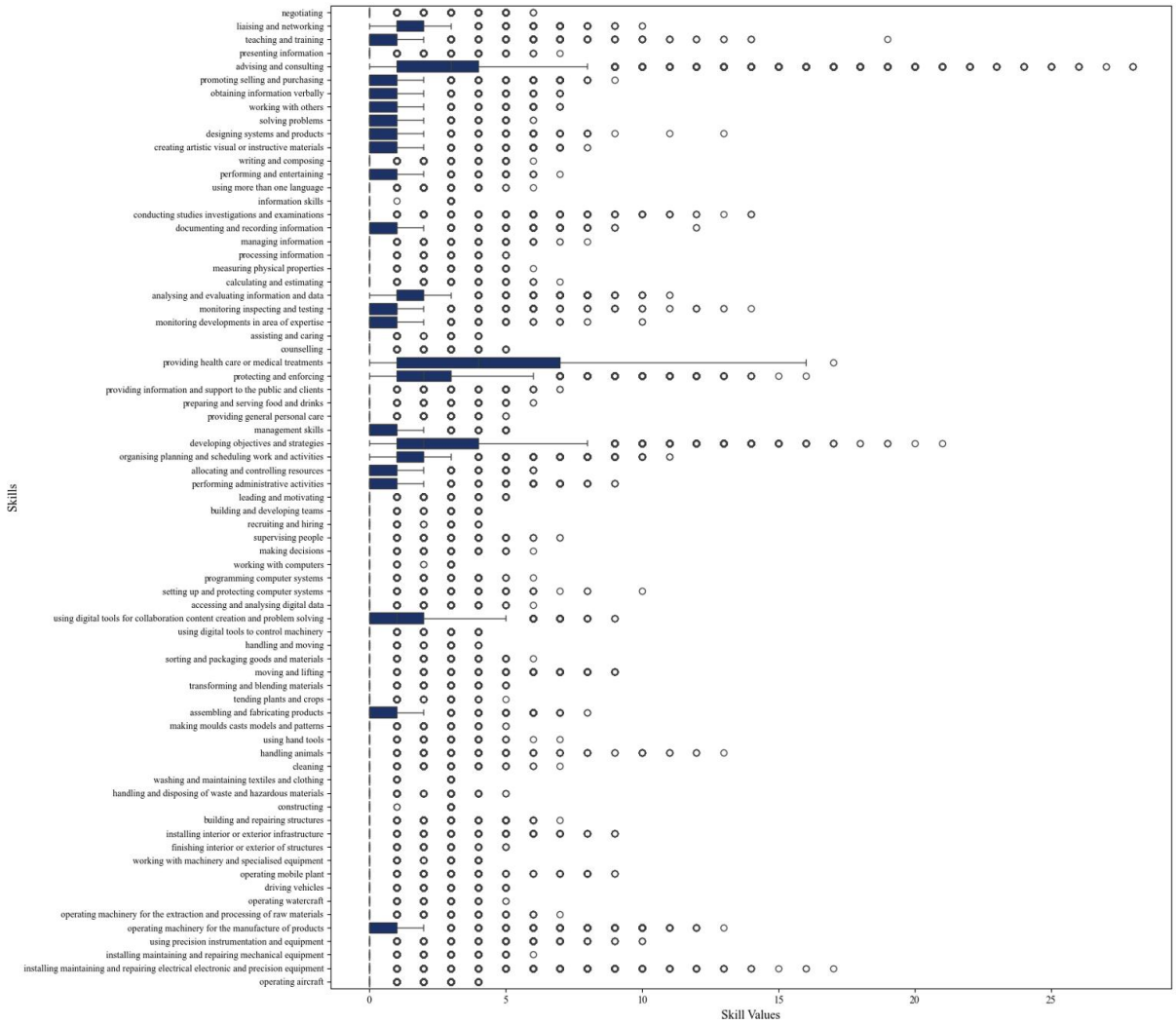


[Figure 9: Box-plot distribution analysis of 74 medium-level ESCO skills across 7th major occupational group – Craft and Related Trades Workers]

(Source: Own elaboration).

In Group 7 (Craft and Related Trades Workers), *providing health care or medical treatments* records an IQR of 6 ($Q_1 = 1, Q_3 = 7$), with whiskers reaching 16 and outliers at 17 and 18, reflecting industrial safety officers and emergency response personnel in hazardous trades. *Developing objectives and strategies* has an IQR of 3 ($Q_1 = 1, Q_3 = 4$), whiskers to 8.5 and outliers from 9 to 21, capturing master craftsmen and

trade contractors. *Advising and consulting* displays an IQR of 3 ($Q_1 = 1, Q_3 = 4$), whiskers to 8.5 and extreme outliers reaching 27, reflecting specialized technical trade consultants. *Using digital tools for collaboration* has an IQR of 2 ($Q_1 = 0, Q_3 = 2$), whiskers to 5 and outliers at 6, 7 and 9, capturing digital fabrication and CAD-enabled craft roles. *Protecting and enforcing* exhibits an IQR of 2, whiskers to 6 and outliers at 7, 8, 11 and 15.

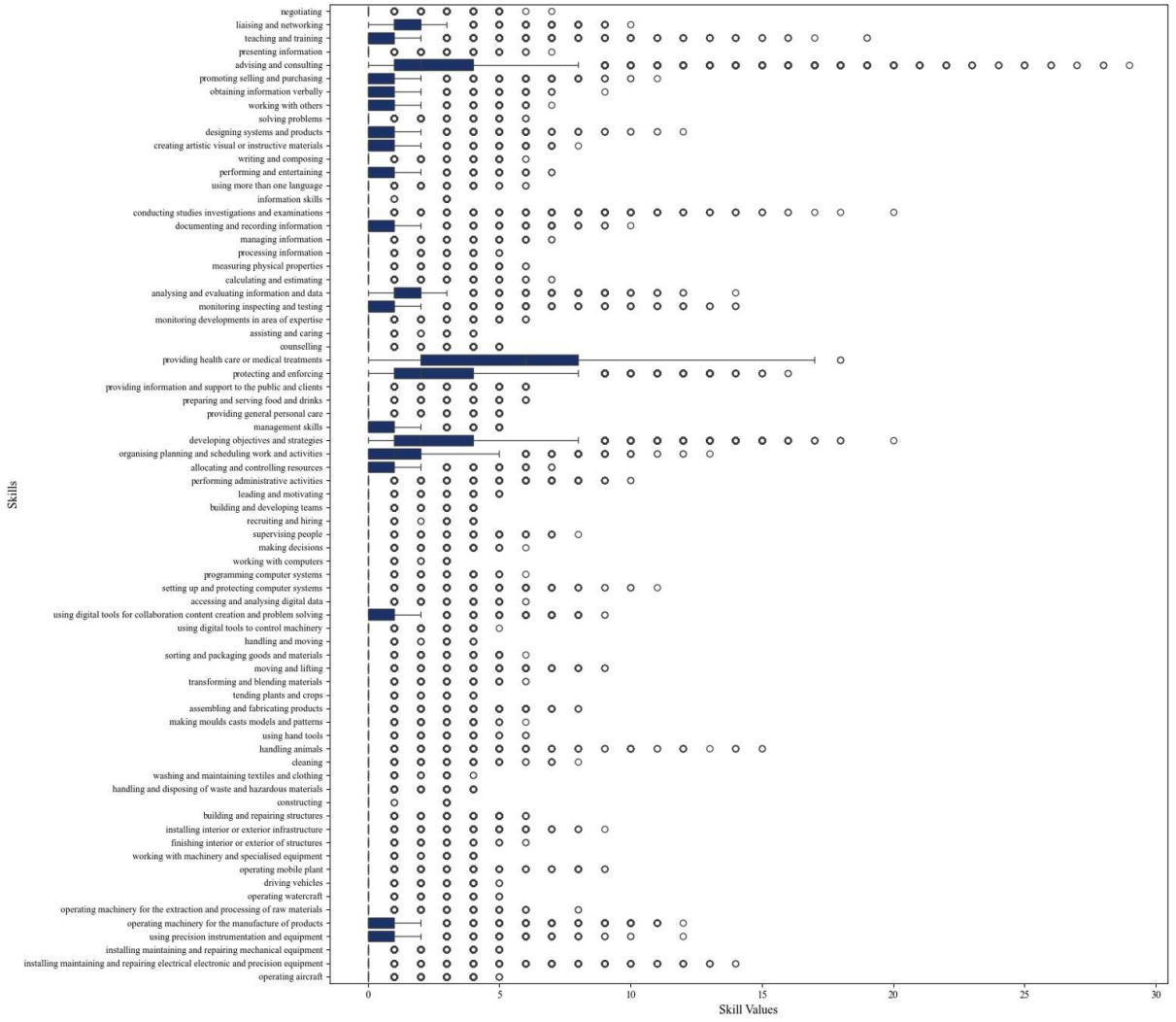


[Figure 10: Box-plot distribution analysis of 74 medium-level ESCO skills across 8th major occupational group – Plant and Machine Operators and Assemblers]

(Source: Own elaboration).

In Group 8 (Plant and Machine Operators and Assemblers), *providing health care or medical treatments* displays an IQR of 6 ($Q_1 = 1, Q_3 = 7$), whiskers extending to 16 and an outlier at 17, reflecting specialized safety and workplace injury response roles on high-risk industrial sites. *Developing objectives and strategies* has an IQR of 3 ($Q_1 =$

$1, Q_3 = 4$), whiskers to 8.5 and outliers up to 21, reflecting line supervisors and operations planners. *Advising and consulting* displays an IQR of 3 ($Q_1 = 1, Q_3 = 4$), whiskers to 8.5 and outliers from 9 to 15, reaching a maximum of 28 for senior manufacturing process specialists. *Using digital tools for collaboration* presents an IQR of 2 ($Q_1 = 0, Q_3 = 2$), whiskers to 5 and outliers at 6 and 9, reflecting automated monitoring and diagnostic system operators. *Protecting and enforcing* displays an IQR of 2, whiskers to 6 and outliers from 7 to 16.



[Figure 11: Box-plot distribution analysis of 74 medium-level ESCO skills across 9th major occupational group – Elementary Occupations]

(Source: Own elaboration).

In Group 9 (Elementary Occupations), *providing health care or medical treatments* records an IQR of 6 ($Q_1 = 2, Q_3 = 8$), whiskers extending to 17 and an outlier at 18, capturing hospital orderlies, home care aides and residential support workers.

Developing objectives and strategies spans an IQR of 3 ($Q_1 = 1, Q_3 = 4$), whiskers to 8 and outliers from 9 to 20, capturing facility support leaders and workflow coordinators. *Protecting and enforcing* displays an IQR of 3 ($Q_1 = 1, Q_3 = 4$), whiskers to 8.5 and outliers at 9, 12 and 16, reflecting facility security, gatekeeping and basic safety monitoring. *Advising and consulting* has an IQR of 3 ($Q_1 = 1, Q_3 = 4$), whiskers to 8.5 and outliers reaching 29, capturing specialized customer service and information assistance workers. *Organizing, planning and scheduling work activities* has an IQR of 2 ($Q_1 = 0, Q_3 = 2$), whiskers to 5 and outliers at 6, 7, 8 and 13.

These box-plot distributions confirm that while core technical capabilities exhibit tight, predictable ranges within formal occupations, strategic planning, advisory consulting, digital coordination and compliance activities demonstrate extensive dispersion and extreme positive outliers across all nine major occupational groups. This widespread presence of high-level outliers demonstrate that modern job postings reflect substantial intra-occupational specialization, skill hybridization and variable competence depth that transcend conventional occupational definitions.

5. Discussion

The empirical findings generated through the semantic parsing of 4.8 million online job vacancies provide critical insights into the structural transformation of labor demand and the evolving nature of work in the digital era. The granular box-plot distributions and outlier dynamics across the 74 medium-level ESCO skill groups provide compelling evidence of intra-occupational heterogeneity and role specialization. While core operational competencies exhibit tight, standardized interquartile ranges, complex cognitive and interactive functions, such as strategic planning, advisory consulting, healthcare provision and regulatory compliance, display wide dispersion and extreme positive outliers reaching values of 20 to 30 across all nine occupational categories. The presence of high-level advisory and strategic outliers within clerical, service, trade craft and elementary occupations demonstrates that formal administrative classifications (KZiS/ISCO) fail to capture the full scope of modern workplace responsibilities. Many positions officially categorized under middle- or lower-tier occupational codes in fact demand advanced project management, client consulting, digital monitoring or legal compliance oversight.

These empirical realities highlight the limitations of conventional occupational taxonomies in technological forecasting and labor market intermediation. Relying on static 6-digit occupational titles generates informational frictions and matching inefficiencies, as job titles alone do not reveal the specific capability mix required by employers. Shifting toward dynamic, skill-first frameworks, supported by semantic NLP models and dense vector architectures, enables a more accurate assessment of structural skill shortages, technological displacement risks and emerging capability requirements in real time.

From an educational and active labour market policy perspective, these findings underscore the necessity of curriculum modernization across both higher education and vocational education and training (VET) systems. Because transversal competencies, specifically digital literacy and effective communication penetrate every occupational group, educational programs must move away from rigid, siloed vocational training toward integrated curricula that embed transversal and digital skills into technical disciplines. Equipping workers with robust transversal capabilities enhances occupational mobility and adaptability, mitigating the risk of technological obsolescence as automation continues to substitute routine operational tasks.

For public employment services (PES) and digital job intermediation platforms, implementing deep contextual NLP matching models (such as Sentence-BERT coupled with FAISS indexing) represents the technological opportunity. Replacing rigid keyword search filters with dense vector semantic search engines allows recruitment platforms to match job seekers to vacancies based on transferable skill profiles rather than formal occupational histories. This capability is particularly vital for facilitating the reallocation of displaced workers from declining routine occupations into growing high-demand sectors, reducing search frictions and fostering a more resilient workforce in the face of ongoing technological change.

6. Conclusions and Limitations

6.1. Synthesis of Findings

This study implemented an advanced Natural Language Processing and vector space framework combining Sentence-BERT (all-mpnet-base-v2) contextual embeddings with

Facebook AI Similarity Search (FAISS) dense vector indexing to extract, standardize and analyze skill demand patterns across 4.8 million online job vacancies in Poland mapped to the ESCO taxonomy. Through evaluating baseline competence distributions and granular medium-level box-plot parameters across nine major occupational groups, the empirical analysis show that demand-side skill requirements vary significantly across major occupational categories in both type and frequency, while displaying strong cross-occupational interdependencies, transversal penetration and intra-occupational hybridization.

The empirical findings confirm that digital competencies, interpersonal communication and workflow management have penetrated the entire occupational hierarchy, functioning as essential capabilities in both high-skill professional domains and traditionally manual trade occupations. Physical handling, machinery operation and construction activities remain tightly coupled, forming an integrated operational core across technical fields. Simultaneously, the presence of extreme positive outliers in strategic planning, advisory consulting and compliance across non-managerial categories highlights the extensive hybridization of modern jobs, demonstrating that administrative occupational titles alone do not adequately reflect the operational complexity of workplace demand.

6.2. Methodological and Data Limitations

Several methodological and data limitations should be acknowledged when interpreting these results. First, the empirical dataset is based on commercial online job postings (SOJO), which naturally overrepresent white-collar, information technology, managerial and formal corporate positions relative to informal, seasonal or low-skilled agrarian employment that frequently relies on offline recruitment channels. Second, the analysis utilized a pooled cross-sectional structure without explicitly estimating dynamic longitudinal time-series models, limiting the direct observation of multi-year temporal trends. Third, the semantic skill extraction pipeline applied a strict 99% similarity cutoff, which maximizes precision by preventing false-positive semantic matches but may exclude nuanced or highly paraphrased competence descriptions, potentially reducing overall recall.

6.3. Future Research Directions

Future research should expand this computational framework in several promising directions. First, applying longitudinal panel architectures to continuous web-scraped vacancy streams will enable the dynamic tracking and forecasting of emerging skill trajectories, identifying early warning signals of technological deskilling and skill obsolescence in real time. Second, integrating multi-modal deep learning architectures that combine raw textual descriptions with structured employer metadata, posted wage offers, regional macroeconomic indicators and educational requirements will provide deeper causal insights into the drivers of skill demand shifts.

Declarations

Competing Interests

The author declares that she has no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data and Code Availability

The Python routines used for web scraping, Sentence-BERT embedding generation and FAISS vector indexing, alongside the processed analytical datasets, are available from the corresponding author upon reasonable request. Public data sources are accessible via the ESCO classification portal and the SOJO repository administered by the Institute of Labour and Social Studies.

Bibliography

Acemoglu, D. (1998). Why Do New Technologies Complement Skills? Directed Technical Change and Wage Inequality. *The Quarterly Journal of Economics*, *113*(4), 1055–1089. https://doi.org/10.1162/003355398555838

Acemoglu, D. (2002). Technical Change, Inequality, and the Labor Market. *Journal of Economic Literature*, *40*(1), 7–72. https://doi.org/10.1257/0022051026976

Acemoglu, D., & Autor, D. (2010). Skills, Tasks and Technologies: Implications for Employment and Earnings. *National Bureau of Economic Research, Inc, NBER Working Papers*, *4*. https://doi.org/10.1016/S0169-7218(11)02410-5

Acemoglu, D., Autor, D., Hazell, J., & Restrepo, P. (2022). Artificial intelligence and jobs: Evidence from online vacancies. *Journal of Labor Economics*, *40*(S1), S293–S340.

Acemoglu, D., & Restrepo, P. (2018a). Automation and new tasks: The implications of the task content of production for labor demand. *Journal of Economic Perspectives*, *33*(2), 3–30.

Acemoglu, D., & Restrepo, P. (2018b). Low-skill and high-skill automation. *Journal of Human Capital*, *12*(2), 204–232.

Acemoglu, D., & Restrepo, P. (2020). Robots and Jobs: Evidence from US Labor Markets. *Journal of Political Economy, The University of Chicago*, *128(6)*. https://doi.org/10.1086/705716

Allen, J. P., & Van der Velden, R. (2001). Educational mismatches versus skill mismatches: effects on wages, job satisfaction, and on-the-job search. *Oxford Economic Papers*, *53*(3), 434–452.

Anderson, R. E. (2008). Implications of the information and knowledge society for education. *International Handbook of Information Technology in Primary and Secondary Education*, 5–22.

Andrews, J., & Higson, H. (2008). Graduate Employability, 'Soft Skills' Versus 'Hard' Business Knowledge: A European Study. *Higher Education in Europe*, *33*(4), 411–422.

Arendt, L., Gałecka-Burdziak, E., Núñez, F., Pater, R., & Usabiaga, C. (2023). Skills requirements across task-content groups in Poland: What online job offers tell us. *Technological Forecasting and Social Change*, *187*, 122245.

Arendt, Ł., & Grabowski, W. (2019). Technical change and wage premium shifts among task-content groups in Poland. *Economic Research-Ekonomska Istraživanja*, *32*(1), 3392–3410.

Arrow, K. J. (1972). Economic Welfare and the Allocation of Resources for Invention. In C. K. Rowlet (Ed.), *Readings in Industrial Economics*. Palgrave. https://doi.org/10.1007/978-1-349-15486-9_13

Asplund, R., Barth, E., Lundborg, P., & Nilsen, K. M. (2011). Polarization of the Nordic Labour Markets. *Finnish Economic Papers*, *24*(2).

Attewell, P. (1990). What is Skill? *Work and Occupations*, *17*(4), 422–448. https://doi.org/10.1177/0730888490017004003

Autor, D. (2013). The “Task Approach” to Labor Markets: An Overview. *SSRN Electronic Journal*, *46*. https://doi.org/10.2139/ssrn.2211349

Autor, D. H. (2019). Work of the Past, Work of the Future. *AEA Papers and Proceedings*, *109*, 1–32.

Autor, D. H., Katz, L. F., & Kearney, M. S. (2006). The Polarization of the U.S. Labor Market. *American Economic Review*, *96(2)*, 189–194. https://doi.org/10.1257/000282806777212620

Autor, D., Levy, F., & Murnane, R. (2003). The skill content of recent technological change: An empirical exploration. *The Quarterly Journal of Economics*, *118*(4), 1279–1333.

Autor, D., & Salomons, A. (2018). *Is automation labor-displacing? Productivity growth, employment, and the labor share*. National Bureau of Economic Research.

Badet, J. (2021). AI, Automation and New Jobs. *Open Journal of Business and Management*, *09*, 2452–2463. https://doi.org/10.4236/ojbm.2021.95132

Battisti, M., Del Gatto, M., & Parmeter, C. F. (2022). Skill-biased technical change and labor market inefficiency. *Journal of Economic Dynamics and Control*, *139*, 104428.

Baumol, W. J. (1967). Macroeconomics of unbalanced growth: the anatomy of urban crisis. *The American Economic Review*, *57*(3), 415–426.

Bejaković, P., & Mrnjavac, Ž. (2020). The importance of digital literacy on the labour market. *Employee Relations: The International Journal*, *42*(4), 921–932.

Beręsewicz, M., Cherniaiev, H., Mantaj, A., & Pater, R. (2024). Text analysis of job offers for mismatch of educational characteristics to labour market demands. *Quality & Quantity*, *58*(2), 1799–1825.

Beręsewicz, M., Cherniaiev, H., & Pater, R. (2021). Estimating the number of entities with vacancies using administrative and online data. *ArXiv Preprint ArXiv:2106.03263*.

Bessen, J. (2016). How computer automation affects occupations: Technology, jobs, and skills. *Boston Univ. School of Law, Law and Economics Research Paper*, (15–49).

Bessen, J. (2019). Artificial intelligence and jobs. *The Economics of Artificial Intelligence: An Agenda*, *291*.

Bishop, C. M., & Bishop, H. (2023). *Deep learning: Foundations and concepts*. Springer Nature.

Brenčič, V., & Norris, J. B. (2010). Do employers change job offers in their online job ads to facilitate search? *Economics Letters*, *108*(1), 46–48.

Brynjolfsson, E., & McAfee, A. (2014). The second machine age: Work, progress, and prosperity in a time of brilliant technologies. In *The second machine age: Work, progress, and prosperity in a time of brilliant technologies.* W W Norton & Co.

Cappelli, P. H. (2015). Skill gaps, skill shortages, and skill mismatches: Evidence and arguments for the United States. *ILR Review*, *68*(2), 251–290.

Carlisle, S., Ivanov, S., & Dijkmans, C. (2023). The digital skills divide: evidence from the European tourism industry. *Journal of Tourism Futures*, *9*(2), 240–266.

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. *Technological Forecasting and Social Change*, *173*, 121177.

Cole, M. (2023). (Infra) structural discontinuity: capital, labour, and technological change. *Antipode*, *55*(2), 348–372.

Council of the European Union. (2018). *Council Recommendation of 22 May 2018 on key competences for lifelong learning (Text with EEA relevance) (2018/C 189/01)*.

Dabla-Norris, M. E., Pizzinelli, C., & Rappaport, J. (2019). *Job Polarization and the Declining Fortunes of the Young: Evidence from the United Kingdom*. International Monetary Fund.

Data, C. J., Laughlin, L., Song, X., Wisniewski, M., & Xu, J. (2024). *From Job Descriptions to Occupations: Using Neural Language Models to Code Job Data*.

David, P. A. (1990). The dynamo and the computer: an historical perspective on the modern productivity paradox. *The American Economic Review*, *80*(2), 355–361.

De Cuyper, N., De Witte, H., Kinnunen, U., & Nätti, J. (2010). The relationship between job insecurity and employability and well-being among Finnish temporary and permanent employees. *International Studies of Management & Organization*, *40*(1), 57–73.

Doogan, K. (2005). Long-term employment and the restructuring of the labour market in Europe. *Time & Society*, *14*(1), 65–87.

Drabowicz, T. (2014). *Determinanty nierówności w kompetencjach cyfrowych dorosłych Polaków*. Wydawnictwo Sejmowe.

ESCO. (n.d.). *Skills Pillar*. European Commission. Retrieved https://esco.ec.europa.eu/en/about-esco/escopedia/escopedia/skills-pillar

European Commission. (2021). *Digital Economy and Society Index (DESI)*. https://digital-strategy.ec.europa.eu/en/library/digital-economy-and-society-index-desi-2021

Evenson, R. E., & Westphal, L. E. (1995). Technological change and technology strategy. *Handbook of Development Economics*, *3*, 2209–2299.

Fernández-Macías, E., & Hurley, J. (2017). Routine-biased technical change and job polarization in Europe. *Socio-Economic Review*, *15*(3), 563–585.

Francis, D., Karalashvili, N., Maemir, H., & Rodriguez Meza, J. (2020). *Measuring Total Factor Productivity Using the Enterprise Surveys*.

Fulton, R., Fulton, D., & Kaplan, S. (2022). Artificial Intelligence: Framework of driving triggers to past, present and future applications and influencers of industry sector adoption. *ArXiv Preprint ArXiv:2204.01518*.

Gao, Q., Zhang, J., & Chen, G. (2023). Firm heterogeneity, biased technological change, and total factor productivity: Evidence from China. *Journal of Productivity Analysis*, *60*(2), 147–177.

Gonzalez Ehlinger, E., & Stephany, F. (2024). Skills or degree? The rise of skill-based hiring for AI and green jobs. *The Rise of Skill-Based Hiring for AI and Green Jobs (February 25, 2024)*.

Goos, M., & Manning, A. (2007). Lousy and lovely jobs: The rising polarization of work in Britain. *The Review of Economics and Statistics*, *89*(1), 118–133.

Goos, M., Manning, A., & Salomons, A. (2014). Explaining job polarization: Routine-biased technological change and offshoring. *American Economic Review*, *104*(8), 2509–2526.

Goos, M., Rademakers, E., & Roettger, R. (2021). Routine-biased technical change: Individual-level evidence from a plant closure. *Research Policy*, *50*(7), 104002.

Grace, K., Salvatier, J., Dafoe, A., Zhang, B., & Evans, O. (2018). When will AI exceed human performance? Evidence from AI experts. *Journal of Artificial Intelligence Research*, *62*, 729–754.

Green, F. (2011). *What is skill?: An inter-disciplinary synthesis*. Centre for Learning and Life Chances in Knowledge Economies and Societies London.

Greenwood, J. (1997). *The third industrial revolution: Technology, productivity, and income inequality* (Number 435). American Enterprise Institute.

Grugulis, I., & Stoyanova, D. (2011). Skill and performance. *British Journal of Industrial Relations*, *49*(3), 515–536.

Gruza, M. (2018). Classification of european skills/competences, qualifications and occupations (esco) and its link with information about occupations. *Polityka Społeczna*, *20*(1 (tematyczny)), 10–15.

Guetterman, T. C., Chang, T., DeJonckheere, M., Basu, T., Scruggs, E., & Vydiswaran, V. G. V. (2018). Augmenting qualitative text analysis with natural language processing: methodological study. *Journal of Medical Internet Research*, *20*(6), e231.

Guo, X., Cheng, Z., & Pavlou, P. A. (2024). Skill-biased technical change, again? Online gig platforms and local employment. *Information Systems Research*.

Handel, M. J. (2003). Skills mismatch in the labor market. *Annual Review of Sociology*, *29*(1), 135–165.

Hardy, W., Keister, R., & Lewandowski, P. (2016a). Do entrants take it all? The evolution of task content of jobs in Poland. *Ekonomia. Rynek, Gospodarka, Społeczeństwo*, *47*, 23–50.

Hardy, W., Keister, R., & Lewandowski, P. (2016b). Technology or upskilling? Trends in the task composition of jobs in Central and Eastern Europe. *HKUST IEMS Working Paper No. 2016-40*.

Harvey, L. (2001). Defining and measuring employability. *Quality in Higher Education*, *7*(2), 97–109.

Hershbein, B., & Kahn, L. B. (2018). Do recessions accelerate routine-biased technological change? Evidence from vacancy postings. *American Economic Review*, *108*(7), 1737–1772.

Hortigüela-Alcalá, D., Palacios Picos, A., & López Pastor, V. (2019). The impact of formative and shared or co-assessment on the acquisition of transversal competences in higher education. *Assessment & Evaluation in Higher Education*, *44*(6), 933–945.

Hötte, K. (2023). Demand-pull, technology-push, and the direction of technological change. *Research Policy*, *52*(5), 104740.

Jakubowski, M. (2015). Opening up opportunities: education reforms in Poland. *IBS Policy Paper*, *1*, 2015.

Jevons, H. S. (1931). The Second Industrial Revolution. *The Economic Journal*, *41*(161), 1–18. https://doi.org/10.2307/2224131

Katz, L. F., & Murphy, K. M. (1992). Changes in relative wages, 1963–1987: supply and demand factors. *The Quarterly Journal of Economics*, *107*(1), 35–78.

Khaouja, I., Kassou, I., & Ghogho, M. (2021). A survey on skill identification from online job ads. *IEEE Access*, *9*, 118134–118153.

Klevorick, A. K., Levin, R. C., Nelson, R. R., & Winter, S. G. (1995). On the sources and significance of interindustry differences in technological opportunities. *Research Policy*, *24*(2), 185–205.

Kucharski, L., & Kukulak-Dolata, I. (2022). Zjawisko polaryzacji na polskim rynku pracy w latach 2008–2016. *Polityka Społeczna*, *577*(04), 2–9.

Lai, N. Y. G., Jayasekara, D., Wong, K. H., Yu, L. J., Kang, H. S., Pawar, K., & Zhu, Y. (2020). Advanced automation and robotics for high volume labour-intensive manufacturing. *2020 International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA)*, 1–9.

Leontief, W. (1983). Technological advance, economic growth, and the distribution of income. *Population and Development Review*, 403–410.

Leontief, W. (1986). *Input-Output Economics. 2nd edition*. Oxford University Press.

Leslie, D., & Perini, A. M. (2024). *Future Shock: Generative AI and the international AI policy and governance crisis*.

Machin, S. (2004). Skill-biased technical change and educational outcomes. *International Handbook on the Economics of Education*, 189–211.

Mayer, C.-H., & Oosthuizen, R. M. (2021). Anxiety and excitement in the fourth industrial revolution: A systems-psychodynamic perspective. *Journal of Industrial Psychology*, *47*, 1813.

McGaughey, E. (2022). Will robots automate your job away? Full employment, basic income and economic democracy. *Industrial Law Journal*, *51*(3), 511–559.

Meta AI. (n.d.). *Facebook AI Similarity Search*. Retrieved https://ai.meta.com/tools/faiss/

Mohajan, H. (2019). The First Industrial Revolution: Creation of a New Global Human Era. *Journal of Social Sciences and Humanities*, *5(4)*, 377–387.

Mokyr, J., Vickers, C., & Ziebarth, N. L. (2015). The history of technological anxiety and the future of economic growth: Is this time different? *Journal of Economic Perspectives*, *29*(3), 31–50.

Mowery, D., & Rosenberg, N. (1979). The influence of market demand upon innovation: a critical review of some recent empirical studies. *Research Policy*, *8*(2), 102–153.

Nägele, C., & Stalder, B. E. (2017). Competence and the need for transferable skills. *Competence-Based Vocational and Professional Education: Bridging the Worlds of Work and Education*, 739–753.

National Centre for O*NET Development. (n.d.). *O*NET OnLine*. Https://Www.Onetonline.Org/. Retrieved https://www.onetonline.org/

Ozkan-Ozen, Y. D., & Kazancoglu, Y. (2022). Analysing workforce development challenges in the Industry 4.0. *International Journal of Manpower*, *43*(2), 310–333.

Palczyńska, M., & Świst, K. (2018). Personality, cognitive skills and life outcomes: evidence from the Polish follow-up study to PIAAC. *Large-Scale Assessments in Education*, *6*(1), 1–23.

Panos, G. A., Pouliakas, K., & Zangelidis, A. (2014). Multiple job holding, skill diversification, and mobility. *Industrial Relations: A Journal of Economy and Society*, *53*(2), 223–272.

PARP. (2015). Bilans Kapitału Ludzkiego w Polsce. Najważniejsze wyniki V edycji badań BKL z 2014 roku. *Pobrane z: Https://Bkl. Parp. Gov. Pl (20.05. 2017)*.

Peeters, E., Nelissen, J., De Cuyper, N., Forrier, A., Verbruggen, M., & De Witte, H. (2019). Employability capital: A conceptual framework tested through expert analysis. *Journal of Career Development*, *46*(2), 79–93.

Poletaev, M., & Robinson, C. (2008). Human capital specificity: evidence from the Dictionary of Occupational Titles and Displaced Worker Surveys, 1984–2000. *Journal of Labor Economics*, *26*(3), 387–420.

Rios, J. A., Ling, G., Pugh, R., Becker, D., & Bacall, A. (2020). Identifying critical 21st-century skills for workplace success: A content analysis of job advertisements. *Educational Researcher*, *49*(2), 80–89.

Rogers, E. M. (2001). The Digital Divide. *Convergence: The International Journal of Research into New Media Technologies*, *7*(4), 96–111. https://doi.org/10.1177/135485650100700406

Romanko, O., & O'Mahony, M. (2022). The use of online job sites for measuring skills and labour market trends: A review. *Economic Statistics Centre of Excellence (ESCoE) Technical Reports; Office for National Statistics: Newport, UK*.

Rutledge, M. S., & Guan, Q. (2015). Job polarization and labor market outcomes for older, middle-skilled workers. *Center for Retirement Research at Boston College Working Paper*, (2015–23).

Sarker, I. H. (2021). Machine learning: Algorithms, real-world applications and research directions. *SN Computer Science*, *2*(3), 160.

Saussay, A., Sato, M., Vona, F., & O'Kane, L. (2022). *Who's fit for the low-carbon transition? Emerging skills and wage gaps in job ad data*.

SBERT. (n.d.). *Pretrained Models. Sentence Transformers Documentation.* Retrieved https://www.sbert.net/docs/sentence_transformer/pretrained_models.html

Schedlbauer, J., Raptis, G., & Ludwig, B. (2021). Medical informatics labor market analysis using web crawling, web scraping, and text mining. *International Journal of Medical Informatics*, *150*, 104453.

Scheerder, A., Van Deursen, A., & Van Dijk, J. (2017). Determinants of Internet skills, uses and outcomes. A systematic review of the second-and third-level digital divide. *Telematics and Informatics*, *34*(8), 1607–1624.

Sears, J. B. (2017). When are acquired technological capabilities complements rather than substitutes? A study on value creation. *Journal of Business Research*, *78*, 33–42.

Sebastian, R. (2018). Explaining job polarisation in Spain from a task perspective. *SERIEs*, *9*(2), 215–248.

Senger, E., Zhang, M., van der Goot, R., & Plank, B. (2024). Deep learning-based computational job market analysis: A survey on skill extraction and classification from job postings. *ArXiv Preprint ArXiv:2402.05617*.

Siena, D., & Zago, R. (2024). Job Polarisation, Labour Market Fluidity and the Flattening of the Phillips Curve. *The Economic Journal*, ueae006.

Śledziewska, K., & Włoch, R. (2015). (Un) Digital Poland: The Gap in the Digital Skills of Human Capital. *ENTRENOVA-ENTerprise REsearch InNOVAtion*, *1*(1), 486–492.

Solomon, L. D. (1987). The microelectronics revolution, job displacement, and the future of work: A policy commentary. *Chi.-Kent L. Rev.*, *63*, 65.

Spitz-Oener, A. (2006). Technical change, job tasks, and rising educational demands: Looking outside the wage structure. *Journal of Labor Economics*, *24*(2), 235–270.

Stevenson, B. (2008). *The Internet and job search*. National Bureau of Economic Research.

Stoneman, P., & Battisti, G. (2010). The diffusion of new technology. In *Handbook of the Economics of Innovation* (Vol. 2, Number 1, pp. 733–760). Elsevier B.V. https://doi.org/10.1016/S0169-7218(10)02001-0

Swargiary, K. (2024). The rise of automated jobs: A comprehensive analysis of industry trends, labor market impacts, and strategies for transition. *Labor Market Impacts, and Strategies for Transition (July 01, 2024)*.

Taalbi, J. (2019). Origins and pathways of innovation in the third industrial revolution. *Industrial and Corporate Change*, *28*(5), 1125–1148.

Tsvetkova, A., D'Amico, E., Lembcke, A., Knutsson, P., & Vermeulen, W. (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*.

Van der Heijden, B. I. J. M., de Lange, A. H., Demerouti, E., & Van der Heijde, C. M. (2009). Age effects on the employability–career success relationship. *Journal of Vocational Behavior*, *74*(2), 156–164.

Van Doorn, L., & van Vliet, O. (2024). Wishing for more: technological change, the rise of involuntary part-time employment and the role of active labour market policies. *Journal of Social Policy*, *53*(3), 751–771.

Van Laar, E., Van Deursen, A. J. A. M., Van Dijk, J. A. G. M., & De Haan, J. (2020). Determinants of 21st-century skills and 21st-century digital skills for workers: A systematic literature review. *Sage Open*, *10*(1), 2158244019900176.

Van Vliet, O., Bakker, V., & Van Doorn, L. (2021). *From Social Protection to Social Investment European Responses to Globalization, Technological Change, Labour Market Flexibilization, and Migration*. https://ssrn.com/abstract=3951093

Varelas, G., Lagios, D., Ntouroukis, S., Zervas, P., Parsons, K., & Tzimas, G. (2022). Employing natural language processing techniques for online job vacancies classification. *IFIP International Conference on Artificial Intelligence Applications and Innovations*, 333–344.

Vasilescu, M. D., Serban, A. C., Dimian, G. C., Aceleanu, M. I., & Picatoste, X. (2020). Digital divide, skills and perceptions on digitalisation in the European Union—Towards a smart labour market. *PloS One*, *15*(4), e0232032.

Verma, A., Lamsal, K., & Verma, P. (2022). An investigation of skill requirements in artificial intelligence and machine learning job advertisements. *Industry and Higher Education*, *36*(1), 63–73.

Violante, G. L. (2008). Skill-biased technical change. *The New Palgrave Dictionary of Economics*, *2*, 1–6.

Ziesemer, T. (2023). Labour-augmenting technical change data for alternative elasticities of substitution: growth, slowdown, and distribution dynamics. *Economics of Innovation and New Technology*, *32*(4), 449–475.