Pillars – Pathways to Inclusive Labour Markets

Report

Quo Vadis Labour Market?

Exploiting AI to Study the Impact of Technology in Reshaping Jobs

This project receives funding from the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement No. 101004703.

Quo Vadis Labour Market?
Exploiting AI to study the impact of technology in reshaping jobs

Fabio Mercorio, Mario Mezzanzanica, Andrea Seveso

February 28, 2023

Contents

Executive Summary

The context of the action. The labour market has radically changed in the last ten years, mainly driven by technological progress, globalisation, new demographic trends, and the environment's transformation. The effects are pretty evident: (i) Increasingly, labour supply and demand are channelled through digital/online platforms; (ii) new professional jobs are emerging, and the skills required of existing ones are changing; (iii) the speed at which these phenomena are changing the dynamics of the labour market requires continuous updating and the ability to identify distinctive factors in "real-time" to support decision-making by institutions, organisations, and individuals. This WP is aimed at exploring Big Data (online job ads), and AI (machine learning and distributional semantics) in online job ads to (i) identify emerging occupations and their skill requirements and (ii) relate them to emerging technologies. The task lasted six months, specifically from M21 to M27 of the project timeline, with the deliverable identifier D5.3.

Main objectives. We employ AI-based techniques (i.e., word-embedding (Mikolov et al., 2013)) to classify OJA related to Information and Communication Technology (aka, ICT) to (i) identify OJA demands for new jobs related to technology (e.g., AI specialist, Cloud Computing, etc.) and (ii) to classify OJA using the CEN (the European Committee of Standardisation) classification of technology-related jobs.

Key activities. To meet the goals, we build an **ensemble learning model** that employs exact techniques and distributional semantics (i.e., word-embedding) to classify OJAs titles to better identify novel and ICT-related jobs according to the CEN classification system. Thanks to this, we can provide information related to (i) the distribution of OJAs related to novel jobs for Italy, Germany, France, The Netherlands and the UK; (ii) the distribution of jobs related to ICT for Italy and UK, along with their top-requested ESCO skills - as they emerge from OJAs - with the relevance for each occupation identified.

Results of the action. This WP has built an ML-based pipeline to classify OJAs from IT, UK, DE, NL, and UK according to the CEN classification, identifying novel technology-related jobs using a sample benchmark dataset of OJAs, including OJA titles. Furthermore, the WP sheds light on the relevance of skills (aka, *skill rates*) within occupations as an estimate of skill type's importance in digital, professional and soft skills.

Relevance with EU actions. The 2023 is the *Year of Skills.*[1] The European Year of Skills 2023 will help companies, in particular small and medium enterprises, to address skills shortages in the EU. It will promote a reskilling and upskilling mindset, helping people get the right skills for quality jobs. Analysing OJA goes towards this direction, enabling a real-time and deep understanding of real labour market expectations and demands regarding skills and competencies.This research's goals and results align with the rationale behind the Year of Skills, that is, to *"focus in our investment on professional education and upskilling. We need better cooperation with companies because they know best what they need. And we need to match these needs with people's aspirations. But we also have to attract the right skills to our continent, skills that help companies and strengthen Europe's growth.*[2]

[1] https://year-of-skills.europa.eu/index_en

[2] Ursula von der Leyen, President of the European Commission, 2022

1 Introduction

In recent years, significant forces and factors have dramatically changed the nature and characteristics of the labour market in both advanced and developing countries. Technical progress, globalisation and the re-organisation of the production process – with outsourcing and offshoring – have radically altered the demand for certain skills. In addition, population ageing in advanced economies intensifies the need for continued training. It is likely to affect the structural demand for certain competencies, particularly those related to the health and care of the elderly.The overall impact of these factors on the labour market is multifaceted. On the one hand, several jobs are disappearing while new jobs are emerging; some are simply a variant of existing jobs, and others are genuinely new jobs that were nonexistent until a few years ago. On the other hand, the quantity and quality of the demand for skills and qualifications associated with the new labour market have changed substantially. New skills are needed not only to perform new jobs but also because the skill requirements of existing jobs have changed considerably.

Which occupations will grow in the future and where? What skills will be demanded the most in the next years? How technology and digitalisation will affect existing and well-consolidated occupations? Those are the questions at the forefront of the policy debate among economists and policymakers. To address these questions, data-driven and real-time analysis of the labour market is needed to catch novelties - in terms of skills and new emerging jobs - as soon as they emerge from the labour market demand.

Emerging occupations have arisen or become more in demand in recent years due to technological changes, society, or the economy. These jobs may require new skills or a combination of existing skills uniquely. Emerging occupations often arise due to the introduction of new technologies, such as artificial intelligence, robotics, and automation, as well as changes in consumer behaviour and market demands. This activity is crucial to allow economists and policymakers to observe up-to-date labour market dynamics using standard taxonomies as a *lingua franca*, overcoming linguistic boundaries (see, e.g. Frey and Osborne (2017); Giabelli et al. (2020); Colombo, Mercorio, and Mezzanzanica (2019)).

As technology and digitalisation are a driver of change, we restrict our analyses to ICT-related jobs, as they are more likely to express changes in the labour market concerning non-ICT jobs. To this regard, recently, the European Committee of Standardisation (CEN) has published the European ICT Professional Role Profiles (See Fig. 1), which offer a standard set of typical roles performed

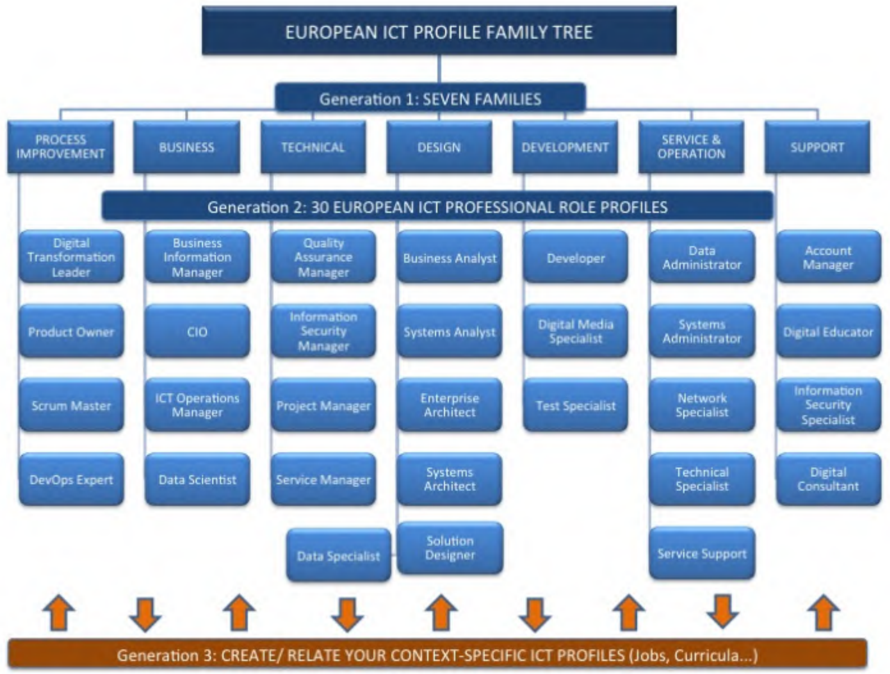

Figure 1: The European ICT profile family tree, showing the novel professions.

by IT professionals in any organisation, covering all aspects of ICT business processes. These profiles are based on the European e-Competence Framework (e-CF) for identifying competencies.

The guides and case studies provide pragmatic explanations to help users effectively apply the profiles and offer practical guidance on using them from different stakeholder perspectives and for various purposes. These may include HR planning, recruitment, supporting digital transformation processes, curriculum design and qualifications promotion, and even adapting the concept for other industries.

We have also analysed seven new emerging profiles based on novel technologies: Artificial Intelligence Specialist, Big Data Specialist, Blockchain Specialist, Cloud Computing Specialist, IoT Specialist, Mobile Specialist, and Robotics Specialist. Given the importance of these novel profiles, it became apparent that a connection between them and the more prominently used ESCO taxonomy is required. In this work, we build a bridge between the two taxonomies.

The ESCO taxonomy provides a standardised system to classify occupations and skills, including those in the ICT sector. However, as technology rapidly evolves, new job roles are constantly emerging. The 16 occupations described in the ESCO taxonomy may not be fine-grained enough to capture all the nuances of these emerging roles.

This is where AI language models come into play. Using these models makes it possible to analyse job ads and identify new potential occupations that the existing taxonomies may not cover. By doing so, we can develop a more fine-grained process of reclassifying roles relevant to the current labour market using the 30 CEN profiles and emerging technologies. By reclassifying these emerging profiles, we can better understand the skills and qualifications needed for these roles, allowing for more targeted education and training programs to meet the demands of the labour market.

2 Data and methods

We use data from the WIH-OJA system of Cedefop and Eurostat, a platform that collects Online Job Ads across Europe-selected portals. It classifies job ads and extracts skills within text according to the ESCO european standard taxonomy following the approach developed by Boselli et al. (2018). The goal of the project SKillOvate[3] is twofold: on the one side, to put OJA into official statistics (aka, *trusted smart statistics*[4]), on the other side, to analyse skills demand within occupations to support and evaluate labour market policy, as well as to monitor the web labour market dynamics in real-time to chance chances as soon as they emerge from the real labour market.

The audit dataset used in this analysis covers the period from the fourth quarter of 2020 to the third quarter of 2021. The job ads are classified according to the ISCO-08/ESCO IV digit standard. We focus on Italy, the UK, Germany, France, and the Netherlands, the countries the consortium partners represent. Only job titles are available for classification without the full description.

Important disclaimer: The number of OJAs presented in the results should not be considered representative of the actual state of the online labour market. This is because the OJA dataset used is from an audit

[3] https://www.cedefop.europa.eu/en/tools/skills-online-vacancies

[4] https://ec.europa.eu/eurostat/cros/content/trusted-smart-statistics-nutshell_en

sample of the WIH provided by Eurostat, including job titles needed for this research. Therefore, it is *recommended to focus on distributions rather than on absolute numbers* to understand the trends and patterns in the labour market.

Using the ESCO taxonomy, we can categorise skills and conduct rigorous analysis. We group skills into categories based on the ESCO v1.0.8 standard, excluding language skills and restricting professional skills to those labelled as essential or optional by ESCO. Our classification distinguishes hard and soft skills specific to a job or task, such as knowledge of particular software or instruments. Soft skills are transversal and refer to an individual's ability to interact with others and their environments, such as communication and problem-solving skills. Within hard skills, we further distinguish between hard digital skills and professional skills. Digital skills encompass a range of abilities related to using and manipulating ICT tools, while professional skills are job-specific. We then employ various measures to compare occupations and explore the composition of their skill sets.

2.1 Bridging ESCO to CEN Profiles

We propose an ensemble learning-based classification pipeline to classify OJA on the European Committee of Standardisation (CEN) profiles and emerging professions. Ensemble learning is a machine learning technique combining multiple individual models and rule-based techniques to enhance the prediction's overall performance, accuracy, and robustness (see, e.g. Giabelli et al. (2022a)). The proposed pipeline utilises four distinct steps, which leverage various techniques ranging from rule-based methods and expert assessments to word-embedding-based classification. By integrating multiple models, our proposed pipeline aims to minimise the impact of biases, errors, and noise in the individual models, thus producing reliable and accurate predictions. Below we provide a bird-eye-view of the approach, depicted in Fig. 2.

2.2 Exact matching

The initial phase of our classification pipeline involves precisely matching the CEN profile label against the job advertisement titles. In other words, it consists of comparing the phrases with each other (title against the CEN profile) suitably preprocessed and marking the comparison as matched when the two strings are perfectly equal. A complete match ensures accurate classification. Additionally,

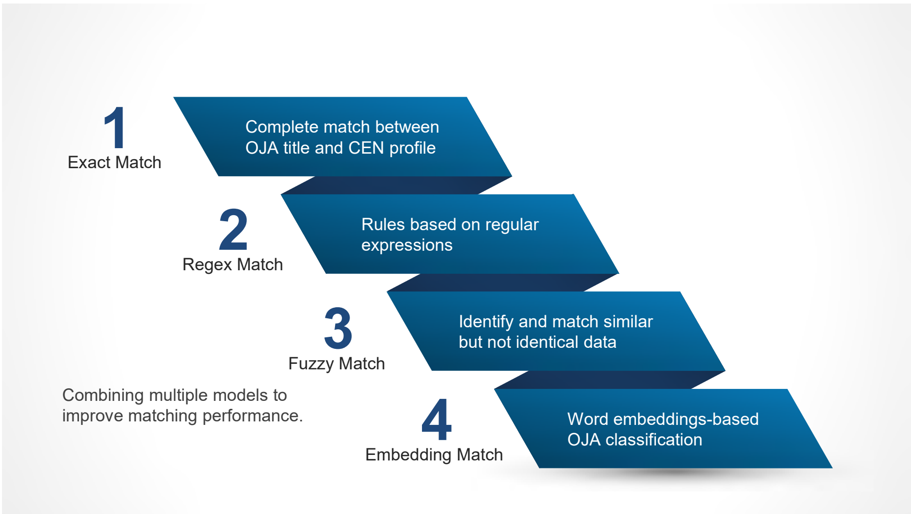

Figure 2: Classification pipeline using ensemble learning.

we incorporate selected search terms recommended by labour market experts and closely related terms identified through word-embedding similarities. This multi-pronged approach improves the accuracy and comprehensiveness of our classification model.

This phase requires carefully avoiding every possible mismatch: edges of the string, spaces, the format of the string itself, and so on. The phase is very fast and, therefore, scalable. It is dependent on the order of comparison.

2.3 Regular Expressions

The subsequent stage builds upon the initial step and employs regular expressions to identify matches between terms and job titles while incorporating additional rules. Regular expressions, or Regex, is a pattern-matching language that facilitates searching, manipulating, and validating text based on predefined patterns.

Surprisingly, the idea behind regexes did not originate as a computer tool. Instead, regular expressions were created in the field of neuroscience. In 1943, Warren S. McCulloch (Neuroscientist) and Walter Pitts (Logician) began developing models that describe how the human nervous system works. Their research focused on understanding how the brain might produce complex patterns using simple cells tied together. In 1956, mathematician Stephen Kleen described McCulloch-Pitts neural models with an algebraic notation that described "regular expressions". Influenced by Kleen's notion, in 1968, mathematician and Unix pioneer Ken Thompson implemented the idea of regular expressions within the text editor, "ed". Its purpose was for end users to be able to make advanced pattern matching in text files. Larry Wall's Perl programming language from the late 1980s has helped regular expressions go mainstream. Perl was originally designed as a flexible text-processing language but has grown to become a full-fledged programming language that remains a natural choice for any text processing. Supported by all modern programming languages, word processing programs and advanced text editors, regex is now used in over a third of Python and JavaScript projects. The search pattern they interpret allows them also to be used in various applications such as OCR recognition.

Regular expressions can also be thought of as a programming language which expresses relevant applications, relating, for example, to the field of text searches or as components of a compiler. Regular expressions are closely related to nondeterministic finite automata (NFA) and can represent a convenient alternative to the notation of NFA to describe software components. Regular expressions define the same languages described by the various forms of automaton: regular languages. Regular expressions, however, offer something more than automata: a declarative way of expressing the strings to accept. Therefore regular expressions serve as the input language for many systems that deal with strings. This feature enables accurate and efficient text processing by identifying and extracting specific parts of a string or text. For instance, the search term may be a subset of the job title, or the words may appear in a different order.

2.4 Fuzzy Matching

We then employ fuzzy matching, also known as approximate string matching, to identify and match data that may not be the same but similar. It involves measuring the similarity between two strings using various algorithms and matching techniques, such as Levenshtein distance, to find the best match.

The Levenshtein distance, or edit distance, measures the difference between two strings. Introduced by Russian scientist Vladimir Levenshtein in 1965, it determines how similar two strings are. In our situation, the Levenshtein distance is a text similarity metric that measures the distance between 2 words or two sentences. It has several applications, including text autocompletion and auto-correction. For either of these use cases, the word entered by a user is compared to words in a dictionary to find the closest match, at which point a suggestion (s) is made. In the case of two words, the Levenshtein distance is usually calculated by preparing a matrix of size (M+1)x(N+1), where M and N are the lengths of the two words, and looping through said matrix using 2 for loops, performing some calculations within each iteration. Calculating the distance between a word and a dictionary of thousands of words is time-consuming. Fuzzy matching is particularly useful where the data may contain errors, typos, variations in the spelling, singular/plural words and other minor differences that might have caused the previous steps to miss a potential match.

We used Rapidfuzz, which is a complex and sophisticated string comparison Python library developed by Max Bachmann[5] that contains different comparison modes within it. Different techniques of the fuzzy match are used, e.g. token sort ratio and token set ratio.

2.5 AI-based word embeddings classifier

Lastly, we resort to a machine-learning classifier to find missing matches. It is based on FastText (Bojanowski et al., 2017), a technique for representation learning which builds word embeddings considering sub-word information by representing each word as the sum of its character n-gram vector. It uses an unsupervised learning technique to generate word embeddings by representing words as vectors in a high-dimensional space, where words with similar meanings are closer to each other in the vector space. In the CEN classification task, FastText uses these word embeddings to create feature vectors for entire OJAs by averaging the embeddings of individual words in the text. These feature vectors are then used to train a classifier to predict the CEN profile of the OJA. This classification method using word embeddings is highly effective in various natural languages processing tasks, such as sentiment analysis, topic classification, and spam detection. It can also handle out-of-vocabulary words (not seen during training) by breaking them down into sub-word n-grams, allowing for better coverage and generalisation.

[5] https://github.com/maxbachmann/RapidFuzz

2.6 Metrics for Analysing Skills

Our metric is based on the concept of Revealed Comparative Advantages (RCA), which was initially developed in international trade theory. RCA uses post-trade measurements to deduce a region's comparative advantage in producing traded goods in a multilateral context (Balassa, 1965). In recent years, the economic geography field has widely employed RCA to characterise a region's specialisation in productive activities (see van Dam et al. (2022) for a comprehensive review of the literature and the latest developments). We apply a similar approach, using measures of skill frequency in each occupation as derived from online job advertisements and computing the extent to which each occupation is specialised in each skill. As a result, we can identify emerging occupations and skills in high demand.

Our approach to measuring occupations' skills specialisation builds on the method proposed by Alabdulkareem et al. (2018), which utilises the O*NET skills classification and relevance as key factors in calculating revealed comparative advantages. However, to apply this method to European online job ads data, we need an alternative measure of skill relevance as it is not directly available in the ESCO skill classification system. To overcome this limitation, we adopt the approach proposed by Giabelli et al. (2022b), which measures skill relevance based on the frequency of skills in online job ads. Specifically, we compute the skill relevance as the ratio of the frequency of a particular skill in job ads for a given occupation relative to the frequency of the same skill in job ads for all other occupations. We define a set of occupations $\bar{O}_t = o_k, k = 1, \ldots, K_t$ and a set of skills for each occupation $\bar{S}_{kt} = s_j, j = 1, \ldots, J_t$ for each year of observation t. Finally, we calculate the skill frequency as follows:

$$sf(o_k, s_j) = \frac{\sum_{i=1}^{n} I(o_i = o_k) \cdot I(s_i = s_j)}{\sum_{i=1}^{n} I(o_i = o_k)} \tag{1}$$

where I denotes the indicator function and $\sum_{i=1}^{n} I(o_i = o_k) \cdot I(s_i = s_j)$ the count of the occurrences of the skill s_j for occupation o_k. The term $\sum_{i=1}^{n} I(o_i = o_k)$ represents the total number of observations of occupation o_k.

Iterating over skills and occupations, we obtain a matrix $\underset{m \times p}{M}$ of the skill frequency for each pair of occupations $i \in \bar{O}$ and skills $s_j \in \bar{S}$. The revealed comparative advantage, rca, for occupation i and s_j is defined as:

$$rca_{ij} = \frac{sf_{ij} / \sum_{j=1}^{J_t} sf_{ij}}{\sum_{i=1}^{I} sf_{ij} / \sum_{i=1}^{I} \sum_{j=1}^{J_t} sf_{ij}} \tag{2}$$

with a range between $[0, +\infty)$.

The distribution of RCA is not evenly distributed around its neutral point of 0. To account for this, we utilise a transformation method proposed by Laursen (2015), which involves calculating a non-linear transformation of the RCA. This transformation is referred to as Revealed Symmetric Comparative Advantages:

$$srca_{ij} = \frac{rca_{ij} - 1}{rca_{ij} + 1} \tag{3}$$

The $srca_{ij}$ is uniformly distributed over a continuous interval of $[-1, 1]$ and has a neutral point of 0. The $(0, 1]$ values indicate that the occupation i is specialised in skill j, while the values in the $[-1, 0)$ range indicate that the skill is requested.

3 Findings

3.1 Exploring the Results: Descriptive Analytics

In this section, we present the descriptive statistics of the CEN classification results for the audit sample collected by the WIH-OJA system of Cedefop and Eurostat.[6] The sample includes online job advertisements from five European countries: Italy, the UK, Germany, France, and the Netherlands. The job advertisements have been classified according to the above pipeline on the CEN and emerging profiles.

ICT in the Audit Sample. We first report the distribution of the number of job advertisements of the audit sample in each target country. To perform the classification, we select only 16 ESCO professions that are related to ICT, with the following ISCO-08/ESCO IV digit codes: Information and communications technology service managers (1330), Systems analysts (2511), Software developers (2512), Web and multimedia developers (2513), Applications programmers (2514), Software and applications developers and analysts not elsewhere classified (2519), Database designers and administrators (2521), Systems administrators (2522), Computer network professionals (2523), Database and network professionals not elsewhere classified (2529), Information and communications technology operations technicians (3511), Information and communications technology user support technicians (3512), Computer network and systems technicians (3513), Web technicians (3514), Broadcasting and audiovisual technicians (3521) and Telecommunications engineering technicians (3522). Table 1 shows the frequency and percentage of job advertisements in each country and the number and percentage after filtering on ICT codes. The UK has the

[6] The dataset has been exposed as an audit dataset `wih_oja_sample_v1_2021_r20220215`

Table 1: Number of OJAs in the audit sample in each country, and fraction of ICT ads.

Country	% ICT OJA	# Audit OJA
DE	4.4%	839,237
FR	4.8%	1,046,587
IT	5.0%	520,618
NL	4.1%	603,863
UK	5.3%	971,958

highest number of OJAs, followed by France. **The incidence of ICT codes over the audit samples is around 5% in each country.**

3.2 Emerging Jobs Classification

The emerging job profiles considered for this analysis were Artificial Intelligence Specialist, Big Data Specialist, Blockchain Specialist, Cloud Computing Specialist, IoT Specialist, Mobile Specialist, and Robotics Specialist. The complete reclassification results, grouped by individual profile, are shown in Tab. 2 and Fig. 3.

When comparing the results using the percentage of emerging job profiles, the Netherlands have the highest percentage at 8.9%, followed by Germany at 8.6%, and France at 6.1%. The UK and Italy have the lowest percentages, at 4.4% and 3.2%, respectively. These percentages represent the proportion of emerging job profiles relative to the total number of ICT job advertisements reclassified in each country.

3.3 CEN Profiles Classification

The classified job advertisements provide valuable insights into the distribution of ICT jobs across different CEN profiles and countries. In this section, we present the distribution of job ads successfully classified through matching. CEN profiles consist of 96.7% of the reclassified ads in Italy, and 95.6% in the UK, with the remaining 3.3% and 4.4% being emerging profiles, as detailed in the previous section. Table 3 shows the results of the matching process between job ads and the CEN classification in two different countries, Italy (IT) and the United Kingdom (UK).

The results in table 3 show that the most commonly advertised CEN profile

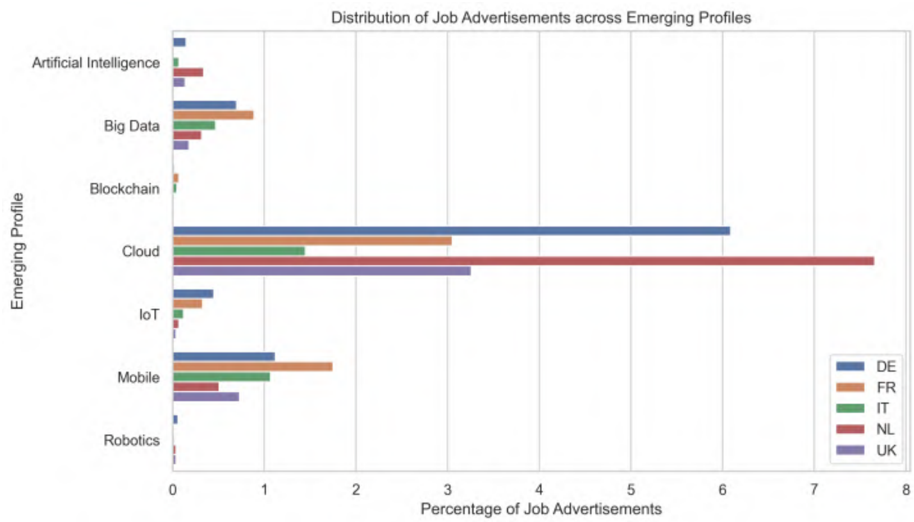

Figure 3: Emerging job profile distribution in each country, showing the ratio of matching job ads from Q4 2020 to Q3 2021.

in both countries is "Developer", with 26.54% ads classified as such in Italy and 15.02% in the UK. The second most common profile is "Digital Consultant," with 13.41% ads and 12.93% in the UK. There are also major differences in the distribution of CEN profiles between the two countries. For example, there are more job advertisements for "CIO" and "Business Information Managers" in Italy. At the same time, there are more job advertisements for "Data Scientists" and "Project Managers" in the UK. Another notable difference is in the "Technical Specialist" profile, which is required in 4.09% of the ads in the UK, more than five times as many as in Italy, where 0.81% of the ads search for this profile. While the UK has a higher proportion of "Digital Transformation Leaders" profiles than Italy, Italian IT job ads have a higher percentage of "Service Support" and "Systems Analyst" profiles than the UK.

This analysis provides an overview of the most common CEN profiles in the job market and their distribution across countries. Furthermore, it allows us to identify potential trends and areas of growth in the job market, as well as areas where there may be a shortage of skilled workers.

3.4 Skill Rate Analysis

Skill rate analysis examines the prevalence and importance of skills required for specific jobs or occupations. It involves calculating the percentage of job

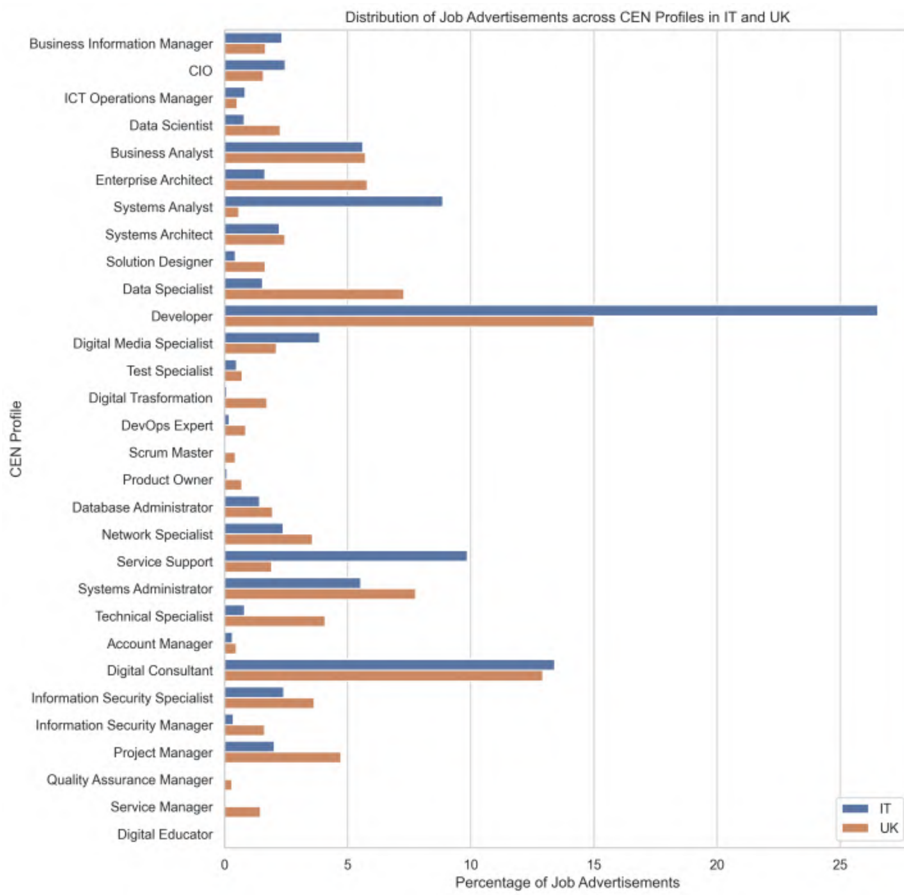

Figure 4: CEN job categories distribution in Italy and the UK, showing the number of job ads from Q4 2020 to Q3 2021.

Table 2: Distribution of job advertisements classified under each emerging job market profile, based on the audit dataset from Q4 2020 to Q3 2021.

Emerging Profile	Description of the emerging profile	DE	FR	IT	NL	UK
Artificial Intelligence Specialist	Guides, the process of application and identification of Artificial Intelligence algorithms, supervises the training activities of the algorithms, identifies the qualitative metrics for the evaluation of the achieved accuracy	0.15%	0.89%	0.07%	0.34%	0.14%
Big Data Specialist	Guides the process of identification and application of architectures for the management of structured and unstructured data; identifies and defines the technical requirements for an effective data processing pipeline; identifies strategies to extract value from data for the organisation's business	0.7%	<0.01%	0.47%	0.32%	0.18%
Blockchain Specialist	Leads the process of defining a distributed architecture based on blockchain develops and defines the protocol for proof-of-work on blockchain	0.02%	0.07%	0.05%	<0.01%	0.01%
Cloud Computing Specialist	Guide and support the company during the process of migrating data from on-premises to the cloud; guarantees the identification of the most appropriate cloud technologies concerning business needs	6.09%	3.05%	1.45%	7.66%	3.26%
IoT Specialist	Organise and model data received from various sources, devices and sensors; supports the organisation in defining IoT services; He is responsible for the creation and identification of the IoT architecture of the company	0.45%	0.33%	0.12%	0.07%	0.04%
Mobile Specialist	Implements and maintains corporate services delivered on a mobile platform	1.12%	1.75%	1.07%	0.51%	0.73%
Robotics Specialist	Defines semi-automatic robotic systems to improve company production efficiency; documents and integrates the robotic systems developed with the organisation's IT systems	0.06%	0.01%	0.02%	0.04%	0.04%

postings that mention a particular skill or set of skills and comparing this to the overall number of job postings. In particular, we use Relative Skill Contribution

Table 3: Distribution of job advertisements classified under each CEN profile in the IT and UK job markets, based on the audit dataset from Q4 2020 to Q3 2021.

CEN Profile	Description	IT	UK
Account Manager	Figure of reference for sales and customer satisfaction	0.32%	0.48%
Business Analyst	Analyse the business domain and optimise business performance through the application of technology	5.62%	5.73%
Business Information Manager	Proposes, designs and manages the functional development of the Information System (IS) focusing on user needs	2.33%	1.66%
CIO	Develops and maintains Information Systems to generate value for the business and meet the organisation's needs.	2.47%	1.58%
Data Scientist	Drive the process for applying data analytics algorithms. Defines and implements the analytics process, representing the data in a visual form	0.8%	2.26%
Data Specialist	Ensures implementation of organisations data management policy	1.55%	7.29%
Database Administrator	Design, implement, monitor and maintain databases, both structured (e.g., relational) and unstructured (text).	1.42%	1.95%
DevOps Expert	Implement processes and tools for development (Dev) and deployment (Ops) throughout the entire software development cycle	0.19%	0.86%
Developer	Designs and develops (coding) software components to respond to specific company solutions	26.54%	15.02%
Digital Consultant	Supports the process for enhancing and understanding the value that digital technologies provide to the corporate business	13.41%	12.93%
Digital Educator	Educates and trains professionals to achieve optimal digital competence to support the organisation's business	<0.01%	0.04%
Digital Media Specialist	Integrates digital technology components for internal and external communication	3.87%	2.11%
Digital Transformation Leader	Provides direction for executing the organisation's digital transformation strategy	0.09%	1.73%
Enterprise Architect	Designs and maintains the holistic architecture of business processes and information systems	1.64%	5.8%
ICT Operations Manager	Manages overall ICT operations, people and assets	0.84%	0.51%
Information Security Manager	Leads and manages the organisation's information security policy	0.36%	1.63%
Information Security Specialist	Ensures the implementation of the organisation's information security policies through the appropriate use of ICT resources	2.41%	3.64%
Network Specialist	Ensures alignment of the network, including telecommunications and/or IT infrastructure to meet the communication needs of the organisation	2.38%	3.57%
Product Owner	Report stakeholder community needs and customer feedback to the development team	0.11%	0.7%
Project Manager	Manage projects for optimal performance and results	2.03%	4.73%
Quality Assurance Manager	Ensures that processes and organizations applying information systems comply with quality policies.	0.04%	0.3%
Scrum Master	Leads and mentors agile development teams	0.01%	0.44%
Service Manager	Plan, implement and manage user solutions	0.06%	1.46%
Service Support	Provides both remote and onsite support to resolve technical issues for internal or external customers	9.86%	1.92%
Solution Designer	Provides translation of business requirements into end-to-end IT solutions.	0.44%	1.65%
Systems Administrator	Administer ICT system components to meet service needs	5.54%	7.76%
Systems Analyst	Analyse organisational requirements and specify system software requirements for new IT solutions	8.87%	0.59%
Systems Architect	Plan, design and integrate ICT system components, including hardware, software and services	2.23%	2.46%
Technical Specialist	Maintains and repairs hardware, software and service applications	0.81%	4.09%
Test Specialist	Design and test applications	0.49%	0.72%

Analysis (RSCA) to calculate the skill rate as in Sec. 2.6. This analysis can provide insights into the demand for certain skills in the job market and help inform decisions related to education and training programs. We focus on Italy and the UK and explore the skills required for each CEN profile.

Firstly, we compute the percentage of skills for each profile, divided into transversal, professional, and digital categories. Transversal skills, transferable or soft skills, are general skills and abilities that can be applied across various jobs and industries. These include communication, problem-solving, teamwork, time management, and adaptability. Professional skills, also known as job-specific or hard skills, are specific knowledge and abilities required for a particular job or industry. These skills can be acquired through education, training, and work experience. Examples of professional skills include accounting and legal knowledge. Digital skills, or digital literacy or IT skills, are the ability to use digital technologies and tools effectively. These include basic computer skills, using software applications, digital communication, and online collaboration. Figure 6 shows the composition of skill rate for each profile and country.

Secondly, we focus on digital skills and divide them into four categories:

- Applied & Management Informatic Skills are skills related to applying computer technology and information management to solve problems and support decision-making processes.

- Basic Information Skills are the essential skills needed to use and manipulate digital information effectively, such as browsing the internet, using search engines, managing email, and using office productivity software.

- ICT Technical Skills are skills related to developing, maintaining, and administering computer and network systems, such as programming, system administration, and database management.

- Information Brokerage Skills are skilled in acquiring, organising, and disseminating information to support decision-making processes, such as data analysis, visualisation, and knowledge management.

As a rule of thumb, Applied & Management Informatic Skills and ICT Technical Skills are more specialised and technical, while Basic Information Skills and Information Brokerage Skills are more general and widely applicable. Figures below show the distribution of these skills over each profile.

3.5 Most Requested Skills in CEN profiles

The section analyses the top skills required for each CEN profile in the UK and Italy. The analysis focuses on the most used digital, professional, and transversal

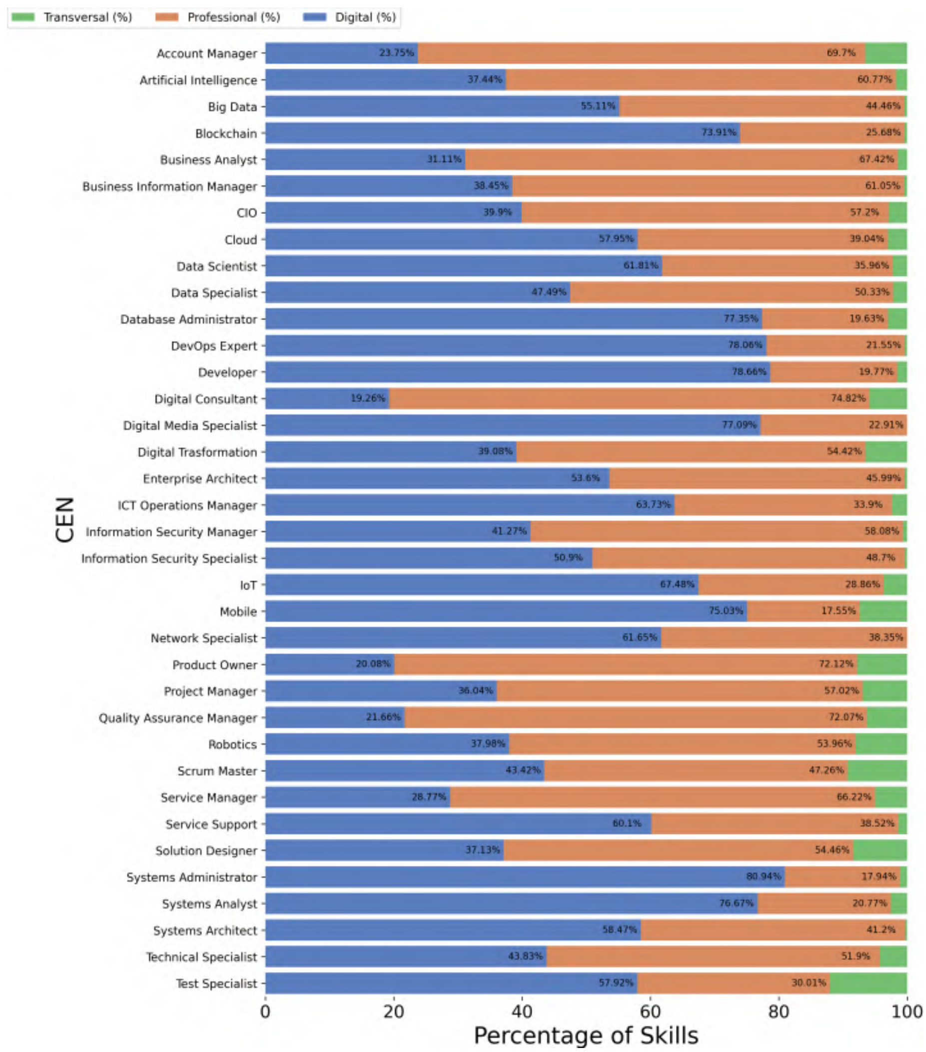

Figure 5: Comparing Digital, Professional and Transversal Skill Rates of IT and UK CEN Profiles: Skill rate in Italy

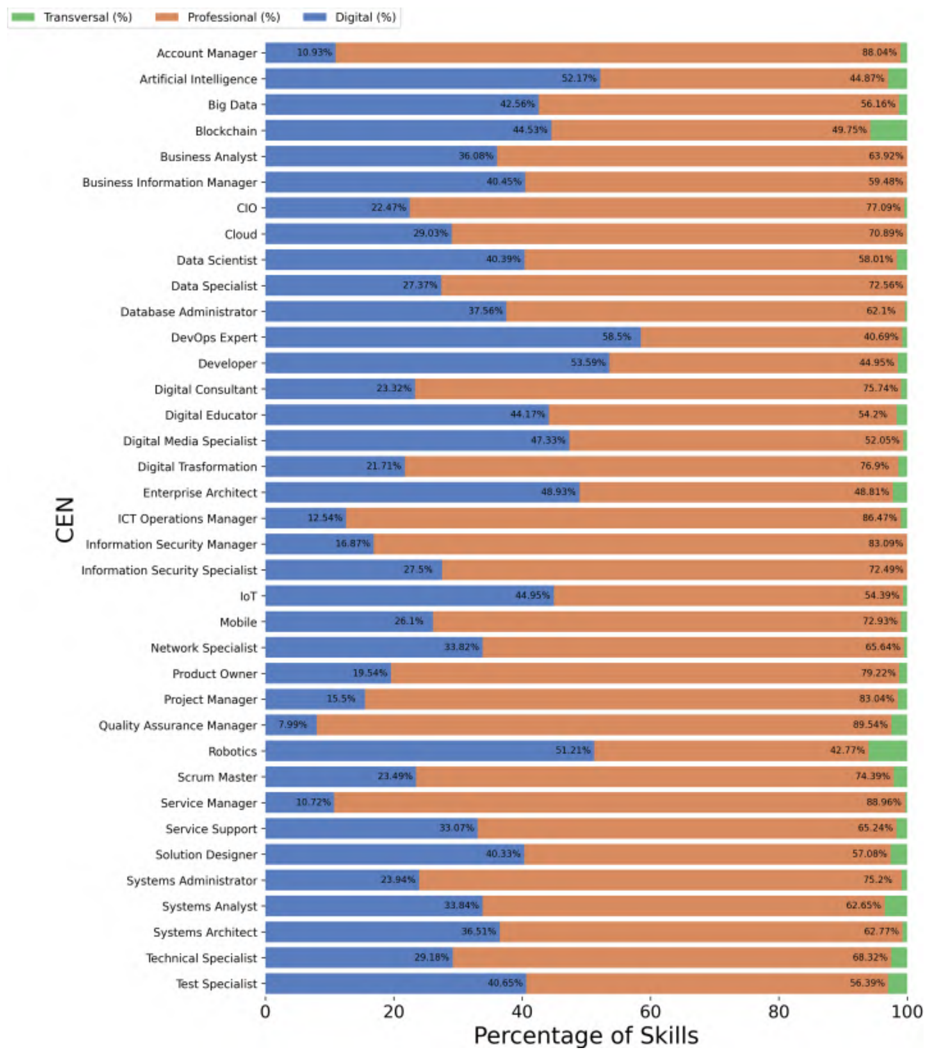

Figure 6: Comparing Digital, Professional and Transversal Skill Rates of IT and UK CEN Profiles: Skill rate in the UK

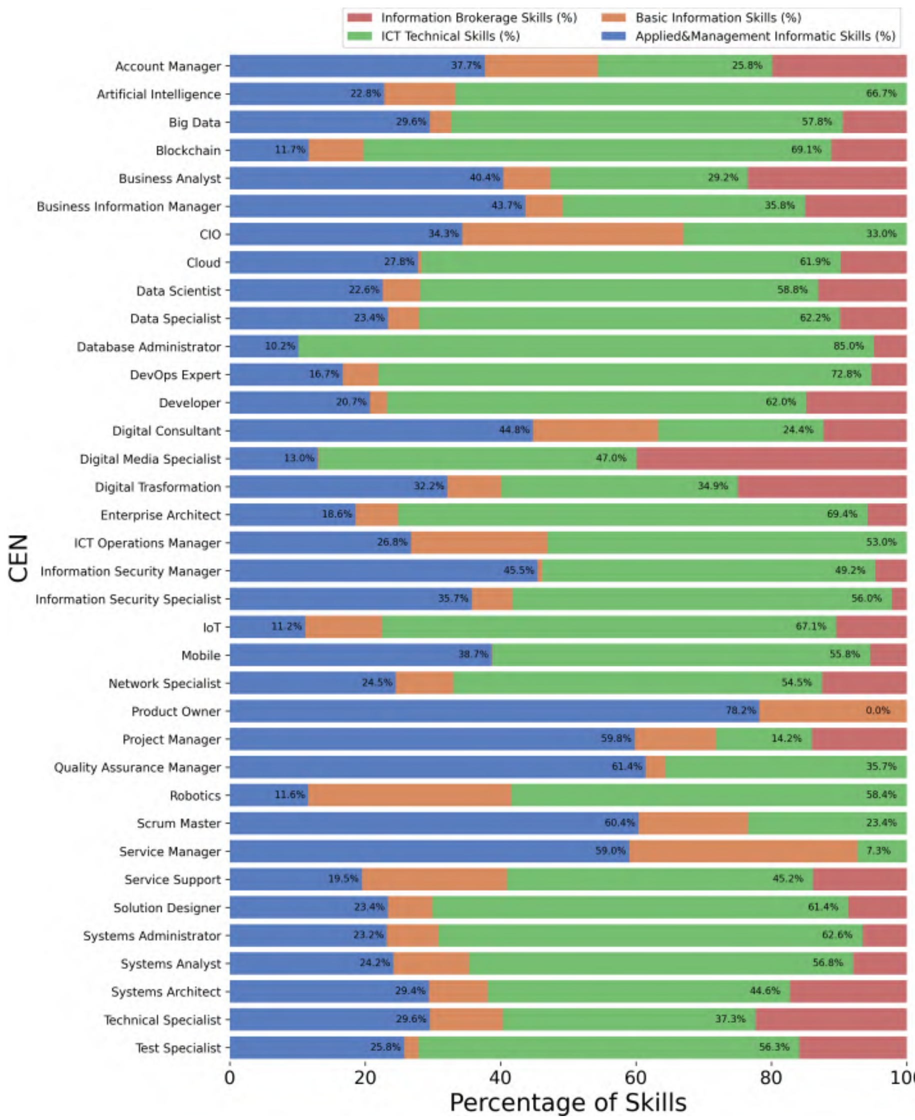

Figure 7: Comparing ICT Skill Rates of IT and UK CEN Profiles: IT skill rate

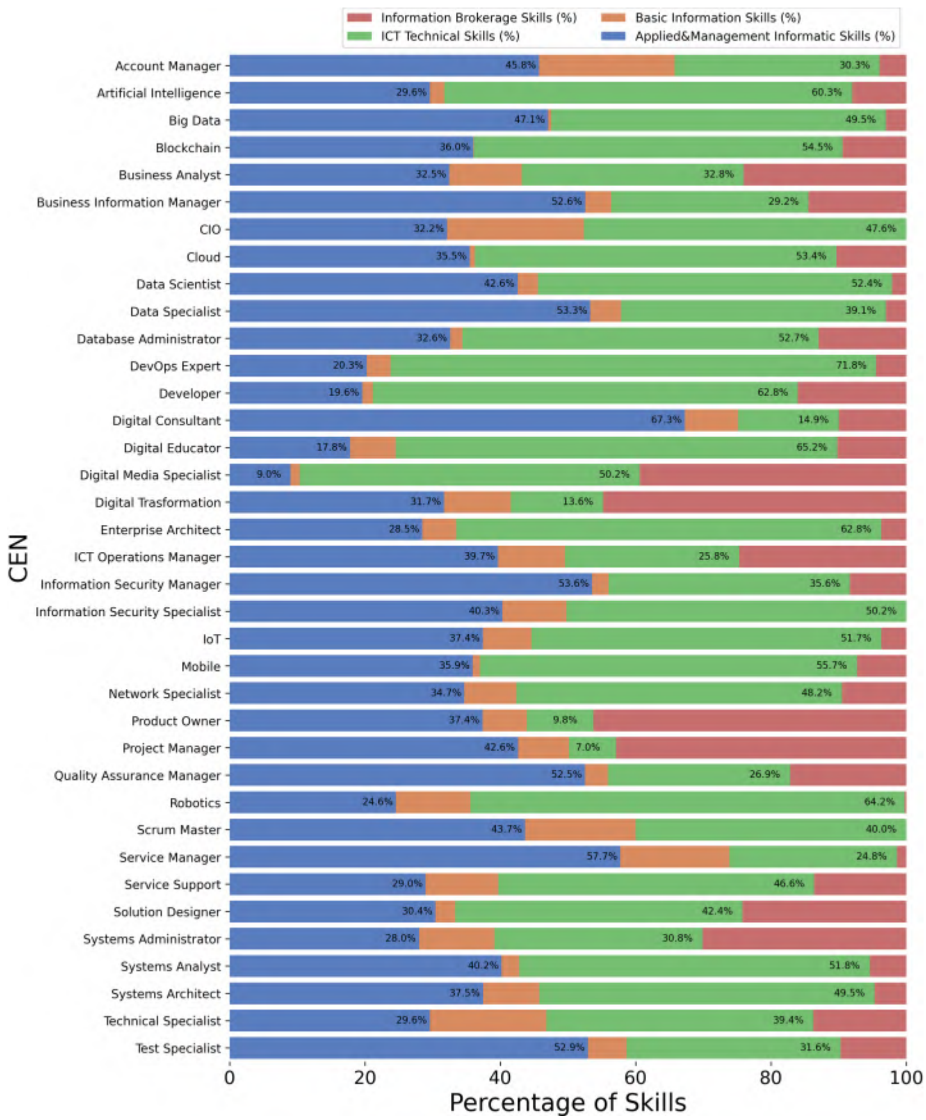

Figure 8: Comparing ICT Skill Rates of IT and UK CEN Profiles: UK skill rate

skills for each profile, based on a measure called Relative Skill Contribution Analysis (RSCA), as described in Sec. 2.6. The results provide insights into the skills in high demand in each country and profile to comprehensively understand the skills required and how they differ between the two countries.

3.6 Skill Analysis for Emerging Jobs

In the paper, we provide examples of top skills for each profile to give readers an idea of the type of information included in the analysis. However, due to the study's many skills and profiles, the full table of top skills for each profile in the UK and Italy were too large to include directly. Therefore, we have provided the full table in the additional material for interested readers to consult. This table allows for a more detailed and comprehensive understanding of the high-demand skills for each profile and how they differ between the two countries.

Table 4 shows an example of the full skill analysis, with the top digital, professional, and transversal skills for two emerging profiles: Artificial Intelligence Specialist and Cloud Specialist in the UK. The RSCA measure is used to identify the most significant skills for these profiles, providing a comprehensive understanding of the skill requirements and how they differ between the two profiles.

For the emerging profile of Artificial Intelligence Specialist, the most required digital skill is *Natural Language Processing*, a subfield of Artificial Intelligence (AI) that focuses on enabling machines to understand, interpret and generate human language. It involves the development of algorithms and computational models that can process and analyse large amounts of natural language data, such as written text and spoken language. Natural language processing is used in various applications, including machine translation, sentiment analysis, speech recognition, and chatbots. In addition, the second most required digital skill is *Sass*, a preprocessor scripting language that is interpreted or compiled into Cascading Style Sheets (CSS). On the other hand, for the emerging profile of Cloud Specialist, the most required digital skill is *Cloud Technologies*, which includes different platforms and technologies related to cloud computing. The second skill for this profile is *Typescript*, a programming language developed and maintained by Microsoft that is a strict syntactical superset of JavaScript.

For the professional skills, the results show that the most requested skills for the AI specialist profile are *mathematics* and *supply chain management*.

This indicates the importance of having a strong foundation in mathematical concepts such as linear algebra and calculus for developing and deploying AI solutions. Additionally, supply chain management is becoming increasingly important in the age of big data as companies strive to optimise their supply chain processes through advanced analytics and AI. On the other hand, for the cloud specialist profile, the most requested professional skills are *service-oriented modelling* and *performing business analysis*. This suggests that cloud specialists must have a strong understanding of service-oriented architectures and be able to identify business requirements and translate them into technical solutions. Service-oriented architectures use distributed systems to provide services that can be reused and combined to create new applications. In contrast, business analysis involves identifying and evaluating business needs and recommending solutions to improve processes and systems.

Both profiles require the transversal skills *work efficiently* and *work independently*, as these are essential skills for any job, particularly for high-skilled emerging profiles. Working efficiently is crucial for meeting deadlines and achieving goals while working independently shows a level of autonomy and self-reliance that is highly valued in the workplace. It is worth noting that these skills are not necessarily innate but can be developed through training and practice. Therefore, job seekers should focus on developing these transversal skills and the specific technical skills required for their chosen profession to increase their employability and career success.

4 Lessons Learned

The study analysed emerging occupations in five different countries using data from the WIH audit dataset, using online job ads. The incidence of ICT codes was found to be around 5% in every country.

We then studied the distribution of emerging occupations and their incidence across countries. The Netherlands had the highest percentage of emerging occupations at 8.9%, while Italy had the lowest at 3.2%. The most commonly advertised emerging occupation was "Cloud Computing Specialist", and the most widespread CEN profile was "Developer".

Skill rate analysis showed each profile's diversified needs of transversal, professional, and digital skills. In the field of ICT, it is common knowledge that digital skills are highly sought-after. However, in this study, we have shown that these skills alone are not sufficient to thrive in the industry. Other complementary skills, both transversal and professional, are also highly valued by

Table 4: Top digital, professional, and transversal skills for two emerging profiles in the UK: Artificial Intelligence Specialist and Cloud Computing Specialist. Score computed using RSCA.

Country	Profile	Skill Type	Skill	Score
UK	Artificial Intelligence Specialist	Digital	natural language processing	●●●●●
			Sass	●●●●●
			implement front-end website design	●●●●
			JavaScript	●●●●
			SQL Server Integration Services	●●●●
		Professional	mathematics	●●●●
			supply chain management	●●●
			information security strategy	●●●
			engineering principles	●●●
			identify customer's needs	●●●
		Transversal	identify with the company's goals	●●●
			work efficiently	●●●
			work independently	●
	Cloud Computing Specialist	Digital	TypeScript	●●●●
			cloud technologies	●●●●
			web application security threats	●●●●
			use content management system software	●●●
			manage ICT virtualisation machines	●●●
		Professional	service-oriented modelling	●●●●●
			perform business analysis	●●●●
			Lean project management	●●●●
			Ansible	●●●●
			advertising techniques	●●●●
		Transversal	work efficiently	●
			work independently	●

employers. It is, therefore essential for job seekers to not only focus on developing their digital skills but also to cultivate a well-rounded set of professional and transversal skills to increase their competitiveness in the job market. Similarly, employers should prioritise hiring individuals with diverse skills and experiences to ensure their teams are equipped to handle the complex and evolving nature of the industry. The most relevant skills required for each CEN profile in the UK and Italy were also provided.

The lessons learned from the study also included the need for ongoing analysis of emerging occupations to support labour market policies and facilitate a smooth transition to new forms of work. To stay up to date with emerging trends, employers should regularly monitor the evolving skill demands of their respective industries. They can do this by using tools such as the present report, staying current on industry news and developments, and seeking feedback from current and potential employees. Employers should also be open to upskilling and reskilling their current workforce to ensure they have the necessary skills to meet emerging demands.

Job seekers should also keep themselves informed about the emerging trends in their field of interest. They can do this by regularly browsing job postings and identifying the skills and qualifications that are in high demand. Job seekers should also be open to upskilling and reskilling themselves through attending workshops, courses, or other educational opportunities, to make themselves more marketable to potential employers.

5 Conclusions

In conclusion, the analysis provides valuable insights into the distribution of job advertisements classified as each CEN profile, emerging jobs and the associated skill requirements. These findings can be used to inform policy-making decisions related to education and training programs, recruitment strategies, and identifying skills gaps and shortages. By understanding the skill demands of each profile in different countries, policymakers and educators can better align training and education programs to meet the job market's needs. Employers can use this information to design more effective recruitment strategies and target their efforts towards specific skills and profiles in high demand. This analysis provides a comprehensive understanding of the skills required for each profile in the job market, which is crucial for maintaining a skilled workforce and promoting economic growth.

6 Limitations and Future Outlook

Several limitations to this study should be considered when interpreting the results. Firstly, the data used for the analysis is limited to the job advertisements in the WIH dataset, which is an audit dataset and only represents a percentage of the overall job advertisements. The results of this study may not be generalisable to all job advertisements or labour market conditions, as the WIH dataset may not fully represent the entire labour market.

Secondly, the WIH data only considers ESCO skills according to version 0.8 of the taxonomy, whilst the novel ESCO v 1.1.1 introduces novel skills. This means the approach developed here should be considered preliminary work. The developed AI-based pipeline should be iterated to improve and enrich job profiles with novel skills over time and consider more OJAs. This is, indeed, the main room for improvement in this work.

References

Alabdulkareem, Ahmad, Morgan R. Frank, Lijun Sun, Bedoor AlShebli, César Hidalgo, and Iyad Rahwan. 2018. "Unpacking the polarization of workplace skills." *Science Advances* 4 (7).

Balassa, Bela. 1965. "Trade Liberalisation and "Revealed" Comparative Advantage." *The Manchester School* 33 (2):99–123. URL `https://onlinelibrary.wiley.com/doi/10.1111/j.1467-9957.1965.tb00050.x`.

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

Boselli, Roberto, Mirko Cesarini, Fabio Mercorio, and Mario Mezzanzanica. 2018. "Classifying online Job Advertisements through Machine Learning." *Future Gener. Comput. Syst.* 86:319–328. URL `https://doi.org/10.1016/j.future.2018.03.035`.

Colombo, Emilio, Fabio Mercorio, and Mario Mezzanzanica. 2019. "AI meets labor market: Exploring the link between automation and skills." *Information Economics and Policy* 47:27–37.

Frey, Carl Benedikt and Michael A. Osborne. 2017. "The future of employment: How susceptible are jobs to computerisation?" *Technological Forecasting and Social Change* 114 (Supplement C):254 – 280.

Giabelli, Anna, Lorenzo Malandri, Fabio Mercorio, and Mario Mezzanzanica. 2020. "GraphLMI: A data driven system for exploring labor market information through graph databases." *Multimedia Tools and Applications* :1–30.

———. 2022a. "WETA: Automatic taxonomy alignment via word embeddings." *Comput. Ind.* 138:103626. URL `https://doi.org/10.1016/j.compind.2022.103626`.

Giabelli, Anna, Fabio Mercorio, Mario Mezzanzanica, Francesco Trentini et al. 2022b. "Capturing the Skill Transformation Using Big Data and Artificial Intelligence: a Novel Change Index to Observe how the Italian Labour Market has been Changing over Time." In *The Relevance of Artificial Intelligence in the Digital and Green Transformation of Regional and Local Labour Markets Across Europe*. Nomos Verlagsgesellschaft mbH & Co. KG, 107–128.

Laursen, Keld. 2015. "Revealed comparative advantage and the alternatives as measures of international specialization." *Eurasian Business Review* 5 (1):99–115. URL `https://EconPapers.repec.org/RePEc:spr:eurasi:v:5:y:2015:i:1:p:99-115`.

Mikolov, Tomas, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. 2013. "Distributed representations of words and phrases and their compositionality." In *Advances in neural information processing systems*. 3111–3119.

van Dam, Alje, Andres Gomez-Lievano, Frank Neffke, and Koen Frenken. 2022. "An information-theoretic approach to the analysis of location and colocation patterns." *Journal of Regional Science* 63 (1):173–213.