Pillars – Pathways to Inclusive Labour Markets

Report

Speed and Direction of Changes of Skills Demand, within Occupations, across European Countries

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

Speed and direction of changes of skills demand, within occupations, across European countries

Emilio Colombo, Anna Giabelli, Fabio Mercorio,
Mario Mezzanzanica, Francesco Trentini

February 28, 2023

Contents

Executive Summary

Context of the action. In a historical moment characterised by radical changes in technology and innovation of productive processes, it is crucial to understand the impact these phenomena bear on job content, to design policies that minimise the negative outcomes on workers while suggesting the directions of up-skilling and re-skilling to involved workers. The study of how occupations change is usually bound to the availability of fine-grained data on job content and, most importantly, on the direction of demand of occupational profiles at the skill level. Online job ads (OJA) constitute a novel source of information that provides timely data on skill requirements by prospective employers. The unstructured nature of job ads induces challenges related to the classification of its content. Still, it is also possible to leverage it to study the relation among cited skills in a large volume of documents. This task explores this latter dimension by using AI-based methods to study skill similarity as it emerges from the texts of OJAs.

Main objectives. The deliverable is part of a work package that aims at developing tools and analytics to predict future trends of occupations and skills demand based on online job ads (OJAs). The specific role of task 5.2 is to apply AI-based methods, such as word-embeddings and machine learning, to study skills similarity and predict the direction of change in the skill set of occupations in different European regions.

Key activities. The desk research was conducted by elaborating a methodology to implement word-embeddings and novel measures of skill change using OJA data. The application is made in 5 European countries: France, Germany, Italy, The Netherlands and the UK.

Results of the action. Language models constructed using word-embeddings provide innovative tools to estimate the similarity among skills and quantify the proximity of novel skills to existing ones. We plan to make them available to the community in all European languages to allow the reproduction of the study. The methodology to estimate changes in the occupations' skill composition develops existing literature to improve the validity of results and allow comparative studies on the topic.

Impact of the action. We expect our contribution to impact the PILLARS consortium as well as the scientific community at large. The methodology is directly accessible within the consortium. Moreover, we plan to publish the trained language models, to make our results reproducible and empower the rest of the community to develop this strand of research.

Value added. We provide the first application of word-embeddings to economic analysis to study skills similarity. This methodology is innovative as it allows to capture of language specificities, which are a usual limitation of comparative studies based on unstructured texts in different languages, and, additionally, provides a way to quantify similarity among skills in a multilateral way, i.e. simultaneously estimating the proximity of each skill against all others.

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 demand expectations regarding skills and competences. The goals and results of this research are aligned 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

This task aims to study the speed and direction of change that occupations undergo regarding skills composition for different European regions. One peculiar aspect of this research is the application of AI methods, such as word embeddings and machine learning to conduct economic analysis. We measure changes in the occupation skill sets based on the words used to advertise job vacancies in online job ads (OJAs) to measure changes between and within European countries, considering occupations' specificity. Secondly, we build upon this novel evidence and estimate changes in the quality of demand for occupations, measured in terms of changes at the intensive and extensive margin, namely in the relevance of existing skills through time and the importance of novel skills that emerge through time.

We study how occupations change in France, Germany, Italy, the Netherlands and the UK over the period 2019-2021 by analysing the evolution of their skill set. We study the evolution of skill demand at the intensive and the extensive margin, namely in the relevance of skills in occupations and the emergence of new skills. We leverage established econometric techniques to estimate the relative impact of digital, professional and soft skills in the change at the intensive margin and innovative AI methods to document the speed of change of the skill set at the extensive margin.

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; then 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 first use case is online job ads (OJAs) collected from the 32 European countries in the European Statistical System. The PILLARS consortium has

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

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

full access to the dissemination area of the Datalab, the experimental microdata for research produced by the WIH-OJA.

Different datasets are used in the analysis, depending on the activity conducted. Considering the task of creating the language models to compute the similarity among existing and novel skills, we used a sample of OJAs that includes text and description.[5]. The sample was created by Eurostat and followed a stratification by language and by seven variables under which the OJAs were classified: occupation, type of contract, salary, working time, education, required experience, and economic activity of the prospective employer. For each stratum, 50 OJAs were selected so that in the first release, the sample counted 9,262,005 observations on 25 languages for 2020Q2 and 2021Q1. Considering the analysis of skill change, we used the classified online job ads presented in the dissemination area of the Datalab[6]. The collected OJAs are classified according to ESCO (European Skills, Competences, Qualifications and Occupations), so that the textual content of the collected web pages is reduced to a standard classification of occupations, skills and other relevant variables. We restrict our analysis to the European countries represented by partners of the consortium, namely France, Germany, Italy, the Netherlands and the UK. We assign each skill to a group based on belonging to one of the hierarchical groups provided by ESCO v1.0.8. We exclude language skills from our analysis and constrain the professional (also called hard non-digital skills) to those labelled as essential or optional by ESCO[7].

The ESCO taxonomy allows grouping skills into categories for better representation and analysis. The first major distinction is between hard and soft skills. Hard skills are typically job-specific skills and competencies needed to perform a specific job or task (examples are knowledge of specific software or instruments, specific manual abilities, etc.) Soft skills, on the other hand, are transversal and refer to the capacity of individuals to interact with others and the environment (examples are communication skills, problem-solving, etc.). Within hard skills, we further distinguish between hard digital skills (HD) and hard non-digital (HND) skills. Digital skills encompass various abilities that allow an individual to use ICT tools at different levels, from using, manipulating and interacting with standard ICT tools to designing, implementing and deploy-

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

[6]The release `wih_oja_blended_v2_2022q3_r20221104` was used.

[7]ESCO skills are linked with occupations at the ESCO 0 level, that corresponds to either ISCO08 5° or 6° digit. Being the Datalab micro data classified at the ISCO08 4° digit, we select the optional or essential professional skills for any occupations belonging to the considered ISCO08 4° digit.

ing complex ICT systems and services. Hard non-digital skills are professional skills that characterising for an occupation and are sufficient to differentiate one occupation from another. Our classification builds upon ESCO groups, namely essential/optional skills, language skills, transversal skills, and digital skills, with a custom classification which extends the ESCO classification for digital and transversal skills.

Moreover, we filter out occupations of the sixth ISCO08 Major Group (Skilled Agricultural, Forestry and Fishery Workers) for lack of observations in the OJAs data. We also filter out the ISCO08 4° digit occupations of the "not elsewhere classified" groups. The composition of our dataset for skill analysis is reported in table 1.

Table 1: Number of OJA used in the analysis

Country	2019	2021	All
DE	5,642,670	6,711,163	12,353,833
FR	2,264716	7,178,613	9,443,329
IT	876,058	2,776,858	3,652,916
NL	462,405	2,408782	2,871,187
UK	2,627,806	8,125,521	10,753,327
All	11,873,655	27,200,937	39,074,592

We use various measures to compare occupations through time and explore the extent and intensity of the change in their skill sets.

2.1 Descriptive and AI-based measures of skill change

Skill change can be measured descriptively by counting the number of unique skills that are observed in 2021 compared to the skill set observed in 2019. The implicit assumption is that all skills bear the same weight in the overall change in the skill set. The first contribution of our study is to provide a methodology and estimates of similarity between existing and novel skills to assess the degree of innovation that occupational profiles are undergoing in the economy, using word embeddings. This method has several advantages in accounting for many orders of problems that emerge in comparing skill composition and skill change among countries and through time. The main issue is a linguistic dimension to be considered in the analysis. Each language has peculiar syntactical and semantic structures that are idiosyncratic and, metaphorically speaking, define the density of the language. Words are therefore embedded in a network of relations with other words. The second motivation for the use of word embedding models trained on OJA descriptions is to broadly capture the proximity among skills in the text, therefore, accounting for their co-occurrence and producing a multilateral measure of skill bundling. The language model, therefore, offer a summary of the proximity relation among skills embedded in the discourse of job postings. As the model trained on the whole corpus of OJA in a language, we capture discourse about skills independently of the occupational groups.[8] Moving from these premises, we attempted an innovative strategy to use language models trained on OJA descriptions to estimate the degree of similarity between skills. In practice, these language models are independent of occupations and skills classifications, the ESCO dictionaries in our case, but can be queried with couples of ESCO skills to obtain the degree of similarity they bear to each other given the trained language model. Similarity is measured using the cosine similarity between each couple of inputted skills.

Considering the language model itself, we trained 100 models for each language and selected the best embedding for the task of skill classification, following Giabelli et al. (2022b,a). The method performs the principal intrinsic evaluation of tasks over different benchmarks and generates a global intrinsic evaluation score. This score summarizes the performance of the word embedding models over different tasks and benchmarks, providing the user with a comprehensive evaluation for selecting the word-embedding model to be used. In our case the task for which the best embedding model was selected was skill classification.

Measures emerging from embedding models are not directly comparable, because, at this stage, we cannot determine the density of the vector space within which each linguistic model is represented and standardise it to accomplish comparability. This issue is well known in the literature under the name of embedding alignment, but there is not a consensus on the best strategy to solve it. One of the most promising directions of research is anchors alignment. This method requires finding several words with the most exact translation possible with a benchmark language so that the correspondence between these words provides an anchoring between the language models. Then the distance between anchors within each language model is computed to determine the vector space's

[8] An interesting extension would be to train language models by occupational groups to test differences in each skill's similarity in different contexts. This process would require a strong assumption that is not required in our case, i.e., to taken as given the classification of occupations.

density and is used to standardise it. The outcome are comparable language models, assuming exact correspondence between anchors and the coherence of the underlying textual corpora on which word embeddings are generated. Unfortunately, this research is still too early to provide methods useful to us. Still, we can overcome this limitation by relying on statistical methods that work in analogy with the standardisation provided by the anchor's alignment method. Computing the cosine similarity between the novel and existing skills, we obtain different distributions of similarities that can be used as proxy of the density of the language-specific vector space of the word embedding. Figure 1 presents the smoothed distribution of similarities between existing and novel skills observed in OJAs in all language model.

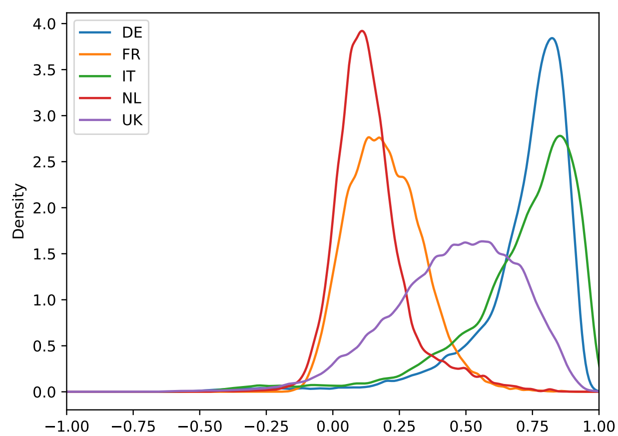


Figure 1: Distribution of similarity between exsting and novel skills, by country.

Despite this is not a statistical test and the representation is run on a selection of the whole language model, level differences emerge between countries. A second-best solution to obtain comparability between language models is therefore to compute a measure of central tendency, such as the median, and calculate the linear distance of each observation from this value.

The relative deviation provide, albeit in qualitative terms, the degree of similarity between existing and novel skills. We summarise the skill similarity by groups of belonging of existing and new skills, so that we can compare hard-digital, hard non-digital and soft skills.

Country	Median similarity
DE	0.771
FR	0.187
IT	0.765
NL	0.122
UK	0.481

Table 2: Median similarity among the observed existing and novel skills.

2.2 RCA-based measures

Our measures are rooted in the revealed comparative advantages (RCA) concept. This class of indicators was first developed in the international trade theory. They used post-trade measures to infer the comparative advantages in the production of traded goods of a region in a multilateral context (Balassa, 1965). More recently, RCA has been widely used in the economic geography literature to describe regions' specialisation in productive activities (see van Dam et al. (2022) for an overview of the literature and the most recent advances). In analogy with this solution, we use measures of skill frequency in each occupation as it emerges from online job ads and calculate the degree of specialisation each occupation has in each skills. We follow Alabdulkareem et al. (2018), who calculate occupations' skills specialisation by leveraging on O*NET skills classification and relevance as factors in the calculation of revealed comparative advantages. To replicate this method on European OJA data, we need an alternative measure to the relevance directly measured in O*NET, a measurement not provided by ESCO skill classification. Following Giabelli et al. (2022c), we measure the skill relevance by using skills' frequencies in OJAs. Then skill relevance is computed as the frequency of a given skill in the job ads of a specific occupation relative to the skill's frequency in job ads in all other occupations. For each year of observation t, we define a set of occupations $\bar{O}_t = \{o_k, k = 1, \ldots, K_t\}$ and a set of skills for each of them $\bar{S}_{kt} = \{s_j, j = 1, \ldots, J_t\}$. We define the skill frequency as:

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

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:

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

which ranges between $[0, +\infty)$.

The distribution of the RCA is not symmetric around its neutral point 0. We adopt the transformation proposed by Laursen (2015) and calculate a non-linear transformation of the RCA called Revealed Symmetric Comparative Advantages:

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

The $srca_{ij}$ is uniform over the continuous interval $[-1, 1]$, and its neutral point is set to 0; values in the interval $(0, 1]$ mean that occupation i is specialised in skill j, while values in the interval $[-1, 0)$ identify the skill is requested.

On the basis of the Balassa's RCA, we build three measures. The first two develop upon the previous research by Giabelli et al. (2022c). The first measure is the change in the core-skills set, which captures the change in the specialisation of occupation in skills that persist over time. The second measure is the change in the novel-skills set, which captures the degree of innovation in the skill repertoire and its relevance in a single indicator. The two measures are described in sections 2.2.1 and 2.2.2, respectively. Finally, we introduce a new measure of job complexity that uses a transformation of the RCA to expose the prevalence of characterising skills of an occupation. The measure is exposed in section 2.2.3.

2.2.1 Changes in the core-skills set

The first set of measure captures changes in the specialisation of the core-skills, namely those that are persistent throughout the windows of analysis. We measure the distance between the core skill set with the weighted Jaccard distance, measured on the intersection set of skills observed in 2019 and 2021. Given that the weighted Jaccard distance can be computed on strictly positive sets, we shift the $rsca_{ij}$ to the positive domain: $rsca'_{19ij} = rsca_{ij} + 1$, $rsca'_{19ij} \in [0, 2]$. Being S_{ti} the skill set in year t for occupation i, our measure of change in the core-skill set is:

$$\Delta S_i^C = 1 - \frac{\sum_{j \in S_{19i} \cap S_{21i}} \min(rsca'_{19ij}, rsca'_{21ij})}{\sum_{j \in S_{19i} \cap S_{21i}} \max(rsca'_{19ij}, rsca'_{21ij})} \tag{4}$$

ΔS_i^C ranges from $[0, 1]$, being 1 complete similarity between years and 0 complete difference.

2.2.2 Changes in the novel-skills set

Analogously we define the measure of changes in the novel-skill set as the share of RCAs pertaining to new skills in occupation i on the sum of RCAs of skills observed in 2019 and 2021 in that occupations:

$$\Delta S_i^N = \frac{\sum_{j \in S_{21i} - S_{19i}} rsca_{ij}}{\sum_{j \in S_{21i}} rsca_{ij}} \tag{5}$$

2.2.3 Job complexity

Starting from RCA we can compute the *effective use* of the skill for the occupation:

$$e_{ij} = \begin{cases} 1 & rsca_{ij} > 0 \\ 0 & otherwise \end{cases} \tag{6}$$

e allows comparing the relative importance of a skill s_j to an occupation o_i (the numerator in equation rca_{ij}) to the expected relative importance of that skill on aggregate (the denominator in rca_{ij}). Intuitively when $rsca_{ij} > 0$ means that occupation o_i relies on skill s_j more then expected given its prevalence in total labour demand.

We use effective use to introduce a new measure of job complexity. Job complexity is a concept difficult to measure and conceptualize. The literature so far has relied mainly on subjective measures obtained from surveys (Bai, Tian, and Liu, 2021) with all the limitations that subjective measures imply (biases, lack of comparability etc.). We provide an operational measure that complies with the conceptualization provided by Wood (1986) by defining job complexity as the share of skills with an effective use higher than 1. In other words, more complex jobs require a larger number of characterising skills, namely those skills that are more concentrated in the considered occupation.

3 Findings

In this section, we present the results of the analysis, separated by the method. Subsection 3.1, we expose the outcomes of the descriptive analysis using word-embeddings, while in subsection 3.2 we present the results of the econometric analysis involving measures of change in the core and novel skill set and in subsection 3.3 the econometric analysis of job complexity.

3.1 Descriptive measures

Tables 3 to 7 show the results by country, occupation and skill groups. Colors in the column "Similarity" are useful to identify qualitative results, given that an exact correspondence between cardinal values between countries cannot be achieved in language models despite standardisation.[9]

Considering the count of unique skills (columns "Skill count" and "New skills" of each table) we observe that there is a richness of unique skills in each country. Overall the largest variety of skills is observed among high skill occupations (Professionals (2) and Associate professionals and Technicians (3), and Managers (1)), followed by middle and low skill occupations (in order, Clerical Support Workers (4), Service and Sales Workers (5), Craft and Related Trades Workers (7), Plant and Machine Operators, and Assemblers (8) and Elementary occupations (9)). The UK leads the rank in terms of the largest variety of observed skills. Nonetheless, UK itself is the lowest scoring in terms of innovation, as it shows the lowest absolute and relative number of new skills for almost all occupations, except ISCO08 Major Group 9, which displays a degree of change comparable to other countries. On the contrary, countries with the lowest number of unique skills in 2019 or 2021, namely Italy and the Netherlands, rank higher in terms of absolute and relative change. The skill set composition in terms of the type of skills varies by occupation at the country level. Nonetheless, there are some general trends. Professionals (2), Associate professionals and technicians (3) are displaying the highest incidence of hard digital (HD) and Hard non-digital (HND) skills, and the change in their skill set is primarily due to HND skills. On the contrary, occupations of ISCO08 Major Groups 1, 4, 5, 7, 8 and 9 are those that are changing more in terms of HD skills. Some notable exceptions are the UK that in general display low change in high skill occupations , and France and Germany who exhibit high degree of innovation on occupations in middle and low skill occupations. It is also interesting to note that Germany also shows an important increase in the variety of soft skills requested for occupations in the same groups.

[9]ISCO08 Major Groups are Managers (1), Professionals (2), Associate professionals and technicians (3), Clerical Support Workers (4), Service and Sales Workers (5), Craft and Related Trades Workers (7), Plant and Machine Operators, and Assemblers (8) and Elementary occupations (9). We also group them in high skill occupations (ISCO08 Major Groups 1 to 3), medium skill occupations (ISCO08 Major Groups 4 and 5) and low skill occupations (ISCO08 Major Groups 7 to 9).

Skill Similarities. We now focus on skill similarity between existing and novel skills (column "Similarity"). High similarity means that the observed skills have high proximity in the corpus of OJA descriptions and are conditional to the language dictionaries that are used to enrich the embedding. Therefore we can interpret negative values, where the cells are green colored, as the cases in which new skills are more innovative in comparison to existing ones, and the darker the color, the highest the degree of innovation; on the contrary, positive values, where the cells are red colored, indicate cases of highest innovation compared to existing skills. In terms of the method of reading such results, it is relevant to separately consider the comparison between existing and novel skills within the same group, represented by the diagonal elements of each 3x3 Major Group-occupation sub-matrix, from the between-groups similarities.

The first evidence to show is that overall the similarity between existing and new skills belonging to the same group scores higher than median, which is expected, given that any new additional skill is, overall, more similar to skills in the same domain to skills in other domains. Nonetheless, notable exceptions in magnitude vary with occupations and countries. In France the between-groups similarity is the highest observed. It is worth noting that HD skills for Clerical Support Workers (4), Craft and Related Trades Workers (7), and Elementary occupations' workers (9) are the most similar to the HD skills already present meaning that the degree of novelty of these skills is low; on the contrary, the highest novelty is observed in S skills compared to HD skills and S skills compared to HND skills. Occupations in the Professional (2) Craft and Related Trades Workers (7), Plant and Machine Operators, and Assemblers (8) groups share a high degree of novelty of HND skills with skills of all types. New S skills tend to highly innovate compared to existing HD and HND skills.

In Germany, a clear pattern is visible for all ISCO08 Major Groups, highlighting that new S skills bear a lower degree of innovation compared to the existing S skills. On the contrary, there is a simultaneous trend in new HD and HND skills, that tend to innovate compared to existing skills, especially for high skill occupations. In Italy, S skills account for the lowest degree of novelty and innovation compared to existing ones for high skill occupations, namely high skill occupations. The highest degree of innovation is observed for Plant and Machine Operators, and Assemblers (8) and Elementary occupations (9), particularly for HD skills with all types of skills, indicating a radical innovation for these occupations. A similar trend, albeit at a lower degree of intensity, is observed for high skill occupations, and Craft and Related Trades Workers (7) indicating that digitalisation is a characterising trend in the evolution of Italian occupational profiles and is happening in a radical way with regard to existing skills. In the Netherlands, skills innovation in middle and low skill occupations is very limited, both for HD and S skills, differently for what is observed for other countries.

In the UK, overall there is a low degree of innovation among new skills, which strengthens the result on the count of unique skills: change is not only limited in terms of novelty but also the degree of connection of new skills with existing skills is low. The highest novelty is concentrated in S skills. It is worth noting that for Professionals (2) HND skills are the most innovative skills, while differently from other countries HND skills bear a low degree of innovation for Clerical Support Workers (4).

Table 3: Skill composition and change in France, 2019-2021

		Skills count		New skills		Similarity		
ISCO08	**Skill type**	**2019**	**2021**	**N**	**%**	**HD**	**HND**	**S**
	HD	141	151	31	22%	16,2%	-17,7%	-14,7%
1	**HND**	118	122	23	19%	-17,4%	5,0%	-2,5%
	S	120	102	13	11%	-4,1%	20,3%	15,5%
	HD	227	216	10	4%	13,2%	-29,2%	-14,9%
2	**HND**	253	273	43	17%	-32,2%	-7,4%	-23,0%
	S	129	119	20	16%	-8,1%	-17,8%	16,1%
	HD	188	175	14	7%	15,2%	-15,2%	-6,8%
3	**HND**	188	208	39	21%	-22,3%	8,5%	-10,8%
	S	116	113	12	10%	-9,9%	-6,0%	23,1%
	HD	86	94	30	35%	43,5%	-0,3%	-12,5%
4	**HND**	46	49	12	26%	-13,7%	11,4%	8,1%
	S	75	88	21	28%	-2,1%	27,1%	16,9%
	HD	88	118	45	51%	10,4%	-18,4%	-17,2%
5	**HND**	57	57	10	18%	-20,8%	22,7%	-5,1%
	S	79	94	21	27%	-8,7%	20,0%	12,2%
	HD	77	89	29	38%	47,8%	-3,5%	-5,6%
7	**HND**	65	83	21	32%	4,8%	23,1%	-18,2%
	S	63	68	17	27%	-2,3%	-13,1%	2,9%
	HD	52	68	28	54%	6,2%	-25,4%	-30,6%
8	**HND**	43	46	10	23%	-9,2%	16,7%	-8,4%
	S	71	83	21	30%	-4,5%	-33,3%	8,8%
	HD	32	43	25	78%	32,3%	-21,5%	-19,7%
9	**HND**	17	22	6	35%	-0,9%	48,0%	4,4%
	S	42	58	18	43%	0,5%	-32,0%	1,9%

Table 4: Skill composition and change in Germany, 2019-2021

		Skills count		New skills		Similarity		
ISCO08	**Skill type**	**2019**	**2021**	**N**	**%**	**HD**	**HND**	**S**
	HD	183	156	16	9%	-4,8%	-6,8%	-0,3%
1	**HND**	168	140	13	8%	-7,3%	-3,1%	0,5%
	S	163	122	15	9%	-2,9%	0,3%	6,3%
	HD	250	233	5	2%	-5,1%	-8,5%	-3,8%
2	**HND**	356	322	28	8%	-8,8%	-7,2%	-6,6%
	S	186	133	9	5%	1,0%	-1,7%	5,8%
	HD	211	188	7	3%	-4,5%	-8,2%	-4,8%
3	**HND**	229	222	28	12%	-6,5%	-6,7%	-3,7%
	S	138	117	11	8%	0,7%	-1,4%	5,1%
	HD	121	84	11	9%	2,4%	1,8%	0,2%
4	**HND**	67	63	9	13%	-0,4%	3,2%	4,3%
	S	88	79	15	17%	3,8%	5,3%	7,1%
	HD	139	105	17	12%	-1,5%	-2,1%	-0,9%
5	**HND**	72	67	9	13%	-0,6%	0,5%	5,8%
	S	116	93	14	12%	1,1%	4,3%	6,3%
	HD	131	116	19	15%	-1,8%	-4,3%	-3,4%
7	**HND**	88	96	14	16%	-4,0%	-6,5%	-0,7%
	S	79	79	17	22%	2,1%	-0,6%	5,2%
	HD	128	114	15	12%	-2,7%	-3,2%	-0,6%
8	**HND**	64	67	11	17%	-7,3%	-2,5%	-3,4%
	S	102	90	13	13%	0,5%	2,3%	6,3%
	HD	77	68	18	23%	-1,6%	-20,8%	-5,1%
9	**HND**	19	22	5	26%	-0,3%	1,4%	3,6%
	S	46	55	17	37%	0,6%	-4,2%	5,0%

Table 5: Skill composition and change in Italy, 2019-2021

		Skills count		New skills		Similarity		
ISCO08	**Skill type**	**2019**	**2021**	**N**	**%**	**HD**	**HND**	**S**
	HD	103	91	13	13%	-5,0%	-3,9%	5,2%
1	**HND**	112	105	15	13%	-0,2%	0,8%	4,3%
	S	101	88	12	12%	-2,8%	0,4%	4,0%
	HD	217	210	16	7%	-7,6%	-4,4%	1,4%
2	**HND**	252	252	32	13%	3,7%	1,2%	2,9%
	S	117	107	18	15%	-1,7%	-1,2%	2,5%
	HD	163	162	18	11%	-8,1%	-5,3%	1,7%
3	**HND**	163	182	33	20%	-2,4%	-2,0%	2,0%
	S	106	95	11	10%	-3,0%	-4,2%	2,3%
	HD	53	66	30	57%	-1,3%	-4,2%	5,8%
4	**HND**	51	49	5	10%	-2,8%	7,2%	2,1%
	S	67	75	14	21%	1,8%	-1,0%	1,1%
	HD	40	49	27	68%	0,4%	-5,1%	4,0%
5	**HND**	52	63	14	27%	2,5%	2,5%	4,2%
	S	66	73	17	26%	0,4%	-2,2%	-1,0%
	HD	41	53	21	51%	-1,4%	1,7%	0,7%
7	**HND**	64	76	18	28%	-2,7%	0,6%	3,7%
	S	44	57	19	43%	-2,2%	-1,7%	0,2%
	HD	27	30	13	48%	-21,1%	-10,3%	-4,1%
8	**HND**	38	55	20	53%	-7,3%	5,2%	3,9%
	S	54	61	18	33%	-6,2%	3,0%	0,2%
	HD	9	11	6	67%	-16,9%	-33,9%	-16,4%
9	**HND**	22	23	2	9%	2,6%	8,1%	6,5%
	S	30	40	15	50%	-8,4%	-15,6%	-1,0%

Table 6: Skill composition and change in the Netherlands, 2019-2021

		Skills count		New skills		Similarity		
ISCO08	**Skill type**	**2019**	**2021**	**N**	**%**	**HD**	**HND**	**S**
	HD	109	108	21	19%	4,3%	-22,1%	-16,0%
1	**HND**	115	114	20	17%	-11,2%	10,0%	2,3%
	S	117	100	15	13%	-20,0%	16,5%	27,7%
	HD	224	211	8	4%	6,4%	-28,6%	-22,2%
2	**HND**	262	263	42	16%	-14,5%	6,6%	-21,6%
	S	143	107	15	10%	-17,0%	-15,5%	23,1%
	HD	143	144	22	15%	25,2%	-17,9%	-17,4%
3	**HND**	164	186	39	24%	-15,6%	16,1%	-11,7%
	S	109	102	14	13%	-22,0%	-10,3%	27,3%
	HD	59	78	28	47%	24,1%	-15,4%	-7,5%
4	**HND**	56	58	10	18%	-11,8%	13,5%	7,1%
	S	79	94	28	35%	-6,1%	7,1%	33,1%
	HD	59	95	40	68%	11,5%	2,4%	-7,7%
5	**HND**	45	55	12	27%	-22,1%	36,8%	10,9%
	S	80	97	30	38%	-18,4%	33,3%	32,2%
	HD	67	74	23	34%	31,4%	-4,6%	-12,5%
7	**HND**	63	74	15	24%	-5,5%	54,6%	-9,6%
	S	53	61	15	28%	-13,7%	-5,9%	35,6%
	HD	63	81	34	54%	5,0%	-13,1%	-12,9%
8	**HND**	44	48	8	18%	-8,1%	26,9%	0,8%
	S	96	92	21	22%	-23,5%	-14,2%	25,0%
	HD	41	61	26	63%	33,9%	1,4%	-4,3%
9	**HND**	12	16	5	42%	-6,4%	47,4%	17,3%
	S	50	63	20	40%	-23,9%	0,2%	34,4%

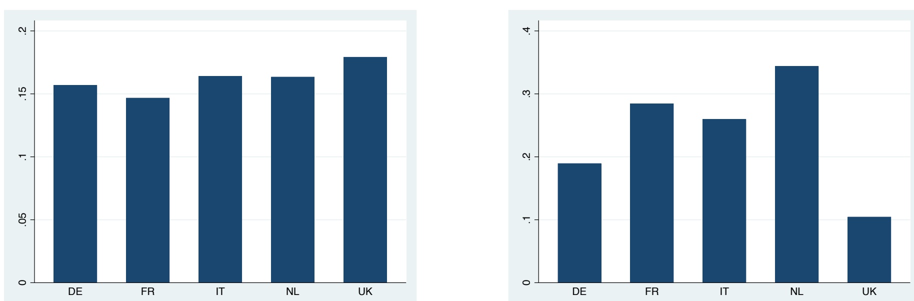
Table 7: Skill composition and change in the UK, 2019-2021

		Skills count		**New skills**		**Similarity**		
ISCO08	**Skill type**	**2019**	**2021**	**N**	**%**	**HD**	**HND**	**S**
1	**HD**	242	194	3	1%	-0,7%	1,2%	-5,6%
	HND	307	232	13	4%	-1,5%	14,2%	-0,6%
	S	384	268	7	2%	2,1%	7,8%	-1,6%
2	**HD**	277	259	1	0%	-0,2%	-6,0%	-5,4%
	HND	667	585	27	4%	-4,0%	-2,7%	-5,1%
	S	399	286	8	2%	4,3%	-2,0%	0,2%
3	**HD**	259	226	2	1%	1,0%	-5,8%	-5,0%
	HND	475	418	35	7%	-2,1%	2,9%	-3,4%
	S	394	284	6	2%	4,1%	-0,5%	0,2%
4	**HD**	154	143	20	13%	5,8%	17,5%	-2,4%
	HND	109	98	4	4%	4,7%	27,3%	-3,1%
	S	336	231	9	3%	4,8%	18,6%	-2,2%
5	**HD**	135	131	18	13%	1,1%	-1,9%	-4,3%
	HND	140	128	11	8%	-1,4%	-4,5%	-6,1%
	S	296	232	17	6%	5,8%	4,0%	-2,2%
7	**HD**	152	149	22	14%	-1,6%	-1,0%	-4,8%
	HND	163	163	23	14%	-0,6%	3,6%	-5,4%
	S	266	197	16	6%	1,5%	1,7%	-2,8%
8	**HD**	213	184	10	5%	-2,1%	-7,4%	-6,9%
	HND	114	118	16	14%	5,2%	9,1%	-3,1%
	S	325	220	11	3%	4,5%	1,9%	-2,2%
9	**HD**	70	87	29	41%	-4,5%	4,1%	-0,8%
	HND	39	52	17	44%	-4,6%	4,9%	-6,9%
	S	210	179	14	7%	0,9%	-0,7%	-1,4%

3.2 Change in core and novel skill sets

We have computed the measures presented above for each 4 digit occupation and then aggregated at country level or by ISCO08 Major Group.

We start by presenting some overall measures of change. Figure 2 shows the core and novel skill set change, by country, between 2019 and 2021. Panel a presents the results for the change in the core set while panel b for the novel skill set. Remember that due to the different computation methods, despite being normalised, the two measures of change are not comparable, although each of them allows cross-country comparison. The figure shows the relatively modest effect of the change in the novel set in the UK. This is probably due to the fact that the UK has a relatively high skill yield as documented in the previous paragraph, therefore the contribution of new skills is limited. This is clear particularly for high skill occupations and less relevant for low skill occupations as documented in figure 3.



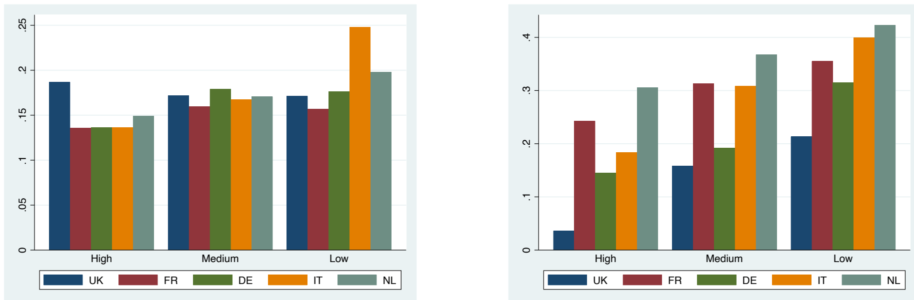
(a) Change in the core set

(b) Change in the novel set

Figure 2: Overall change in the core and novel skill set, by country, 2019-2021

Figure 3 reports the measures of change aggregating across high, medium and low skill occupations.

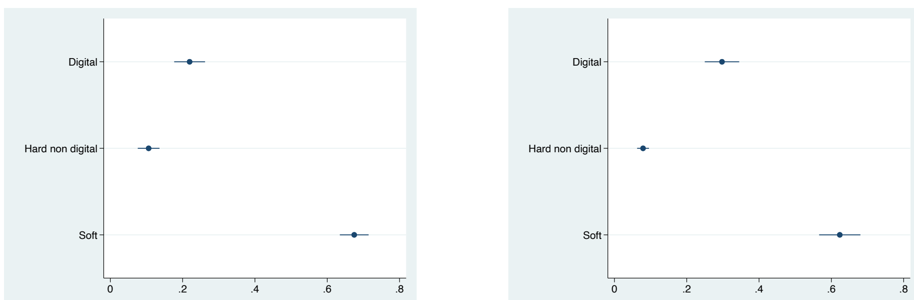
In Figure 4 we decompose the change in the core and skill set by considering the contribution to changes in different groups of skills. As stated in the previous section, one of the limits of measures such as RCA is that they are not linearly additive. Therefore it is not possible to decompose the variation of the core and of the skill set in its components. In order to overcome this problem, we have estimated the contribution of each component to the overall measure of change by estimating a constrained regression model. More in detail we have computed the change in the core and novel skill set for each subgroup of skills (HD, HND and S), and we have regressed the overall measure of change over



(a) Change in the core set

(b) Change in the novel set

Figure 3: Change in core and novel skill set by occupation groups and country



(a) Change in the core set

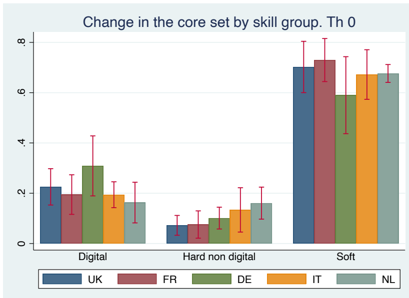
(b) Change in the novel skill set

Figure 4: Determinants of the change in core and novel skill set, overall, 2019-2021

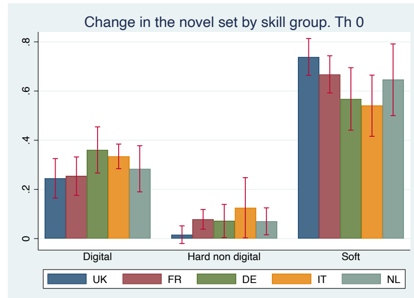
the measures of each group, constraining that the sum of the coefficient of each group add to 1.[10] The figure shows that soft skills are responsible for most of the variation accounting for more than 60% of the change in both the core and novel skill sets. Digital skills account for 22 and 30% respectively of the change in the core and novel skill sets while the contribution of Hard Non Digital skills are limited to 10% for the core set and 7% for the novel skill set. The results are not surprising considering that soft skills are becoming pervasive even for technical occupations , while digital skills include several innovative technologies and programming languages that both account for new skills as well as variations in existing skill sets.

Finally Figure 5 reports the results by country confirming the overall pattern.

[10]In figure 4 errors are clustered at country level

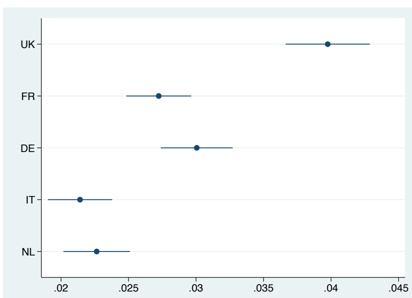


(a) Change in the core set

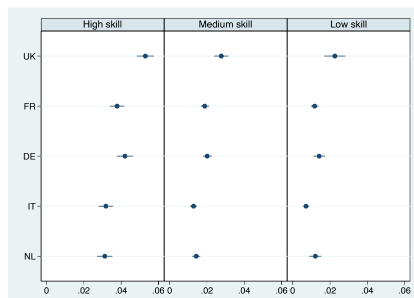


(b) Change in the novel skill set

Figure 5: Determinants of the change in core and novel skill set, by country, 2019-2021



(a) Complexity, by country, 2019



(b) Complexity, by occ. and country

Figure 6: Job complexity, 2019

3.3 Job complexity

In this section, we present the results of job complexity. Figure 6 reports the measure of complexity by country in 2019. Overall the UK is characterised by more complex jobs i.e. jobs that require a higher number of skills *effectively used*. This difference persists across occupation groups (panel b) but is more pronounced for high skill occupations.

Finally figure 7 display the change in job complexity between 2019 and 2021. The figure suggests a possible catching up with countries characterised by lowest levels of complexity in 2019 having the largest increase. In order to better investigate this issue we have performed a simple "convergence" regression by estimating the following regression

$$\Delta Complexity_i = \alpha + \beta Complexity19_i + \gamma X_i + epsilon_i \tag{7}$$

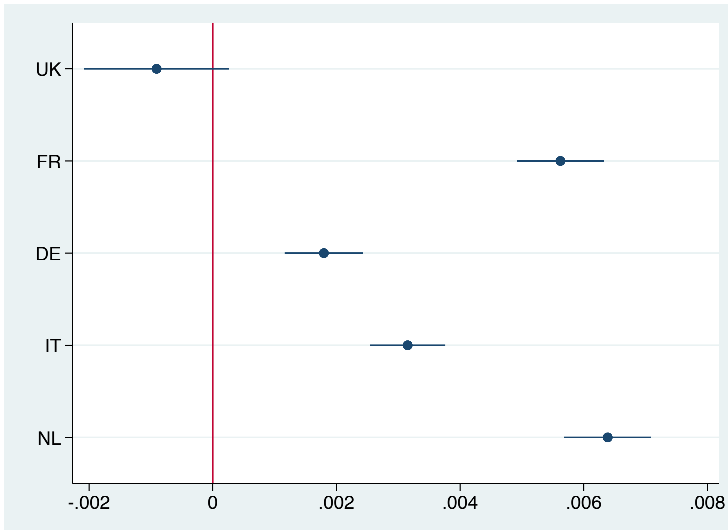


Figure 7: Change in job complexity, by country

where X is a vector of control variables. Table 8 shows that there is indeed some sort of convergence to a higher level of job complexity. Countries and occupations that were in 2019 less complex are becoming more complex faster. The analysis also shows that the increase in complexity is the result of the change in the novel skill set (mainly digital and soft) while the change in the core set is negatively associated with the change in complexity. In other words, the change in complexity is driven by the extensive margin of the set of skills.

4 Conclusions

By analyzing changes in the demand for skills across different industries, regions, and occupations, we can better understand the impact of new technologies on the labour market. While most occupations will not disappear, the skills required for them will change. By analyzing OJAs' descriptions, we can identify these changes and estimate the speed and direction of skill demand within occupations and between regions. In order to explore these different dimensions of change, we rely on descriptive statistics as well as econometric techniques. The relevant innovations in our approach are manifold.

First, we used AI algorithms to create language models trained on OJA descriptions to compute similarities between skills and track the magnitude of their change over time.

Table 8: Change in job complexity

	reg2	reg1
Complexity 2019	-0.091***	-0.107***
	(0.013)	(0.015)
Δ core	-0.009***	
	(0.001)	
Δ core HD		-0.006***
		(0.002)
Δ core HND		-0.001
		(0.001)
Δ core S		-0.017***
		(0.002)
Δ novel	0.009***	
	(0.001)	
Δ novel HD		0.007***
		(0.001)
Δ novel HND		0.000
		(0.001)
Δ novel S		0.012***
		(0.002)
r2	0.288	0.512
N	1695	983

Note: Dependent variable is Change in job complexity. OLS estimates. Robust standard error reported in brackets. Regression include country and Isco 1d dummies (not reported) * p<0.05, ** p<0.01, *** p<0.001

Secondly, we developed existing literature on skill specialisation to suggest a novel measure of skill change at the intensive and extensive margin, providing an econometric solution to the additivity issue that constrain existing measures of skills specialization when considering aggregates such as digital, soft, and hard-non-digital skills. As a result, we were able to estimate the contribution of each group of skills to the overall changes at both margins.

Third, we build a measure of job complexity based on the relevance and variety of the skills used in it.

Overall we find that even in a short period of time (2 years), the skill composition of jobs changes substantially, approx 15% for the core set (intensive margin) and 20% for the novel skill set (extensive margin). This result is driven in particular by soft skills which are becoming pervasive also in technical occupations, and by digital skills. Finally, we document an increase in job complexity driven mainly by novel skills.

Limitations and further developments. The research developed in this task introduces novel methods based on textual analysis for the study of economic phenomena using unstructured data. Moreover, it leverages high-frequency, high-detail data produced from OJA. Both dimensions are still subject to research and will undergo improvements in the future. Embedding alignment is expected to be an important contribution to granting a correct generalisation of results in comparative research. At the same time, our method currently provides the second-best solution to the issue. Further developments are also expected on the number of languages and countries covered: the method is scalable, while language idiosyncrasies are still expected to impinge on its direct generalisation, requiring ad hoc analyses. Considering the use of classified OJA, the RCA-based measures presented here are currently not additive, which implies that contribution to skill change needs to be estimated. Further work will be necessary to develop measures that allow the studying of phenomena at different levels of aggregation, such as regions or aggregates of occupations while preserving the comparability with results at a higher or lower level of aggregation.

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).

Bai, Jing Yi, Qing Tian, and Xia Liu. 2021. "Examining Job Complexity on Job Crafting Within Conservation of Resources Theory: A Dual-Path Mediation Model." *Frontiers in Psychology* 12.

Balassa, Bela. 1965. "Trade Liberalisation and "Revealed" Comparative Advantage." *The Manchester School* 33 (2):99–123.

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`.

Giabelli, Anna, Lorenzo Malandri, Fabio Mercorio, and Mario Mezzanzanica. 2022a. "WETA: Automatic taxonomy alignment via word embeddings." *Comput. Ind.* 138:103626. URL `https://doi.org/10.1016/j.compind.2022.103626`.

Giabelli, Anna, Lorenzo Malandri, Fabio Mercorio, Mario Mezzanzanica, and Navid Nobani. 2022b. "Embeddings Evaluation Using a Novel Measure of Semantic Similarity." *Cogn. Comput.* 14 (2):749–763. URL `https://doi.org/10.1007/s12559-021-09987-7`.

Giabelli, Anna, Fabio Mercorio, Mario Mezzanzanica, Francesco Trentini et al. 2022c. "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.

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.

Wood, Robert E. 1986. "Task complexity: Definition of the construct." *Organizational Behavior and Human Decision Processes* 37 (1):60–82.