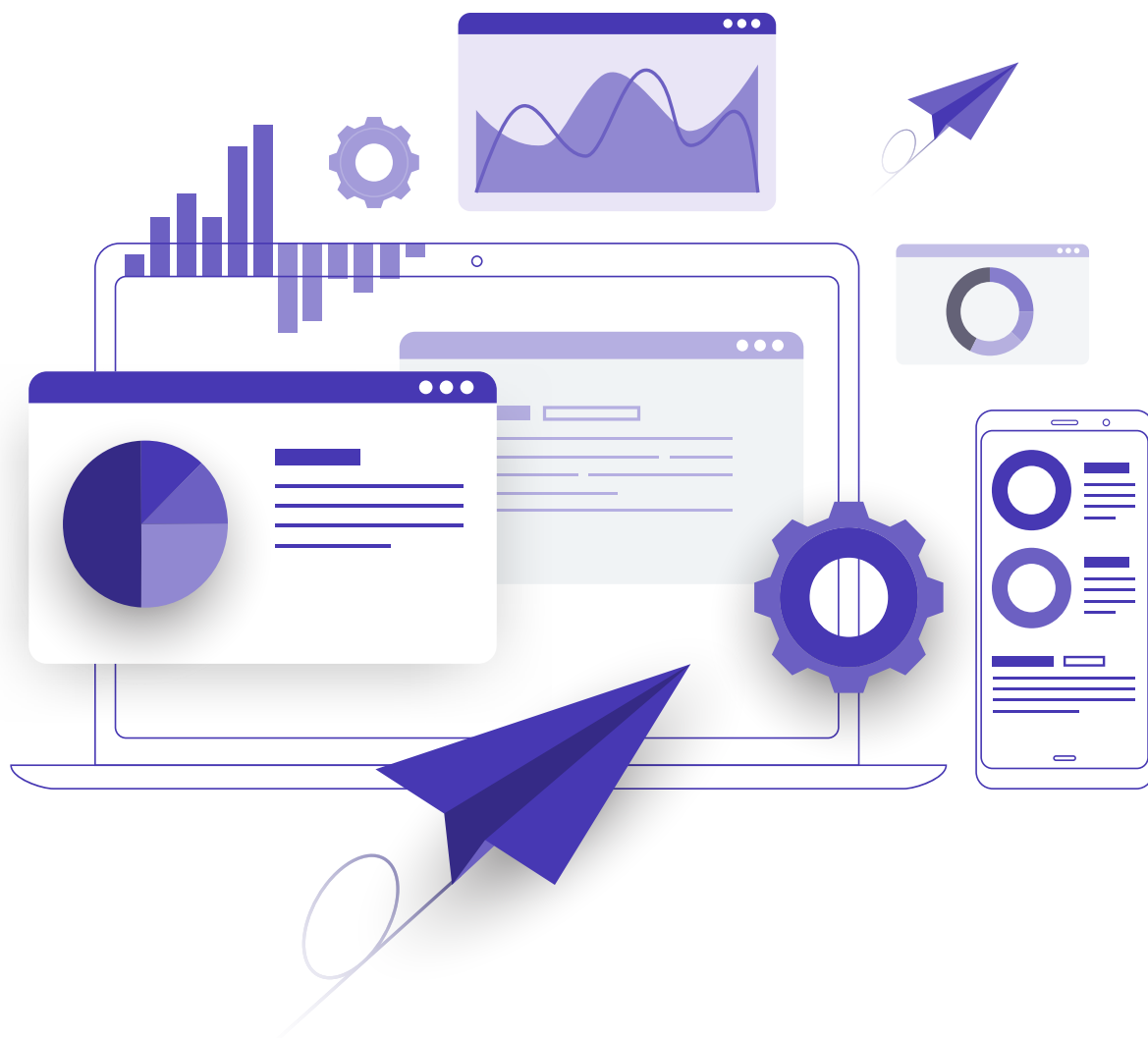




Harnessing web data for next-generation **skills intelligence**

BOOK OF ABSTRACTS



Disclaimer: The opinions and arguments expressed herein are those of the authors and do not necessarily reflect the official views of European Centre for the Development of Vocational Training and the European Commission.
The content has not been edited nor proofread by Cedefop's editing service.



© Cedefop, 2026

The reuse of this document is authorised under the [Creative Commons Attribution 4.0 International \(CC BY 4.0\) licence](https://creativecommons.org/licenses/by/4.0/). This means that reuse is allowed, provided appropriate credit is given and any changes made are indicated. For any use or reproduction of photos or other material that is not owned by Cedefop, permission must be sought directly from the copyright holders.

DOI [10.5281/zenodo.20656597](https://doi.org/10.5281/zenodo.20656597)

Please cite this publication as:

Cedefop. (2026). *Harnessing web data for next-generation skills intelligence*
Book of abstracts. Cedefop. <https://doi.org/10.5281/zenodo.20656597>

Table of contents

Foreword.....	7
Acknowledgements.....	8
A Hybrid Semantic Normalization Framework for Mapping Multilingual Job Postings to European Taxonomies	9
A methodological focus on skills analysis: experimentations with lexicometry and Natural Language Processing	11
A Vacancy Data-Driven Institutional Ecosystem: Bridging the Education-Employment Gap through Real-Time Skills Intelligence	12
Agent-Based Knowledge Graph-Augmented Skills Extraction from Online Job Advertisements	14
AI-Complementary Skills: Real-Time Evidence from Italian Job Postings.....	16
Beyond Composite Indices: Using Online Job Advertisements to Reveal Gender-Blind Spots in Skills Intelligence.....	18
Bridging Skills Demand and Vocational Qualifications: Linking Online Job Postings with National Skills and Qualification Frameworks	20
Building a Common Language for the Twin Transition: Reference Frameworks for Digital and Green Skills.....	22
Beyond Green Jobs: How Climate Policies Reshape Skill Demand.....	24
Digital Skills Demand in Greece: Labour Market Analysis Using Online Job Vacancies	25
Disruption or Augmentation? The Changing Demand for AI Skills in the Age of Generative AI.....	27
From Web Data to Actionable Skills Intelligence: A Three-Stage Architecture for Europe's Advanced Digital Skills	28
Harnessing web data for next-generation skills intelligence in the UK	30
Imbalance and talent shortage in the Catalan labor market: survey insights and comparison with OJAs	32
Leveraging Machine Learning for Personalized Labor-Market-Relevant Training Pathway Recommendations in Serbia.....	34
Mapping Digitalization through Skill Demand: Firms, Sectoral, and Regional Labor Market Insights in Italy	36

Mind the AI gap: A Stage-Based Human Capital Framework for the Fourth Industrial Revolution	37
Occupation classification algorithms at the Luxembourgish Public Employment Agency (ADEM)	39
Skills Demand and Regional Productivity	42
Skills intelligence analysis in the Offshore Renewable Energy sector in Europe: A value-chain approach.....	44
Skills Intelligence Emilia-Romagna (SI-ER): Leveraging Web Data and AI for Next-Generation Regional Skills Intelligence	46
Skills Intelligence for informing HEIs curricula: UPAL Labour Market tools	48
Spatial Matching of Online Job Vacancies and Training Provision Data	50
The elephant never left the room: did gender discrimination went implicit in French job posting?	51
The Hidden Innovators: Uncovering Capabilities and Industrial Dynamics in European Rural Areas	53
The ROI of Guidance: Measuring the Impact of Human-in-the-Loop Curation on STEM Talent Mobility	55
The role of education and training offer data in skills intelligence	57
Tracking Upskilling and Reskilling via Skill Evolution in GitHub Profile READMEs	58
Using Online Job Adverts Data on Skills to Identify Potential Career Pathways	60

Foreword

High-quality, timely, and user-centred skills intelligence is a vital component of skills development and matching policies. It guides policy makers, employers, workers, learners, and other skills ecosystem stakeholders, facilitates skills matching, and up- and re-skilling. Cedefop's skills forecast, skills surveys, the big data-powered SkillsOVATE platform, and other skills intelligence tools, are essential assets in understanding changing skills requirements on the European labour markets. They shed light on labour market trends in sectors and occupations and give insight into how skills upgrading and new emerging skills transform jobs. Cedefop's skills intelligence also identifies the skills shortages and bottlenecks standing in the way of just green and digital transitions, competitiveness, innovation, and growth.

The rapid progress in Big Data analysis and Artificial Intelligence in the last decade offers opportunities to generate more timely and granular skills intelligence. Web data is increasingly used for skills intelligence and in labour market research. Cedefop and Eurostat have played a leading role in building capacity to use online job advertising data for skills intelligence purposes. Both organisations have also stimulated debates among policymakers, researchers, practitioners and other stakeholders.

The 2026 research conference [*Harnessing web data for next-generation skills intelligence*](#), co-organised by Cedefop and Eurostat, created space for researchers to present their recent work powered by web data. More than 35 researchers presented their cutting-edge work in front of a highly engaged audience. The attendees greatly contributed to the success of the conference through their active participation and stimulating questions.

This publication presents (in alphabetical order) research abstracts that were presented in the conference. Its aim is to present in short format the key methods and findings of state-of-the-art, web-based driven skills intelligence. We encourage readers to use this book of abstracts as a resource for gaining insight into the current state of the art in skills intelligence and for engaging with the community of experts to further explore the field. We trust the overview of approaches and use cases in the field can serve as a basis for further future developments and contribute to expanding and further developing Cedefop's next generation skills intelligence.

Jürgen Siebel
Executive Director

Antonio Ranieri
Head of Department

Acknowledgements

This publication was produced by the Department for VET and Skills of the European Centre for the Development of Vocational Training (Cedefop), under the supervision of [Antonio Ranieri](#).

The publication originates from the research conference [Harnessing web data for next-generation skills intelligence](#), which was held on 28 and 29 May 2026 in Thessaloniki. Cedefop would like to express its gratitude to Eurostat for co-organising the event, to Michail Skaliotis for his insightful Keynote speech and to all the authors who presented their novel work at the conference and those whose abstracts are included in this publication.

We wish to acknowledge the invaluable contributions of the Scientific Committee members who reviewed abstracts submitted to the conference. We are also grateful to the session chairs for their dedication and professionalism in moderating the conference sessions and fostering scholarly exchange. We would like to also thank the technical support team for their invaluable assistance in the preparation and delivery of the conference and the proceedings volume.

A Hybrid Semantic Normalization Framework for Mapping Multilingual Job Postings to European Taxonomies

Panagiotis Zervas * [0000-0003-2710-1066](#)

Giannis Tzimas * [0000-0002-4073-7256](#)

Konstantinos Giotopoulos ** [0000-0001-5989-6313](#)

Łukasz Sienkiewicz *** [0000-0002-0653-8755](#)

Flavia Pesce **** [0009-0004-6829-169X](#)

* Department of Electrical and Electronics Engineering, University of Peloponnese, Patras, Greece

** Department of Management Science and Technology, University of Patras, Patras, Greece

*** Department of Human Capital, Faculty of Management and Economics, Gdańsk University of Technology, Gdansk, Poland

**** Training, Labour, Community and Local Development Area, Institute for Social Research, Bologna, Italy

Online Job Advertisements (OJA) have exploded in the entire European labor markets resulting significant opportunities and challenges for the development of skills intelligence systems. OJA data provide valuable real-time insights into policy and employer demand, however, different languages, industries and national contexts pose significant obstacles to proper comparative analysis. This contribution presents a novel hybrid normalization framework mapping job posting metadata to standardized European taxonomies, ESCO (European Skills, Competences, Qualifications and Occupations), NACE (Statistical Classification of Economic Activities), and ISCED/EQF (International Standard Classification of Education / European Qualifications Framework).

The normalization methodology employs a three-tier semantic matching architecture combining vector similarity search over HNSW-indexed embeddings, domain-specific post-processing filters, and LLM fallback classification. Each normalization task addresses a distinct challenge. For skills mapping, the system applies curated false positive blocking patterns and label-length preference heuristics to counteract systematic errors in embedding-based matching, where semantically unrelated terms (e.g., software tools and language proficiencies) occupy proximate regions in the embedding space. A tiered confidence routing

mechanism directs uncertain matches to LLM arbitration while rejecting low-confidence candidates outright. For occupation classification, the system addresses the known degradation of embedding quality on short text sequences by expanding abbreviated job titles into descriptive text through LLM-based query expansion prior to vector search. For industry classification, the system employs context-aware LLM prompting that leverages job description, extracted skills, and company information to disambiguate generic sector terms against NACE taxonomy.

An innovative feature of the system is the dynamic lexicon generation functionality for new and emergent categories like job responsibilities and employee benefits. By leveraging density-based clustering, the system automatically detects semantic groups within unlabeled data, produces clear, human-readable cluster labels using an LLM, and saves cluster centroids to enable efficient matching. The proposed hybrid static-dynamic approach allows the system to be in line with standard taxonomies and adapt to an evolving labor market terminology.

The system has been evaluated incorporating a multi-layered assessment strategy using Greek OJA. The evaluation covered the entire processing pipeline, including translation to English, deduplication, metadata extraction and taxonomy normalization. To perform the evaluation a ground-truth dataset was used and an LLM-as-a-judge approach combining this with standard classification metrics. The results were promising, demonstrating high performance of the different system units and the entire system itself. The translation component achieved an average semantic accuracy and fluency above 4.2 on a five-point scale. Semantic deduplication reached 0.96 F1-score, metadata extraction and normalization components demonstrated high precision and recall in the key fields of occupation, skills, industry classification and education level. This confirmed the strength and analytical reliability for a policy-oriented labour market intelligence application.

This work is part of the EU-funded EU-ALMPO project (Grant Agreement No. 101178736), which aims to establish a European Observatory for Active Labour Market Policies supported by AI-based labour market intelligence tools. It demonstrates a scalable, production-ready approach to harmonising heterogeneous OJA data for cross-national skills intelligence, supporting skills gap analysis, regional labour market monitoring, and education-employment alignment in line with the European Commission's priorities for evidence-based policy design.

A methodological focus on skills analysis: experimentations with lexicometry and Natural Language Processing

Eline Delmarcelle * [0000-0001-6915-1024](#)

Elise Ugeux *

* Bruxelles Formation, Belgium

As the research department of Bruxelles Formation, the public provider of vocational training in Brussels and coordinator for French-speaking vocational training in the region, the Research and Statistics Department (SES) aims to provide reliable data about the Brussels job market. Our service is therefore tasked to provide insights on skills intelligence in fast-changing sectors, such as the ICT sector.

Based on a previous study about this sector, this contribution aims to compare the advantages and disadvantages of the lexicometric approach and evaluate what shortcomings could be solved through the introduction of machine learning. Our first large-scale analysis of OJA (online job advertisements) in the ICT sector in Belgium, collected via web scraping, resulted in a dataset of 5,000 OJAs. We explain in the first part of our contribution how lexicometric analysis enabled us to draw conclusions about the language skills and education levels required in this sector, while confirming assumptions about contract types and the spatial concentration of employers around Brussels. But despite its many advantages, the lexicometric approach proved less effective for skills intelligence.

Our next project involves creating a pipeline which uses machine learning to sort, prepare, and analyze OJAs to provide insights into skills in a trilingual job market. The second part of our contribution is thus an open discussion on how machine learning and particularly Natural Language Processing can more effectively prepare and analyze data for skills intelligence, while maintaining standards of transparency and bias mitigation, in the perspective of explainable AI.

A Vacancy Data-Driven Institutional Ecosystem: Bridging the Education-Employment Gap through Real-Time Skills Intelligence

Carmen Pagés * [0000-0002-6164-8383](#)

Adrià Lòpez *

Rafael Sosa * [0009-0005-2818-0669](#)

* Universitat Oberta de Catalunya, Spain

As the digital and green transitions accelerate, educational institutions face the challenge of how to remain relevant in a rapidly shifting labour market. This contribution explores how the Universitat Oberta de Catalunya (UOC) is addressing this challenge. By leveraging Online Job Advertisements (OJAs), the university has built a comprehensive framework for skills intelligence and institutional alignment. As a pioneer in online higher education and lifelong learning, UOC's initiative goes beyond mere data collection. It is a holistic ecosystem designed to translate high-frequency web data into tangible institutional value and student success.

The engine of this initiative is a robust Labour Market Intelligence (LMI) infrastructure managed by a specialised Skills Intelligence Unit. The scale is significant. The unit has developed a custom pipeline to process and analyse a longitudinal dataset of over 10 million Spanish job vacancies from 2018 to 2024. The workflow begins with the reception of raw data already standardised according to international classifications (such as ESCO and Lightcast). A critical phase of the process is the subsequent investment in quality assurance, where the unit performs rigorous validation to ensure data reliability for research and academic purposes. Finally, this verified information is structured into a flexible star schema model within a central data warehouse. This architecture provides a high-performance foundation that guarantees scientific integrity and enables the unit to seamlessly adapt its findings to diverse institutional reports and apps. Through this granular analysis, the university can pinpoint emerging demands for digital, green, and transversal competencies. This evidence-based approach changes the internal conversation, allowing faculty and leaders to move past anecdotal observations.

But data is only as useful as its application. To bridge the gap between analysis and action, UOC deployed targeted digital interfaces for different stakeholders. Students now have access to a virtual orientation platform that automates skill-gap analysis and maps out personalised reskilling and upskilling pathways. Meanwhile, faculty use interactive dashboards to ensure that learning outcomes stay synchronised with shifting employer requirements. The impact also extends to the public, offering valuable information on structural trends such as telework and regional skill shortages through a regional labour market observatory.

Can this model scale across borders? We are currently testing its portability through the Erasmus+ OpenEU initiative. Working with partners like Universidade Aberta in Portugal and FernUniversität in Germany, UOC is building a pan-European framework for data-driven university management. Our presentation will dive into the technical hurdles of processing OJA data for HE skills alignment and the methodologies required for the cross-mapping of web-based taxonomies with formal qualification frameworks. We will also share insights into the strategic change management and governance shifts necessary to integrate these tools into traditional academic workflows. Ultimately, the UOC experience may serve as a useful reference for higher education institutions aiming to bridge the gap between academic provision and the shifting economy through proactive, data-driven alignment.

Agent-Based Knowledge Graph-Augmented Skills Extraction from Online Job Advertisements

Mikołaj Tym * [0009-0006-9866-6266](https://orcid.org/0009-0006-9866-6266)

* Poznań University of Economics and Business, Poland

Reliable, large-scale extraction of skills from Online Job Advertisements would significantly improve labour market analysis. Analyzing trends, seasonality, reacting to changes before negative effects occur, and adapting regulatory and educational policies, all of this requires extraction methods that are reliable, and capable of transforming unstructured text into structured data compliant with a standardized ontology, i.e., European Skills, Competences, Qualifications and Occupations (ESCO).

Traditional Natural Language Processing methods fail to achieve such reliable extraction. Approaches relying on keyword matching are unreliable because skills in natural language are often expressed differently than in the standardized ontology required for statistical aggregation. Due to the rich morphology, many skills remain undetected because of word inflection. Additionally, many skills are expressed as multi-word sequences rather than single words. Finding the exact ontology sequence in a job ad is rare, making keyword detection impossible. Moreover, job ads often contain company descriptions. Keyword matching frequently misclassifies a company's business activities as required job skills.

Attempts to bridge the gap between rigid ontology labels and real-world language often rely on detecting frequent n-grams. This strategy identifies sequences of words that appear repeatedly in ads and maps them to specific skills, theoretically capturing common synonyms not present in the ontology. However, this approach fails to capture less common requirements or whenever the wording deviates even slightly from the most common phrasings.

Large Language Models with Decoder architecture (like GPT) also have limitations. While they can extract skills, they do not output labels strictly according to the ESCO standard. Even when explicitly prompted to return ESCO labels, these models may tend to classify skills into the most common skill groups encountered during pre-training or even hallucinate non-existent skills outside the official taxonomy. A robust solution must cover the entire spectrum of the ontology, not just the most statistically probable terms. More importantly, using these models to

process hundreds of millions of ads to generate reliable statistics would be computationally infeasible.

To solve these problems, this work proposes a new approach for robust skill extraction. The goal is to automatically extract required skills and map them exactly to ESCO Knowledge Graph. This methodology entails the deployment of autonomous agent, powered by Large Language Models, that utilizes the ESCO Knowledge Graph to reason about skill connections. This agent employs Retrieval-Augmented Generation (RAG) and specific graph traversal tools, so it can compare a candidate skill from an ad with its potential parent groups, siblings, or descriptions within the ESCO graph to verify the semantic fit. This allows the system to distinguish between ambiguous terms based on context and graph structure.

This new approach aims to support the analysis of skills demand by occupation and provide labour market insights, including skill trends across economic sectors.

AI-Complementary Skills: Real-Time Evidence from Italian Job Postings

Greta Sofia Lampis * [0000-0002-0643-9845](https://orcid.org/0000-0002-0643-9845)

* Fondazione Bruno Kessler - Institute for the Evaluation of Public Policies, Italy

This paper investigates how artificial intelligence (AI) is reshaping skills demand in the Italian labour market using online job advertising (OJA) data from Lightcast for 2018-2024. We construct a vacancy-level dataset combining online postings with ESCO-based skill extraction and a dedicated AI-skill taxonomy. The data are geocoded at NUTS 3 level, enabling mapping by territory, occupation and sector and timely monitoring of emerging requirements, including generative AI. Beyond tracking single skills, we use OJA to identify skill bundles - sets of competences that systematically co-occur with AI requirements - thus translating real-time labour market intelligence into actionable insights for firms, policy makers, and education systems.

The study aims to:

- (a) investigate the diffusion of AI-related requirements across Italian occupations and regions and identify AI-complementary skills;
- (b) move beyond descriptive patterns by estimating how broader AI shocks translate into local changes in skill demand.

To establish whether AI adoption causes changes in local skill demand we adopt a counterfactual evaluation framework and implement an instrumental variables (IV) design at the occupation-location-year level. Local AI intensity is measured as the share of vacancies requiring at least one AI skill. Identification exploits national occupation-by-year AI shocks, common to all locations, interacted with predetermined local occupational structures (base-period shares). This generates quasi-experimental variation: areas more exposed ex ante to occupations experiencing stronger national AI uptake receive larger "AI shocks." We estimate reduced-form and IV effects on the demand for specific skills, controlling for occupation and time fixed effects.

Descriptives show that AI-related postings represent a small share of total vacancies, but exhibit heterogeneity across time, geography, and occupations, motivating sectoral and place-based analysis. A stable technical backbone dominates (e.g., machine learning and data tools), while generative-AI terms become visible after 2022.

Preliminary results suggest that higher AI exposure is associated with stronger growth in transversal skills linked to coordination and organisational adaptation: most notably “work in teams” and “adapt to change”, alongside baseline digital skills (e.g., office software and spreadsheet use). Co-occurrence networks reinforce this, revealing a persistent organisational “core” that connects multiple AI skills to teamwork, baseline digital tools, problem-solving and adaptability. Incidence measures in AI postings relative to overall incidence show that AI demand is disproportionately associated with software/data production competences. Changes between 2018 and 2024 suggest that scripting and query skills as well as interface-oriented competences (front-end design, graphics) become increasingly distinctive of AI vacancies. This suggests a two-layer interpretation: AI is increasingly embedded in organisational practices, yet it relies on a specialised set of tools that differentiates AI postings.

The paper contributes to the debates on vocational training and education, providing a replicable approach to identify which complementary competences expand as AI demand rises and where these shifts are strongest. The outputs can inform the design of professional training content and help assess whether training priorities should be targeted by occupation/sector or complemented with place-based initiatives. Italy serves as the pilot case and the approach will be extended to Germany and other EU countries as the AI-skill classification becomes available.

Beyond Composite Indices: Using Online Job Advertisements to Reveal Gender-Blind Spots in Skills Intelligence

Irina Dimitrova *

* National VET Team Bulgaria, Human Resources Development Center; ITPIO, Bulgaria

Composite indicators such as the Women in Digital (WiD) Index and the Gender Equality Index provide valuable, comparable insights into women's participation in digital education, employment, and leadership across Europe. However, their strength at macro level also constitutes a limitation: these indicators offer limited visibility into occupation-specific skill demand, transition points between education and employment, and the concrete skill signals that shape labour market entry and progression. This creates a methodological blind spot for skills intelligence and policy design, particularly in countries that perform well on aggregate indicators but still face persistent skills mismatches and attrition along ICT pathways.

This paper explores how online job advertisement (OJA) data can complement existing gender and skills indices by providing granular, real-time evidence on skills demand in ICT-related occupations. Using OJA data as the primary empirical source, the study analyses the structure and evolution of skill requirements across selected ICT roles, with a focus on entry-level and early-career positions. The analysis is anchored in a country case study (Bulgaria), where WiD data indicate strong performance in the ICT education phase but weaker outcomes in progression and leadership, raising questions about how education-stage advantages translate into labour market outcomes.

The study applies natural language processing techniques to extract and classify skill requirements from OJAs, grouping them into technical, transversal, and hybrid skill clusters. These clusters are analysed by occupation, seniority level, and sector, and then compared with education-stage indicators from the Women in Digital Index and national ICT training profiles. This mixed approach allows the identification of patterns that remain invisible in composite indices alone, such as skill inflation in entry-level roles, increasing emphasis on security, infrastructure, or hybrid business–technical skills, and divergence between formal education outputs and market-facing skill signals.

The findings demonstrate that while macro indicators capture participation and attainment, they do not reflect the complexity of skills demand shaping real labour market transitions. OJA-based analysis reveals where skill expectations may act as hidden barriers to entry or progression, and where training provision risks lagging behind evolving occupational profiles. The paper argues that integrating web-based labour market data into skills intelligence systems can improve the explanatory power of existing indices and support more targeted education and training responses.

From a policy perspective, the paper illustrates how OJA data can inform curriculum updates, modular training design, and early warning mechanisms for skills mismatches, complementing established monitoring tools such as the Women in Digital Index. By bridging macro-level indicators and micro-level skills signals, the proposed approach contributes to next-generation skills intelligence capable of supporting evidence-based education, training, and labour market policies.

Bridging Skills Demand and Vocational Qualifications: Linking Online Job Postings with National Skills and Qualification Frameworks

Mauro Pelucchi *

Sophie Gvasalia *

Anna Clara Gatti *

Ettore Colombo *

Luca Agostoni *

Simone Perego *

* Lightcast, Italy

The increasing availability of online job postings has significantly expanded the potential for timely and granular skills intelligence. However, translating web-based signals into actionable insights for vocational education and training (VET) systems remains a methodological challenge, particularly in contexts where qualifications are structured around nationally defined standards rather than occupations alone.

Vocational education systems are increasingly expected to respond rapidly to changing labour market needs driven by digitalisation, technological change, and sectoral transformation. While online job postings provide a rich and timely source of information on emerging skills demand, they are often analysed in isolation from institutional education and training frameworks. This limits their usefulness for policy design, particularly in countries where vocational provision is organised around qualifications and learning outcomes rather than job titles. The Italian Atlante del Lavoro offers a unique opportunity to bridge this gap. Acting as an intermediate semantic layer (linking occupations, sectors, areas of activity, skills, and qualifications) it enables a structured integration of web-based labour market signals with formal VET standards. Applying this framework to leFP, ITS and IFTS pathways makes it possible to conduct an evidence-based assessment of whether vocational qualifications equip learners with the skills that are demonstrably in demand.

This study presents an integrated framework that links online job postings data with the Italian Atlante del Lavoro (INAPP, 2024) and the National Repository of Qualifications to analyse skills demand and supply within Education and Vocational Training (leFP), Higher Technical Education (ITS) and Higher Technical Education

and Training (IFTTS). Using a multi-layer mapping approach, job postings are connected to Areas of Activity (ADA) through occupations (CP2021) and sectors (ATECO, compatible with NACE). Each ADA is then linked to formally defined vocational qualifications, enabling a direct comparison between skills demanded by employers and skills embedded in qualification standards.

By extracting and harmonising skills from both job postings and qualification descriptions, the framework enables a systematic skills gap analysis at the level of vocational pathways rather than individual occupations. Building on prior methodological work that pioneered the measurement of alignment between online job postings and regional training catalogues (OECD, 2023), we introduce the VET Alignment Index (VAI), a composite indicator that quantifies the degree of correspondence between the skills demanded by employers in online job postings and the skills embedded in the educational content of leFP, ITS and IFTS qualification standards. This approach provides new evidence on the alignment, and misalignment, between labour market demand and vocational curricula, offering concrete insights for curriculum design, qualification updates, and policy monitoring. The paper contributes to the next generation of skills intelligence by demonstrating how web data can be structurally embedded within national qualification systems to support evidence-based VET policy.

The paper pursues four main objectives:

- (a) To demonstrate how online job postings data can be systematically linked to national qualification systems using occupations, sectors, and Areas of Activity as semantic connectors.
- (b) To quantify and characterise skills gaps between employer demand (as observed in online job postings) and the skills embedded in the educational content of leFP, ITS and IFTS qualification standards, through the proposed VET Alignment Index (VAI).
- (c) To assess the extent to which vocational training pathways reflect current and emerging labour market needs, moving beyond occupation-centric analyses.
- (d) To provide a replicable analytical framework that can be adapted to other countries with structured qualification repositories, supporting comparative and forward-looking skills intelligence across Europe.

References

- INAPP. (2024). *Repertorio nazionale qualificazioni*. Atlante Lavoro. <https://www.inapp.gov.it/atlantelavoro/repertorio-nazionale-qualificazioni>
- OECD. (2023). *Big data intelligence on skills demand and training in Umbria*. OECD Publishing. <https://doi.org/10.1787/4b5bbfd6-en>

Building a Common Language for the Twin Transition: Reference Frameworks for Digital and Green Skills

Diana Ripkova Puhovichova *

Mario Lelovsky * [0009-0003-7163-9602](tel:0009-0003-7163-9602)

Martin Martinkovic *

* Digital Coalition, Slovakia

The twin transformation—digital and green—is fundamentally reshaping the European economy, labour markets, and business models. While the debate often focuses on technologies, regulation, and investment, it is increasingly evident that the success of this transformation ultimately depends on the quality and preparedness of the workforce. Changing skills requirements, new forms of work, and rising expectations placed on both employees and employers highlight the need for a systematic, transparent, and comparable approach to the identification, assessment, and development of digital and green skills.

Despite the growing availability of labour market information, including data from online job advertisements (OJAs), Europe still lacks a comprehensive and practically applicable framework that translates changing demand for digital and green skills into measurable and certifiable levels. Existing approaches are often fragmented, descriptive, or occupation-based, with limited linkage to the assessment and development of individual skills.

This paper presents an innovative approach that has been systematically implemented for the first time in the Slovak Republic: reference frameworks for digital and green skills, inspired by the methodology of language frameworks such as the Cambridge framework. Based on the logic of common reference levels, these frameworks establish a shared language for employees, employers, education providers, and policymakers. They enable transparent comparison of skill levels based on a unified competency model.

The reference frameworks were developed within the national project Digital Future, implemented in cooperation with the Ministry of Investments, Regional Development and Informatics of the Slovak Republic, and the Digital Coalition. The project responds to dynamic changes in the labour market and lays the foundation for a coordinated approach to skills development in the context of the digital and green transformation. The frameworks are directly linked to a testing and

certification system that enables skills verification, benchmarking, and validation. Pilot testing and certification are currently being carried out at the managerial level and will run until June 2027 through a web-based interface.

The paper outlines the conceptual design of the reference frameworks, their methodological foundations, and their alignment with labour market needs. It also highlights the potential of a systemic and standardised approach to the development of digital and green skills as a key instrument for addressing the challenges of the twin transformation and as an opportunity for application in a broader European context.

Beyond Green Jobs: How Climate Policies Reshape Skill Demand

Angelica Bertucci *

Emilio Colombo **

Patrizio Tirelli *

* University of Pavia, Italy

** Università Cattolica del sacro Cuore, Italy

This paper studies the effects of climate change policy (CCP) shocks on regional labour market dynamics in Italy, exploiting three independent measures of CCP shocks. We focus on the skill composition of online vacancy postings, distinguishing between postings that require green skills and those that do not. CCP shocks generate persistent contractionary effects on employment, but vacancy dynamics display a markedly different pattern. Both green and—albeit to a lesser extent—brown vacancies increase following the shocks. However, we do not find a statistically significant rise in postings associated with occupations typically classified as green. Instead, vacancies linked to brown occupations increasingly demand green skills. These findings indicate that firms' reorganisation in response to climate policies entails a complex reallocation of labour demand that goes well beyond a simple shift from brown to green jobs. Finally, we find no evidence to support a “twin” green–digital transition: the increase in demand for green skills is significantly stronger in non-digital occupations.

Digital Skills Demand in Greece: Labour Market Analysis Using Online Job Vacancies

Petros Stavropoulos * [0000-0003-1664-6554](#)

Maria Pateli * [0009-0005-4774-9814](#)

Haris Papageorgiou * [0000-0002-7352-2403](#)

* ATHENA Research and Innovation Center - Institute for Language and Speech Processing (ILSP), Greece

We present a web-data approach for producing labour market and skills intelligence from online job vacancy postings in Greece, using data collected from [kariera.gr](#) for the full calendar year 2022. The work was developed by Athena Research Center (ARC) in collaboration with the General Secretariat for Digital Governance and Process Simplification, with the objective of supporting reliable analysis of real labour market data and strengthening standardised evidence on digital skills requirements across occupations and regions.

A key challenge addressed is the limited standardisation in how vacancies describe job roles and required skills, which reduces comparability and weakens evidence for policy design in the context of the digital and green transitions. To respond to this, the project implements an end-to-end pipeline that transforms unstructured vacancy pages into a structured dataset aligned with ESCO and organised by Cedefop sectors. Data are harvested in HTML and structured into fields including job title, posting date, and location, enabling both occupational and skills-level analysis.

For 2022, the corpus comprises 50,335 vacancy postings. A targeted analytical subset includes 20,285 postings from six Cedefop sectors (Financial and insurance activities; Information and communication; Accommodation and food service activities; Wholesale and retail trade, including repair of motor vehicles and motorcycles; Manufacturing; Professional, scientific and technical activities). From 1,248 ESCO occupation titles represented in the dataset, 60 occupations (10 per sector) were selected for detailed skills analysis based on frequency, availability of skills-related content, and reduced ambiguity in sector assignment. The final analytical database includes 14,866 normalised postings and 77,083 ESCO-matched skill mentions after filtering.

Methodologically, Natural Language Processing is applied to normalize job titles and skill mentions to ESCO through pre-processing (e.g., noise removal and lemmatisation), similarity-based matching to ESCO labels and descriptions (Greek and English), ranked candidate selection, and post-filtering to exclude unreliable matches. At the skills level, additional filtering separates genuine skills from qualifications, recruitment constraints, and residual “noise” (e.g., benefits or application instructions). Digital skills are identified using ESCO’s digital skill set (1,201 ESCO skills flagged as digital), and the dataset is enriched with a hard/soft skill distinction and geographic coding to NUTS levels.

Results show that digital requirements are widespread: 69% of postings (10,312/14,866) include at least one ESCO digital skill, though intensity varies strongly by sector (Information and communication: 90% digital postings; Financial and insurance: 86%; Professional/scientific/technical: 81%; Wholesale/retail: 67%; Manufacturing: 60%; Accommodation/food services: 19%). Spatial distributions are highly concentrated: Attica accounts for 68.7% of all postings and 75.3% of digital postings, with the Central Athens area dominating at NUTS3 level.

Across sectors, the most frequent transversal skills include English, communication, teamwork, and Microsoft Office, while top digital skills reflect sector-specific needs (e.g., spreadsheets/ERP/SAP in finance; programming, JavaScript and SQL in ICT; automation technologies and AutoCAD in manufacturing). Overall, the workflow demonstrates how web vacancies can be converted into comparable, classification-aligned indicators to support continuous monitoring of occupational demand and digital skills needs in Greece.

Disruption or Augmentation? The Changing Demand for AI Skills in the Age of Generative AI

Loris Vergolini *

Chiara Binelli *

Teresa Luca *

Gabriele Marconi ** [0000-0002-2957-0870](https://orcid.org/0000-0002-2957-0870)

* University of Bologna, Italy

** Agence pour le Développement de l'Emploi, Luxembourg

We study how the introduction of generative AI (GPT-3) has impacted the demand for AI-related skills in the European labour market. Using a novel dataset on the universe of online job postings in twenty six European countries and developing a detailed classification of AI skills and tasks, we exploit the release of GPT-3 in November 2022 as a natural experiment and apply a difference-in-differences estimation to assess the shifts in the demand for AI skills for occupations that are exposed to ChatGPT relative to not-exposed occupations. We find that GPT-3 had a negative and statistically significant impact on the share of AI job adds for occupations in the treated group relative to the control group so that the availability of generative AI decreased the demand for occupations whose most frequent core task is automatable through ChatGPT. To the best of our knowledge, this is the first paper that proposes a detailed classification framework and robust identification strategy to study the impact of generative AI on labour demand.

From Web Data to Actionable Skills Intelligence: A Three-Stage Architecture for Europe's Advanced Digital Skills

Anu Passi-Rauste *

Brendan Rowan **

Harri Ketamo *

Ivana Pesic *

Julie Byrne ***

Na Fu ***

* HeadAI, Finland

** BluSpecs, Spain

*** Trinity College Dublin, the University of Dublin, Republic of Ireland

Rapid technological change is reshaping skill requirements across labour markets, education and training systems, and innovation ecosystems. Yet skills intelligence in Europe remains fragmented, static, and slow to capture emerging signals. At the same time, large volumes of web data—such as job advertisements, education and training descriptions, research outputs, and investment records—offer timely and granular evidence of changing skill needs. In practice, however, the heterogeneity of these data sources and the lack of semantic alignment limit their direct usefulness for policy and planning. This paper presents an integrated, AI-driven approach that translates large-scale web data into actionable skills intelligence through a three-stage project architecture, developed within the LEADSx2030 – Leading European Advanced Digital Skills to 2030 initiative.

Stage 1 focuses on data modelling, transforming raw web data into structured and comparable skills intelligence. The Advanced Digital Skills (ADS) Data Model enables aggregating and harmonising large-scale unstructured data into structure enabling statistical analysis and meta-analyses between other skills demand activities. The raw data include over 3 million EU-wide online job advertisements (2024), of which more than 210,000 explicitly reference advanced digital skills, alongside over 10 million open-access scientific publications and approximately 100,000 European investment and innovation records. Using focused language model and graph machine learning -based natural language processing, the model organises skills into dynamically evolving skills pockets—semantically clustered groups of related skills that reflect how competencies co-occur and evolve in real

labour-market and innovation contexts. This approach moves beyond static taxonomies, enabling cross-sector and cross-country comparison, longitudinal tracking of skills trends (2021–2026), and early identification of emerging and declining skills.

Stage 2 operationalises the model through a federated Data Hub, providing the infrastructure required to sustain and scale skills intelligence. Aligned with European data-space principles and Gaia-X governance, the Data Hub enables policy bodies, education providers, industry, and research organisations to contribute and access non-personal skills data under transparent governance and a “give-to-get” mechanism. Continuous web-based data streams are combined with partner-contributed datasets, including education and training metadata from the SO4 Advanced Digital Skills Cluster, ensuring data sovereignty, quality assurance, regulatory compliance, and continuous enrichment of the skills intelligence ecosystem. The Data Hub ensures access to up-to-date big data sets.

Stage 3 translates analytics into decision-support tools through a public, interactive skills intelligence dashboard. The dashboard exposes insights derived from millions of job advertisements and related data sources, visualising skills demand trends, skills-pocket evolution, and technology-specific patterns across sectors, regions, and time horizons. Forecasting functions combine short-term labour-market signals with mid- and long-term signals from investment and research data, supporting anticipatory skills planning to 2027 and 2030 rather than reactive responses. This enables policy makers, education planners, and labour-market institutions to align education and training provision, funding priorities, and workforce strategies with evidence-based projections.

Together, these stages demonstrate how large-scale web data, AI-driven modelling, European data infrastructures, and user-centred visualisation can be integrated into a scalable system for next-generation skills intelligence. The approach provides a practical blueprint for evidence-based skills anticipation, education and training alignment, and policy design in Europe’s rapidly evolving digital economy.

Funding acknowledgement:

This research is funded by the European Union’s Digital Europe Programme under the LEADSx2030 – Leading European Advanced Digital Skills to 2030 project.

Harnessing web data for next-generation skills intelligence in the UK

Jeisson Cardenas-Rubio *

Rosie Day **

Neil Bachelor **

Peter Elias **

Andy Dickerson ***

* Cambridge Econometrics, United Kingdom

** University of Warwick, United Kingdom

*** Sheffield University, United Kingdom

Online job postings offer timely and detailed insights into employer demand for skills. Transforming these unstructured texts into actionable intelligence requires a national taxonomy that is coherent, granular, and linked to occupational and qualification standards. This paper presents our experience developing the UK Standard Skills Classification (SSC) and testing its robustness using a large corpus of UK vacancies. Led by the Warwick Institute for Employment Research with support from Skills England, this work combines methodological innovation with practical implementation in a real-world data environment.

We describe the end-to-end design of the SSC, a four-level taxonomy comprising 3,343 occupational skills nested within 606 skill groups, 106 skill areas, and 22 domains. The SSC is linked to the Standard Occupational Classification (SOC) 2020 at the six-digit level and includes conceptual mappings between skills, tasks, knowledge, and core competencies. Its development integrated multiple sources, including ONET, ESCO, and UK vacancy texts, combining AI-based semantic clustering with expert review to ensure conceptual soundness, empirical grounding, and usability for policy and practice. We detail design choices, quality assurance procedures, and trade-offs made to balance specificity with generality.

We empirically tested the SSC using 869,646 UK job postings from the year ending April 2025. Vacancies were standardised and coded to SOC2020 Sub Unit Groups using CASCOT, filtering out low-certainty postings. Sentence-level embeddings were used to extract skills from vacancy texts and compared against SSC-defined skills. For each occupation, we computed an Index of Coincidence to measure the share of SSC skills observed in vacancies. Reliability thresholds were applied, and sensitivity to data sparsity examined.

Results show strong alignment between SSC profiles and employer-described skills for well-represented occupations. Lower alignment occurs mainly in niche roles with few postings or where employer language diverges from taxonomy terminology. At the domain level, high demand is observed in Advising and Supporting, Managing and Directing, Evaluating and Inspecting, Diagnosing and Treating in healthcare, and Educating and Training. We also identify rarely observed SSC skills, often procedural or context-specific, highlighting priorities for refinement.

A second analysis removed occupational restrictions to map all skills mentioned in vacancies, revealing cross-occupational patterns such as healthcare collaboration, client relationship management, risk and quality management, and technical skills in programming and digital deployment. Comparisons with ESCO-based rule extraction show that the SSC combined with semantic methods captures specific, context-rich skills, while rule-based approaches yield more generic labels, limiting policy and curriculum relevance.

This paper provides a practical blueprint for building and applying national skills taxonomies using web data. Iterative testing with vacancy texts ensures the SSC remains aligned with evolving employer language and supports better workforce planning, training provision, and occupational standards.

Imbalance and talent shortage in the Catalan labor market: survey insights and comparison with OJAs

Adrià López Carrasco *

Carmen Pagés Serra * 0000-0002-6164-8383

Paula Pedro Aznar * 0009-0003-5225-8337

* Universitat Oberta de Catalunya, Spain

SMEs in Catalonia represent 99.8% of companies in Catalonia and employ around 70% of the workforce (Pimec, 2024), highlighting their crucial role in the Catalan economy. Despite their importance, relatively little is known about the talent needs and difficulties faced by SMEs. In this study, a representative survey of SMEs in Catalonia was conducted by Pimec as part of the [Baròmetre](#) collaboration project between the Open University Of Catalonia (UOC) and Pimec. The results underlie the difficulties faced by SMEs in the region in recruiting talent for their companies.

Specifically, more than 70% of the companies surveyed that opened a selection process took more than three months to fill the vacancy, and nearly half of the open vacancies (49.3%) were considered difficult to fill by companies, meaning that nearly half of the companies that opened a selection process (48.6%) reported recruitment problems. The main reason cited for limiting the filling of vacancies is the lack of candidates, which was cited as a factor in 77.4% of hard-to-fill vacancies. Other reasons cited were a lack of technical knowledge (in 33.9% of hard-to-fill vacancies), a lack of transferable skills (in 27.7%), and a lack of the required qualifications or level of training (in 15.6%). This suggests a mismatch between labor supply and demand in the Catalan labor market, primarily due to the general shortage of personnel, and secondly, to the difficulty in finding personnel who meet the skill and/or knowledge requirements necessary to access the jobs on offer. This contrasts with the high unemployment rates in Catalonia and Spain at the time of the survey (close to 10%) compared to Europe as a whole (INE, 2024; Eurostat, 2024), and reinforcing the existence of a structural mismatch.

This study also compares SMEs' talent demands with data from Online Jobs Advertising (OJAs), revealing some discrepancies in terms of skills and occupations sought. Capturing SMEs' talent needs can be challenging because they often use recruitment channels other than online job portals. As a result, SMEs' talent needs and demands may be partially excluded from vacancy

analysis. Therefore, while OJAs analysis is a necessary and very powerful tool for monitoring labor market trends, they should be complemented with surveys, interviews, or other methods focused on SMEs to better capture their specific needs—particularly in contexts such as Catalonia and Spain, where SMEs dominate the economy.

Acknowledgements. The work was conducted in collaboration with Silvia Miró and Hèctor Pérez from Pimec. The authors would like to express their gratitude to the Pimec's technical team, responsible for carrying out the survey and to Miriam Durán Martínez for her contributions to this study.

Leveraging Machine Learning for Personalized Labor-Market-Relevant Training Pathway Recommendations in Serbia

Milica Jelic Mariokov *

* The Qualifications Agency of the Republic of Serbia, Serbia

Online job advertisements provide timely, relevant and dynamic information on skill demands in a constantly changing labor market. At the same time, information on officially recognized training programs, which can be aligned with qualification standards, is available through national registers of accredited training providers. Despite the availability of both data sources, the connection between labor market demand and formally recognized training offer remains fragmented in Serbia.

This study explores how machine learning techniques can be employed to link online job advertisements with accredited training programs in order to generate personalized training pathway recommendations in Serbia. The main objectives of this research are:

- (a) to identify and structure qualification, skills and competencies requirements embedded in online job advertisements;
- (b) to map these requirements to official qualification standards and learning outcomes associated with accredited training programs listed in the national register;
- (c) to develop a machine-learning-based recommendation framework that proposes individualized training pathways leading towards relevant competencies and targeted occupations.

The overarching goal is to enhance the responsiveness of formal training systems to labor market needs, strengthening the link between employment demand and the existing lifelong learning offer.

The methodology proposes integration of natural language processing (NLP), ontology-based skill modeling, semantic representation learning and recommender system design. Job advertisement texts could be analyzed using NLP techniques such as tokenization, skill and occupation extraction, and contextual embeddings to capture semantic meaning. Accredited training program data, retrieved from the national register of accredited training providers, include standardized information on qualification levels, learning outcomes, occupations,

prerequisites and awarded certifications. These structured datasets could enable the construction of a shared skill ontology that aligns labor market terminology with formal qualification-specific language. Machine learning models, including similarity-based matching and hybrid recommendation algorithms, can be used to compare job-derived skill requirements with training-related learning outcomes, identify competency gaps and generate personalized training pathway recommendations.

The data sources for this study include online job advertisement datasets, such as those collected from Serbian public or private employment platforms, data obtained from the National Qualifications Register of Serbia, i.e. a Sub-register of Publicly Recognized Organizers of Adult Education Activities (Sub-register of PROA), which contains information on accredited organizations and training programs, and anonymized learner profile data describing prior education, work experience and certifications.

Key findings indicate that machine learning can enable mapping of skill gaps, highlight mismatches between emerging labor market demands and existing accredited training offerings, as well as facilitate training pathway recommendations in Serbia.

Mapping Digitalization through Skill Demand: Firms, Sectoral, and Regional Labor Market Insights in Italy

Sladana Krgovic *

Giulia Canzian *

Erica Delugas *

* CSIL, Italy

We map digitalization in Italy at the firm level, and explore its spatial and sectoral patterns. We proxy firms' level of digitalization using their demand for digital skills based on the crucial role of skills in enabling firms' effective digital transformation and the assumption that changes in digitalization are reflected in changes in labor market skill demand.

Our analysis is based on a comprehensive and highly-granular vacancy-level dataset from Lightcast covering the 2017-2024 period. To assess the intensity and distribution of digital labor demand, we first construct a Skill-Based Digitalization Index (SBDI) based on vacancy-level digital scores, defined as the weighted average of digital skills required in each vacancy, with weights based on the skill level. We also develop a Digital Divide Index (DDI), drawing on Generalized Entropy Indices, to analyze the dispersion in digital scores across vacancies. The relative importance of (i) changes in vacancies composition and (ii) changes in their digital scores in explaining the evolution of SBDI over time is evaluated by applying a shift-share decomposition.

Our findings indicate that companies with the highest SBDI predominantly belong to the Information and Professional Services sectors, which have consistently been top-ranked. Regionally, Lazio and Lombardy have consistently held the highest rankings, although Lombardy experienced a decline in 2023 relative to prior years. Changes in the SBDI over time have been predominantly driven by shifts in vacancy composition, characterized by rising demand for basic and intermediate digital jobs and declining demand for advanced digital jobs. When focusing on vacancy inequality, measured by the DDI, we observe that regions and sectors with the highest SBDI values exhibit the lowest vacancy inequality, indicating that their vacancies are more uniformly digital. At the national level, inequality is primarily driven by variation within regions or sectors rather than between them.

Mind the AI gap: A Stage-Based Human Capital Framework for the Fourth Industrial Revolution

Mikel Albizu * [0000-0003-1679-5007](#)

Alba Rodríguez-Agüero ** [0009-0000-0608-7264](#)

Alejandro Sainz-Aja ** [0000-0002-3233-0640](#)

* Orkestra – Universidad de Deusto (DBS), Spain

** University of Alcalá – Possible Lab, Spain

This paper explores the evolving role of artificial intelligence (AI) in the Fourth Industrial Revolution, with a particular focus on the human capital necessary for its successful integration into production systems and business models. Human capital plays a key role in the development and shaping of industrial revolutions. During the previous industrial revolutions, a workforce with advanced, emerging technology-related skills was essential in driving technological innovation and industrial deployment. These workers generated numerous incremental micro-inventions that, when integrated into production processes, transformed industries. Based on this the study develops a comprehensive framework to examine the dynamic interaction between emerging and specialized AI occupations and the stages of technological evolution across industrial revolutions (1. Ideation and Primary Design, 2. Hatching and Deployment, and 3. Maintenance and Sustainability).

The research methodology takes a multi-phase approach. First, an AI occupational taxonomy is created based on the European Commission's ESCO database. As there are no specific AI-related occupations listed in the ESCO database, a series of skills associated with these occupations were identified to establish an initial list of occupations. Second, this preliminary list was then refined by cross-referencing ESCO with occupational frameworks from the International Professional Organisation for the Evaluation of Competence (OPIIEC), as well as studies developed by the Basque Artificial Intelligence Centre (BAIC) and the Association to Promote the Use of Data and Artificial Intelligence in Companies and SMEs in the Spanish Industry (IndesIA). This ensured that the list was up to date. Third, qualitative research methods were also employed to validate the list, including interviews with companies and experts. This process resulted in the identification of 23 key AI-related occupations and the set of skills linked to them.

These were categorised into three groups: Generators, Enablers, and Generator/Enabler hybrids. These roles were analysed according to their relevance at different stages of technology development within industrial revolutions.

Key findings indicate that we are in the early stages of the Hatching and Deployment phase, with a growing emphasis on hybrid profiles capable of understanding business needs and implementing AI solutions tailored to those needs. The study also highlights the need for a flexible educational system that can understand and adapt to the evolutionary nature of AI technology integration. It concludes that fostering competitiveness requires not only the training of specialized professionals but also the embedding of AI user skills across the entire workforce, ensuring broad adoption and effective application of AI technologies.

This research advances our understanding of how human capital is integrated into the development and diffusion of AI, by providing a framework that explains how occupational needs evolve according to the stage at which the technology is deployed. Additionally, it makes a second significant contribution by developing an AI occupational/skills taxonomy that can be used to harness the potential of AI, a catalytic technology of the Fourth Industrial Revolution. The paper also offers practical recommendations to help educational institutions and policymakers align training systems with evolving needs related to the integration of AI.

Occupation classification algorithms at the Luxembourgish Public Employment Agency (ADEM)

Gabriele Marconi * [0000-0002-2957-0870](https://orcid.org/0000-0002-2957-0870)

* Agence pour le Développement de l'Emploi, Luxembourg

This paper discusses classifying approaches for identifying occupations in online job ads (OJAs). In Luxembourg, this task has at least two important practical implications. First, jobseekers can be sanctioned for refusing “acceptable” vacancies (Crépon & van den Berg, 2016; The Government of the Grand Duchy of Luxembourg, 2001), with “acceptability” partly determined by the match between the occupation code assigned to the jobseeker and the OJA. Second, recruitment from abroad follows a simpler procedure for vacancies tagged with an occupational code that falls on a “shortage” list (The Government of the Grand Duchy of Luxembourg, 2023). Therefore, ADEM is striving to develop occupation classification algorithms with a satisfactory level of accuracy and of explainability (i.e. the possibility for a human observer to “elaborate, control and possibly contest” the algorithm’s results). The goal is to predict the code among 1680 Codes d'Appellation in the ROME taxonomy.

Based on a scan of the literature, an occupation classifier can be preliminarily described based on the following questions:

- (a) Does it transform the source text data into text embedding (Gonzalez-Garcia, et al., 2026), or work on the tokens directly (Ikudo et al., 2019)?
- (b) Does it learn from annotated data (Savic et al., 2020), or does it only work based on occupation descriptors?
- (c) Does it classify job titles or job descriptions?

After testing two open-source alternatives, three classification methods were developed at ADEM:

- (a) The Centroids methods classify the OJA description after creating embeddings based on a TF-IDF-weighted word2vec model (Khan & Rekha, 2023). The word2vec model was trained on the WIH-OJA-NLP dataflow thanks to a collaboration with the responsible team at Eurostat.
- (b) The Duplicate method, inspired by previous research (Gweon et al., 2017), classifies job titles based on the past codes assigned to that job title (or the

closest one) in our database. Optimal string alignment and S-BERT are used to calculate proximities between job titles.

- (c) Following the successful application of Naïve Bayes classifiers in the WIH-OJA project, a Semi-naïve Bayes Model (Bielza & Larrañaga, 2014) was developed (Bayes Network). In this model, every occupation is associated to a set of words, and every word has two different coefficients: a weaker coefficient applicable if it is found in isolation; a stronger coefficient applicable if it is found together with another word associated to that occupation.

To evaluate these classifiers, our back-office agents rated 5 predictions (2+2+1 by the centroids methods, Bayes Network and duplicate method, respectively) for 7076 OJAs published by ADEM in May and June 2026, indicating if these were “bad”, “ok” or “very good”. Only in 5% of cases the best of the five predictions was rated as “bad”, which was considered satisfactory by our back-office agents. This led to the decision to put the models in production, for which work is planned. The Bayes Network was the most accurate algorithm (54% of “very good” predictions when applied to job titles), and also the most explainable (the words that determine the model’s decision can be identified with precision).

Future developments could aim at using LLMs from the latest (genAI) generation, which recently became accessible for applications within the Luxembourgish public sector. In a first application, a RAG-augmented “Mistral-medium” model achieved 42% of “very good” predictions on the dataset described above.

References

- Bielza Concha, & Larrañaga Pedro. (2014). Discrete Bayesian Network Classifiers: A Survey. *ACM Computing Surveys*, 47, 1–43. <https://doi.org/10.1145/2576868>
- Crépon, B., & van den Berg, G. J. (2016). Active Labor Market Policies. *Annual Review of Economics*, 8(1), 521–546. <https://doi.org/10.1146/annurev-economics-080614-115738>
- Gweon, H., Schonlau, M., Kaczmirek, L., Blohm, M., & Steiner, S. (2017). Three Methods for Occupation Coding Based on Statistical Learning. *Journal of Official Statistics*, 33, 101–122. <https://doi.org/10.1515/jos-2017-0006>
- Ikudo, A., Lane, J. I., Staudt, J., & Weinberg, B. A. (2019). Occupational classifications: A machine learning approach. *Journal of Economic & Social Measurement*, 44(2/3), 57–87. <https://doi.org/10.3233/JEM-190463>
- Khan, Z. A., & Rekha, V. (2023). Fake News Detection Using TF-IDF Weighted with Word2Vec: An Ensemble Approach. *International Journal of Intelligent*

- Systems and Applications in Engineering*, 11(3), 1065–1076. Retrieved from <https://ijisae.org/index.php/IJISAE/article/view/3366>
- Lino Gonzalez-Garcia, Miguel-Angel Sicilia, & Elena García-Barriocanal. (2026). Classification of Job Offers into Job Positions Using NET and BERT Language Models. *Computers, Materials & Continua*, 86, 1–15. <https://doi.org/10.32604/cmc.2025.070813>
- Savic, N., Bovio, N., Gilbert, F., & Canu, I. G. (2020). *Procode: the Swiss Multilingual Solution for Automatic Coding and Recoding of Occupations and Economic Activities*. <https://doi.org/10.48550/arXiv.2012.07521>
- The Government of the Grand Duchy of Luxembourg. (2001). *Le Plan d'Action national en faveur de l'Emploi 2001 (PAN)* - Dossier de presse. The Government of the Grand Duchy of Luxembourg. https://gouvernement.lu/fr/actualites/toutes_actualites/articles/2001/05/16biltgen/20010516biltgen_pan.html
- The Government of the Grand Duchy of Luxembourg. (2023). *Communiqué du ministère des Affaires étrangères et européennes et du ministère du Travail, de l'Emploi et de l'Economie sociale et solidaire*. The Government of the Grand Duchy of Luxembourg. Ministère des Affaires étrangères et européennes, de la Défense, de la Coopération et du Commerce extérieur; Ministère du Travail. <https://gouvernement.lu/dam-assets/documents/actualites/2023/09-septembre/01-penurie-main-oeuvre/20230901-communiqu-mteess-et-maee.pdf>
- The Government of the Grand Duchy of Luxembourg. (2025). Accelerating digital sovereignty 2030: Luxembourg's AI Strategy. The Government of the Grand Duchy of Luxembourg. <https://gouvernement.lu/fr/publications/rapport-etude-analyse/minist-digitalisation/2025-luxembourg-ai-strategy-en.html>

Skills Demand and Regional Productivity

Alexandra Badort * [0009-0004-2073-0374](#)

Bernardo Caldarola ** [0000-0002-4970-1583](#)

Ciarli Tommaso *** [0000-0003-0362-8496](#)

Sidharth Rony *** [0009-0008-6828-4704](#)

* Luiss Guido Carli, Italy

** European Commission - Joint Research Centre, Spain

*** United Nations University (UNU) – MERIT, Netherlands

This paper examines how changes in regional skill demand are associated with labour productivity across European regions. The objective is to understand what structure and composition of skills demanded by regional labour markets help explain persistent differences in productivity levels and growth. The analysis is motivated by ongoing technological change and by the digital and green transitions, which are reshaping skill requirements and may generate divergent regional productivity trajectories across space.

The study uses more than 100 million online job advertisements from Lightcast, covering ten European countries at the NUTS-3 regional level between 2014 and 2020. From these postings, 1,747 unique skills are extracted and systematically mapped to the ESCO framework, allowing a distinction between knowledge, skills, and transversal competences. Based on this mapping, the paper constructs multidimensional regional skill portfolios that capture how skills are combined within local labour markets. Using network based methods, three indicators are derived to characterise these portfolios: skill coherence, reflecting how closely related demanded skills are; skill fitness, capturing the sophistication of regional skill demand; and entropy, measuring how evenly skill demand is distributed across different skills. In addition, the analysis separately identifies the regional intensity of digital and green skills to assess their specific associations with productivity.

These skill measures are combined with regional labour productivity data, measured as value added per hour worked, together with standard regional controls. The empirical strategy relies on two complementary approaches. First, two way fixed effects panel regressions are used to analyse how changes in skill demand relate to productivity levels over time. Second, long difference regressions are employed to examine the relationship between shifts in skill demand and

productivity growth over the full period. This design allows the analysis to distinguish between short run associations and longer term productivity dynamics.

The results indicate a nuanced relationship between regional skill demand and productivity. Regions with more coherent skill portfolios, particularly in terms of knowledge, tend to display higher productivity levels. Higher skill fitness is associated with lower productivity in the short run but with higher productivity growth over time, suggesting adjustment costs followed by longer term gains. Regions that move towards a more even distribution of skill demand also tend to exhibit higher productivity levels. With respect to specific skill types, digital skills are positively associated with productivity levels, while green skills are linked to lower productivity growth, reflecting their concentration in relatively lower productivity activities.

Overall, the findings suggest that regional productivity depends on how skills are structured and combined within regional labour markets. The paper highlights the relevance of considering skill demand when designing regional development policies.

Skills intelligence analysis in the Offshore Renewable Energy sector in Europe: A value-chain approach

Eleftherios Sdoukopoulos * [0000-0003-4122-5102](https://orcid.org/0000-0003-4122-5102)

Lucía Fraga Lago **

* Hellenic Institute of Transport (HIT) - Centre for Research and Technology Hellas (CERTH), Greece

** Centro Tecnológico do Mar, Spain

The offshore renewable energy (ORE) sector in Europe is expanding rapidly and constitutes a key pillar in Europe's ambition to become the first carbon-neutral continent by 2050. With the ORE market in the Mediterranean region currently unfolding, following the recent adoption of maritime spatial plans in Spain, Italy and Greece, the sector is at a pivotal stage in its growth trajectory. Conducting regular skills intelligence analysis is therefore more timely than ever, supporting knowledge hubs in the region to introduce new or properly adapt existing educational and training programmes so that the sector is equipped with local, well-qualified and competent personnel. The present communication recaps how these skills intelligence activities have been addressed by the Skills Observatory of the ORESkills large scale partnership in the Pact for Skills, presenting results from the last exercise.

First, two key reference layers were set-up to guide the data collection process and structure the reporting of results. The ORE value chain accounted for the first layer, comprising four key phases i.e. (a) planning, (b) development & construction, (c) operation & maintenance, and (d) decommissioning or recommissioning, as well as a number of cross-cutting activities: specialized transport, research & development, data management, education & training, and other supporting services. The ESCO database was the second reference layer, used to identify occupational profiles (i.e. ESCO Code $\geq$ Level 4) engaged in each activity of the value chain (174 in total). Both layers were validated via a targeted workshop with strong industry participation. Setting-up those reference layers allowed to draw some useful insights: two occupational profiles highlighted by industry experts (i.e. metocean analyst and marine spatial planner) were found not to be included in ESCO.

When analysing skills demand data, in terms of value chain phases and activities addressed vis-à-vis the maturity level of ORE projects in the market addressed, employers' requests were compared with the information reported in ESCO with regard to essential and optional skills, competences and knowledge, considering also the alternative labels listed there for each occupational profile. All those results were reported to advocate for their consideration in future updates of the database.

The analysis considered skills demand, skills offer and skills forecasting. Following a desk research activity, online job vacancies analysis was performed for assessing skills demand (981 offers collected addressing 10 EU countries). The ArcGIS Survey 123 tool was used to formalize the data collection and automatize data aggregation processes at group levels of occupational profiles and at value chain activities considering phases as mentioned before. For the skills supply analysis, 398 educational and training programmes were identified in 13 EU countries (EQF Level 3-8), which once processed were integrated in the MarineTraining platform for facilitating public access. Skills foresight analysis combined desk research with a series of consultation activities (workshops and a two-round Delphi process). As a result, 18 emerging trends and paradigm shifters were identified and grouped. Their impact on skills development was assessed in the short (2030), medium (2040) and long-term thus indicating new occupational profiles likely to emerge.

Skills Intelligence Emilia-Romagna (SI-ER): Leveraging Web Data and AI for Next- Generation Regional Skills Intelligence

Lara Porciatti *

Alain Marenghi *

Nicolò Pranzini *

Francesca Carnevali *

Mauro Pelucchi ** [0009-0000-1354-7556](https://orcid.org/0009-0000-1354-7556)

* ART-ER, Italy

** Lightcast, Italy

Skills Intelligence Emilia-Romagna (SI-ER) is a regional skills intelligence system launched in 2021 by the Emilia-Romagna Region and implemented by ART-ER with support from the European Social Fund Plus (ESF+). The initiative represents a next-generation approach to skills intelligence, combining web-based labour market data, administrative sources, and artificial intelligence to deliver timely, actionable insights for skills policy, education and training design, and workforce planning.

At the core of SI-ER is an open-access digital platform that systematically exploits online job vacancy data as a primary source of skills intelligence. Using machine learning techniques, the platform extracts, classifies, and analyses large volumes of web data from online job advertisements, transforming unstructured information into structured insights on occupations, skills, experience requirements, and emerging trends. Web data are integrated with regional administrative data on employment contracts (SILER) and forward-looking employer survey data from Unioncamere's Excelsior system, enabling triangulation between real-time demand signals, observed hiring outcomes, and expected labour needs. Data are updated on a quarterly basis and structured using the ESCO taxonomy to ensure consistency, transparency, and comparability across European contexts.

SI-ER goes beyond traditional labour market observatories by embedding web data analytics within a strategic and policy-oriented framework. Skills demand is analysed in direct relation to the region's Smart Specialisation Strategy (S3), allowing policymakers and practitioners to identify skills gaps and emerging profiles in key innovation ecosystems such as digital industries, green economy, advanced

manufacturing, agri-food, health, and cultural and creative sectors. The platform supports evidence-based decisions for universities, VET providers, and training institutions in updating curricula, designing upskilling and reskilling pathways, and strengthening career guidance and talent attraction measures.

A distinctive feature of SI-ER is the systematic integration of quantitative web data with qualitative intelligence through permanent multi-stakeholder engagement. Ten thematic focus groups (following the lines of our S3 strategy) are held twice a year. These involve companies, clusters, education providers, employment services, and policymakers, and aim to validate data-driven findings and capture emerging skills not yet visible in online sources. The results feed into 20 sectoral skills reports published annually, reflecting the outcomes emerged from the working group discussions, and containing insights on the subject and proposal for measures deemed a priority to mitigate the skills gap.

The SI-ER platform generates a structured set of concrete outputs designed to support policy design and operational decision-making. These include quarterly interactive dashboards on occupations and skills demand, more than 20 annual sectoral skills intelligence reports aligned with S3 priority areas, thematic briefs on emerging profiles and transversal skills, and open-access datasets for institutional and research use. Outputs are actively used by regional policymakers, universities, VET providers, and employment services to inform programming, funding priorities, curriculum updates, and guidance activities.

The SI-ER experience demonstrates how web data, when combined with AI, administrative records, and participatory governance, can power a robust, user-centred, and sustainable skills intelligence system. It provides a transferable model for regions seeking to harness next-generation data sources to anticipate skills needs and support forward-looking skills policies.

Skills Intelligence for informing HEIs curricula: UPAL Labour Market tools

Natalí Basilico * [0000-0003-4082-0727](#)

Gonzalo Cubillos * [0009-0007-4128-3640](#)

Carmen Pagés * [0000-0002-6164-8383](#)

* Universitat Oberta de Catalunya, Spain

The global labour market is undergoing a profound transformation, driven by megatrends that place unprecedented pressure on workforce resilience. According to recent projections by the World Economic Forum (WEF, 2025) and the International Labour Organization (ILO, 2025), this environment will trigger significant global turnover. International bodies—including the OECD, Cedefop, and the McKinsey Global Institute—emphasise that this transformation is fundamentally reconfiguring occupational profiles, skill demands, and work organisation. In this context, aligning workforce development with industry needs has emerged as a critical challenge for Higher Education Institutions (HEIs).

The Universitat Oberta de Catalunya (UOC)—a pioneer in online education with 97,000 students and more than 6,000 teaching staff—created the Skills Intelligence Unit (UPAL) in 2022, as part of its institutional strategy. The UPAL's mission is to provide scientific knowledge regarding: a) the alignment between teaching programmes and labour market requirements; b) the continuous training needs of workers; and c) the talent demands of employers. One of its initiatives is a skills-intelligence project aimed at academic leaders, teachers, and other community members, which combines an interactive big data visualisation tool (“LMI4E”) with a “Future of Work” report to help align academic programmes with a fast-changing labour market. It integrates various data sources, including 9 million online job advertising (OJA) data from Lightcast and an exhaustive review of international reports on emerging trends.

The “LMI4E” tool allows users to visualise labour market indicators in a detailed, flexible, and personalised manner, while the “Future of Work” report contextualises the data to help identify current and future occupational shifts. A key innovation is the direct linkage between labour market data and academic programmes. Each university programme is matched—with direct input from programme directors—to a specific set of occupations where graduates typically work (using the ESCO taxonomy). This enables academic staff and directors to “deconstruct” degree programmes into skill components, allowing labour market

trends to be tracked not only by occupation but also by academic programme or area. Users can explore trends, shifts, and programme-specific indicators by filtering for programmes, occupations, or skills, thereby facilitating evidence-based decisions for curriculum updates and the design of micro-credentials. The collaboration between the UPAL, faculties and other academic strategic areas ensures the tool remains relevant and programme specific.

Taken together, the UPAL Labour Market tools transform raw data into actionable strategies, enhancing employability and social impact through a scalable framework for curriculum agility. By integrating real-time OJA data into academic governance, the project aims to reduce the lag between market shifts and educational response, ensuring higher education remains responsive to emerging workforce needs. Ultimately, this initiative exemplifies how HEIs can leverage data technologies to build agile, skills-driven learning ecosystems that navigate modern labour market transformations, fostering long-term resilience for both the institution and its graduates.

Spatial Matching of Online Job Vacancies and Training Provision Data

Nermin Oruc * [0000-0001-5026-9153](https://orcid.org/0000-0001-5026-9153)

Aleksandar Kostadinov **

* Centre for Development Evaluation and Social Science Research,
Bosnia and Herzegovina

** University of Ljubljana, Slovenia

Timely and granular skills intelligence is increasingly essential for aligning education and training systems with labour market needs. While online job vacancy (OJV) data provide real-time signals of skills demand, far less attention has been paid to systematically comparing these signals with the structure and spatial distribution of training provision. This paper proposes a big-data approach that integrates OJV data with web-based information on training programmes to analyse skills matching and spatial misalignment across regions.

The study is grounded in skills ecosystem theory and labour market signalling frameworks, conceptualising skills formation as the outcome of interactions between employers, training providers, and learners within both local and regional labour markets. Building on spatial mismatch theory, the analysis explores whether training programmes reflect local employment demand or are implicitly oriented toward international labour markets.

Empirically, the research combines web-scraped vacancy data from online job portals with structured data on vocational, adult-learning, and digital training programmes. AI models are used to match the job descriptions with ISCO classification. Spatial analysis maps demand–supply alignment across municipalities, while matching indicators quantify overlaps, shortages, and potential oversupply. Preliminary results indicate that in some regions training provision prioritises skills strongly demanded in international online labour markets.

The paper demonstrates how integrating multiple web-based data sources can enhance next-generation skills intelligence and support evidence-based skills governance, regional planning, and targeted investment in training systems.

The elephant never left the room: did gender discrimination went implicit in French job posting?

Ylenia Curci * & **

Anne-Sophie Bruno ***

Nathalie Greenan **

Mauro Pelucchi ****

* Université de technologie de Belfort Montbéliard, France

** Conservatoire National des Arts et Métiers (CNAM), France

*** Sorbonne Paris 1, France

**** Lightcast, Italy

In France, since the anti-discrimination law of 2008 (LOI n° 2008-496) and its integration into the labour code in the same year, no person may be excluded from a recruitment process based on their sex. The law stipulates 25 prohibited discrimination criteria (Article L1132-1 of the French Labour Code), including age, sex, origin, family situation, pregnancy, physical appearance, disability, and sexual orientation, amongst others. Since then, job offers cannot be formally directed to a specific category of candidates, and practices for writing more inclusive job postings have become widespread amongst HR recruiters. Nevertheless, gender segregation remains an unresolved—or rather, unaddressed—problem in many sectors in France, calling into question the effectiveness of eliminating formal discrimination in hiring.

In this work, we analyse online French job postings from the perspective of implicit employment norms: unstated expectations or assumptions about employee characteristics, behaviours, or availability that often originate in the private sphere and become embedded in workplace practices without explicit articulation. These implicit norms are frequently gendered, reflecting structural differences in labour market expectations for women and men regarding physical characteristics, temporal availability, geographical mobility, work dedication, and family responsibilities.

The analysis focuses on three key dimensions: physical risks; working time and work-life balance; and forms of employment and career development. Our methodology employs a comparative benchmark approach, testing multiple methods to classify job postings as gendered, including naïve human reading,

expert human reading, dictionary-based analysis, and the use of large language models (LLMs). We perform this benchmark analysis on a corpus of job postings from firms included in the REPONSE survey, a representative survey that allows us to control for selection bias due to the limited use of online job postings in France, especially in gender-segregated sectors such as construction and in micro-enterprises (DARES, Enquête Offre d'emploi et recrutement 2016), which are not subject to any form of gender equality monitoring.

By merging this information with census data on the labour force (DADS database), we examine whether implicit gender norms reinforce gender segregation in the French labour market. Specifically, we employ a matching methodology that links individual job postings to actual hiring outcomes at the firm level. Using unique firm identifiers present in both the online job posting corpus and the DADS administrative data, we observe the gender composition of new hires within specific occupational categories and time windows corresponding to each posting. This allows us to test whether job postings containing gendered implicit norms are associated with gender-differentiated hiring patterns, thereby providing empirical evidence of whether and how linguistic signals in recruitment advertisements translate into actual gender segregation in employment outcomes.

The Hidden Innovators: Uncovering Capabilities and Industrial Dynamics in European Rural Areas

Simone Sasso * [0009-0007-5645-4828](#)

Tommaso Ciarli ** [0000-0003-0362-8496](#)

Šimon Trlifaj ** [0000-0002-1519-8369](#)

Bernardo Caldarola * [0000-0002-4970-1583](#)

Emanuele Pugliese ** [0000-0001-7505-4183](#)

Alessio Bumbea *** [0009-0005-1358-5823](#)

* European Commission, Belgium

** United Nations University (UNU) – MERIT, Netherlands

*** Universitas Mercatorum, Italy

Regional disparities, and particularly the innovation gap between urban and rural regions, remain a central challenge for both competitiveness and cohesion in Europe. Policy frameworks such as Smart Specialisation Strategies often rely on conventional indicators, such as patents and R&D expenditure, that tend to capture innovation in urban contexts while undercounting forms of innovation typical of rural economies (Balland et al., 2020; Eder, 2019). This raises the risk of misrepresenting the innovation capabilities present outside metropolitan areas. In this paper, we address this gap by investigating the skill portfolios and capabilities that characterise European regions, and how these co-evolve with firm dynamics to shape regional development.

Our conceptual framework builds on the idea that regional innovation is the outcome of coevolutionary dynamics between skills, entrepreneurship and other regional characteristics (Pugliese et al., 2019; Coad et al., 2021). Rather than focusing on traditional innovation outputs, we analyse innovation potential through the structure of regional skill demand, entrepreneurial activity, and firm dynamics. This approach enables us to identify “hidden innovators” and to better understand how different regions combine capabilities over time.

Empirically, we construct a granular panel dataset for European regions at NUTS3 and LAU level by combining multiple data sources. To measure skills and occupations, we leverage millions of Online Job Vacancies (OJV), which provide detailed information on firms’ skill demand. Using name-entity recognition, we match vacancies to firms in Orbis and Dealroom, linking firm dynamics and skills

demand. For the regional context, we integrate indicators on demographics, infrastructure, and finance from Eurostat, ARDECO, and the JRC Rural Observatory.

To characterise regional innovation capabilities, we develop skill-based indicators inspired by the economic complexity literature (Neffke and Henning, 2013; Aufiero et al., 2024; Tacchella et al., 2012; Hidalgo and Hausmann, 2009). Using OJV data, we compute measures of diversity, coherence, and sophistication that capture different dimensions of local capabilities. In addition, we identify occupations and skills associated with innovative activity identifying skills that co-occur within innovative firms. Initial results on regional skill portfolios point to a central role for geographic proximity and local spillovers, with neighbouring regions displaying more similar skill structures, alongside marked spatial heterogeneity.

The analysis provides a new geography of innovation potential in Europe, uncovering rural regions with substantial latent capabilities that remain invisible in traditional metrics. By leveraging web-based skills data and integrating them with firm-level and regional information, the paper contributes to next-generation skills intelligence and provides a new empirical basis for place-sensitive innovation and regional development policies.

The ROI of Guidance: Measuring the Impact of Human-in-the-Loop Curation on STEM Talent Mobility

Harshit Jaiswal * [0009-0008-4821-2447](https://orcid.org/0009-0008-4821-2447)

* Flow Community Inc, United States of America

Traditional labour market analysis treats the application process as a “black box”: a job is posted, and a candidate applies. This binary view ignores the critical intermediate layer: Guidance. For high-stakes technical roles in Computer Science (CS) and Electrical Engineering (EE), the primary barrier to entry is often not a lack of skill, but a lack of contextual awareness and personalized application strategy.

This paper presents empirical findings from Flow (withflow.co), a live talent marketplace that has operated for over a decade. Our insights are not theoretical; they are derived from 10 years of operational data connecting 15,000+ engineers with a decentralized network of career counsellors. As a mature product, Flow has evolved into a “Digital Orchestration Layer” for career development services, aggregating high-signal opportunities (“Flow Cards”) and distributing them to vetted counsellors who “bid” to offer 1-on-1 or 1-to-many video guidance.

Methodology & Data Sources:

We analyze the interaction data from this “Managed Guidance Marketplace” to quantify the value of human intervention:

- (a) The Curation Feed: We track the downstream impact of providing career centers with a pre-vetted feed of CS/EE opportunities, measuring how “curation” reduces the administrative burden on counsellors and increases their capacity to coach.
- (b) Intervention Metrics: We compare application success rates between three cohorts:
 - (i) unassisted applicants;
 - (ii) applicants who attended a “Flow AMA” (1-to-many video session);
 - (iii) applicants who received 1-on-1 personalized guidance.
- (c) Counsellor-Led Intelligence: We capture the qualitative feedback loop where counsellors identify why a candidate was rejected (e.g., “portfolio format” vs. “coding ability”), providing granular “failure data” that OJA scraping cannot see.

Key Findings & Contribution:

Our findings suggest that Guidance is the missing variable in skills intelligence.

- (a) The “Confidence Gap”: We find that for CS/EE roles, personalized guidance increases application completion rates by 300%, validating that “human-in-the-loop” support is essential for unlocking talent mobility.
- (b) Scalable Career Services: We demonstrate a model where career centers can leverage “Web Data” (our curated feed) to scale their impact, moving from reactive advising to proactive, opportunity-driven coaching.

Relevance to Policy:

This research offers a roadmap for modernizing Public Employment Services (PES) and University Career Centers. By integrating a “Guidance Marketplace” layer, institutions can move beyond generic advice and offer targeted, data-driven support for specific high-demand sectors. We argue that the next generation of skills intelligence must measure not just who is hired, but who helped them get there.

The role of education and training offer data in skills intelligence

Alice Bertoletti * [0000-0001-8539-2098](#)

Judith Cosgrove * [0009-0009-8675-2900](#)

Montserrat Lopez Cobo * [0000-0003-3351-1352](#)

* European Commission - Joint Research Centre (JRC), Spain

Emerging technologies are reshaping labour markets by transforming job tasks, altering employment structures, and driving new patterns of adoption across sectors. Tracking emerging skill needs is therefore essential for effective workforce planning and matching labour and societal demand with education and training offer. This raises the question of whether education and training systems have the capacity to equip individuals with the skills required to adopt, use, and adapt to these technologies. Addressing this issue requires robust indicators that capture developments in tertiary education, as well as in lifelong learning opportunities.

In this study, we describe characteristics of key data sources and indicators on education and training skills-related offer. We then provide examples of additional indicators that we derived on the basis of web data from universities' programmes syllabi. Finally, we present recent results obtained in relation to two emerging digital technology domains: artificial intelligence and virtual reality, demonstrating their relevance to EU policy initiatives. We offer suggestions for future work on skills intelligence for education and training offer by considering the substantial measurement challenges that often arise when assessing indicators related to skills provision as part of a skills intelligence ecosystem, and possible ways to address these challenges. Moreover, our findings aim to contribute to the broader debate on emerging digital and technological skills mismatches.

Tracking Upskilling and Reskilling via Skill Evolution in GitHub Profile READMEs

Maria Papoutsoglou *

Emmanouil Rigas ** [0000-0002-8042-9135](#)

Georgios Meditskos * [0000-0003-4242-5245](#)

Georgia Kapitsaki ***

Efstratios Kontopoulos ****

Nick Bassiliades * [0000-0001-6035-1038](#)

Christoph Treude ***

* School of Informatics, Aristotle University of Thessaloniki, Greece

** School of Medicine, Aristotle University of Thessaloniki, Greece

*** Computer Science Department, University of Cyprus, Cyprus

**** Choreograph, United States

The recent massive uptake of generative AI across occupations is accelerating the pace at which workers adapt their skills. Alongside long-standing technical competences, new skills are rapidly emerging and becoming mainstream, including prompt-related practices and adjacent capabilities that reshape daily workflows. In this context, it is increasingly important to:

- (a) identify skill changes as they occur over time
- (b) qualify these changes as upskilling or reskilling processes, depending on whether they deepen an existing occupational pathway or signal a shift towards new competence domains.

Online web sources provide an ideal environment for observing such changes. Their content is dynamic, continuously refreshed, and frequently supports regional or community-based perspectives that reflect economic and geographic differences. Compared to static surveys or infrequent reporting cycles, web data can capture early signals of change and offer a scalable foundation for next-generation skills intelligence.

This paper focuses on a well-known and representative web source for developer labour market intelligence: the profile README files of GitHub users. These README files function as public, self-curated CV repositories where developers describe themselves, outline activities they seek, and report expertise in free text. Importantly, each README includes an edit history, making it possible

to reconstruct the time evolution of the self-reported skill set and detect when skills are introduced, emphasised, rephrased or removed.

We collected 37,898 publicly available developer profile READMEs with accessible revision histories. Using state-of-the-art natural language processing pipelines, we extract skill mentions and normalise them to support aggregation across profiles and time. To interpret skill changes in a structured way, we align extracted skills with the ESCO classification and exploit ESCO's RDF/TTL "knowledge graph" version as a time-aware semantic backbone to support consistent mapping across related labels and hierarchies. Based on the detected transitions in ESCO space, we operationalise a distinction between upskilling (skill enrichment within the same occupational/competence area) and reskilling (movement towards different competence areas and skill clusters). This enables the quantification of change patterns over time, by skill group and by occupational segment.

The contribution of the work is directly linked to skills demand by occupation: we map required and emerging skills for specific jobs and occupational groups using NLP and classification techniques, enriched by graph-based reasoning over ESCO. Early findings indicate rapid shifts in the composition of skill portfolios, with clear signals of newly emphasised competences and hybridisation trends in developer profiles over short time windows.

The proposed approach demonstrates how dynamic web data and skill taxonomies can be combined to detect and interpret skill evolution, supporting timely evidence for training design, workforce planning, and skills policy.

Using Online Job Adverts Data on Skills to Identify Potential Career Pathways

Nora Condon *

* Skills and Labour Market Research Unit, SOLAS, Ireland

Recent research from the OECD (2023) indicates that 33% of workers in Ireland report that their skills exceed the requirements of their current roles. Skills mismatch in Ireland (and to a lesser extent, qualifications mismatch) is particularly prevalent among those working in medium skilled (e.g. administrative) and lower skilled (e.g. elementary) occupations. This paper aims to identify the skills necessary to transition between occupations, and in particular the skills required to move from lower-skilled to higher-skilled occupations.

Using the Ireland subset of online job adverts (OJAs) data, available from the Web Intelligence Hub, we first group detailed occupations together based on the extent to which they are associated with similar sectors or areas of business; for example, various construction occupations (e.g. civil engineers, carpenters, and construction operative/elementary workers) are grouped to form an occupational family. We then identify the top skills in demand for each occupation within a given family and compare the skills in demand by occupational skill level (e.g. operative level occupations compared to technician level occupations).

The results highlight:

- (a) Common skills: skills in demand across most occupations within a family, indicating the must-have skills (e.g. economics, tolerate stress and communication were key skills across all financial occupations)
- (b) Unique skills: skills present in higher skilled occupations that are absent in medium or lower skilled occupations; for example, project management and lead a team were among the top skills mentioned in professional financial occupations, but absent from most associate professional and administrative financial occupations, pointing to possible areas for skills development for an individual to progress to higher level occupations, and possibly a better match to their skill set.

Such an analysis of skills data is only possible in large datasets. Other datasets for which skills data are available (e.g. company skills audits) are useful but tend to be small in volume and refer to a limited range of job types. The examination of OJAs, when taken in the broader labour market context (e.g.

employment levels, growth, etc.) can set out key areas of focus for education and training to foster skills development and career progression.

The accelerating pace of change in the labour market means that workers will need to develop new skills throughout their working lives. It is therefore critical that labour market intelligence include regular and timely analyses of skills demand. The analysis presented here represents only one step in better aligning workforce skills with demand. Much work is needed in the areas of career guidance and lifelong learning: not only do lifelong learning opportunities need to be offered to workers, workers need to be further encouraged, facilitated (time, cost etc), and just as importantly made aware of and guided towards relevant education/training provision.



Harnessing web data for next-generation **skills intelligence**

BOOK OF ABSTRACTS

**CEDEFOP**European Centre for the Development
of Vocational Training

Europe 123, Thessaloniki (Pylaia), Greece

Postal: Cedefop service post, 57001 Themi, Greece

Tel. +30 2310490111, Fax +30 2310490020

Email: info@cedefop.europa.eu

visit our portal www.cedefop.europa.eu