Co-funded By the European Commission

AI-based Industry 5.0 skill demand forecast methods.

WP3 Data retrieval and AI-analysis of vacancies and Industry 5.0 organisational practices

Deliverable D3.2 - AI-based Industry 5.0 skill demand forecast methods.

Authors: Jeisson Cardenas-Rubio et al.
Associated Work Package: WP3
Lead Beneficiary – IER
WP leader: Jeisson Cárdenas-Rubio (IER)
File: BRI_WP3_D3.2 AI-based Industry 5.0 skill demand forecast methods_20250424.docx
Version 0.9

Funded by the European Union under grant agreement No 101069651. The contents of this publication are however the sole responsibility of the BRIDGES 5.0 project consortium only and do not necessarily reflect those of the European Union or HADEA. Neither the European Union nor HADEA can be held responsible for them.

Document Summary

Funding:	Horizon Europe
Call:	HORIZON-CL4-2021-HUMAN-01
Grant Agreement:	101069651
Acronym:	Bridges 5.0
Project title:	Bridging Risks to an Inclusive Digital and Green future by Enhancing workforce Skills for industry 5.0
Project start-end:	1-1-2023 – 1-1-2027
Period covered:	1-1-2023 – 1-2-2024

Document type:	Public report
Title:	AI-based Industry 5.0 skill demand forecast methods
Author/s:	Jeisson Cardenas-Rubio / Pascal Kampert / Wouter Korteling
Reviewer/s:	Steven Dhondt, Peter Totterdill
Date:	May 2025
Document status:	Version 1.0
Keywords:	Large Language Model; skills; Industry 5.0; forecasting
Version:	1.0
Document level:	Public

Document description

This report presents a proof-of-concept for an AI-based methodology to forecast future skill demands and to explore their relationship with organisational practices related to Industry 5.0. The research addresses three key questions: (1) How can an AI-based methodology be designed and implemented to forecast future skill demands and to analyse their relationship with practices characteristic of Industry 5.0 organisations?; (2) To what extent can the predictions of future skill demand produced by the AI model be explained?; (3) What are the methodological and data-related limitations of the proposed approaches, and how can these be addressed or mitigated in future applications?

Cite this deliverable as:

Cárdenas-Rubio, J., Kampert, P., Korteling, W. (2024). AI-based Industry 5.0 skill demand forecast methods (BRIDGES 5.0 deliverable D3.2/ version 1 – December 2024). Warwick: BRIDGES 5.0. (Retrieved from: TBC).

BRIDGES 5.0 partners*

TNO, The University of Warwick
*all participating in this Deliverable

Funded by the European Union under grant agreement No 101069651. The contents of this publication are however the sole responsibility of the BRIDGES 5.0 project consortium only and do not necessarily reflect those of the European Union or HADEA. Neither the European Union nor HADEA can be held responsible for them.

Version management

Version	Date	Action
0.1	1-3-25	Fully reviewed report, to be submitted to EB.
0.2	14-4-25	Approval EB
1.0	9-5-25	Submitted to EU Portal

List of abbreviations

BRIDGES	Bridging Risks to an Inclusive Digital and Green future by Enhancing workforce Skills for industry 5.0
CNAM	Conservatoire National Des Arts et Métiers
EU	European Union
GPT-3	Generative Pre-trained Transformer 3
HADEA	Research Executive Agency
I4.0	Industry 4.0
I5.0	Industry 5.0
LMIS	Labour Market Information Systems
LLM	Large Language Models
Llama	Large Language Model Meta AI
TNO	Nederlandse organisatie voor toegepast natuurwetenschappelijk onderzoek TNO
WP(s)	Work Package(s)

Table of Contents

Executive Summary.

This report presents a proof-of-concept for an AI-based methodology to forecast future skill demands and to explore their relationship with organisational practices related to Industry 5.0. The research addresses three key questions.

1) How can an AI-based methodology be designed and implemented to forecast future skill demands and to analyse their relationship with practices characteristic of Industry 5.0 organisations?
2) To what extent can the predictions of future skill demand produced by the AI model be explained?
3) What are the methodological and data-related limitations of the proposed approaches, and how can these be addressed or mitigated in future applications?

To answer these questions, this report first outlines the need and current state-of-the-art on predictions of skills demand in the labour market. It then builds on this by proposing a machine learning (ML) method, outlining how it can be used for insights and implementing it on two datasets as a proof of concept. The datasets we use are the Skills-OVATE dataset from Cedefop and a UK OJA dataset developed in the Bridges 5.0 programme. The resulting conclusions are threefold.

1) The trained models were able to identify which skill categories are likely to grow or decline in relevance. Using hypothetical scenarios of increased Industry 5.0 adoption, the model could estimate corresponding changes in skill needs for different occupation groups. However, the current implementation forecasts only a one-year time horizon; future studies should explore multi-year predictions when higher-quality data becomes available.
2) Scenario inputs function as “what-if” explanations, showing how model outputs shift under different assumptions, such as varying assumptions on how organisational practices will change in the future. Further explainability can be achieved using feature importance methods or by replacing complex models like XGBoost with more interpretable alternatives, such as Decision Trees, especially when the feature set is limited.
3) The data-driven methods discussed in this report should be complemented by theoretical and qualitative approaches. Further, a key methodological limitation of the Time Series-Based method that was applied is the reliance on aggregated time-based data. We expect that this can be mitigated by shifting to a **company-level prediction framework**. However, the primary data limitation lies in the lack of company-level contextual information in existing datasets such as Skills-OVATE OJA. The performance and usability of the applied model will likely significantly increase if such company-level information is included.

To address these limitations, we recommend that future research efforts focus on **enriching (vacancy) datasets with company-level information,** such as sustainability ratings, customer reviews, technological maturity, and other organisational indicators. The **Bridges UK OJA dataset** provides a promising starting point for such enrichment. Ultimately, a **company-level ML approach**, where

individual companies serve as the learning unit, will enable deeper insight into how Industry 5.0 practices influence skill requirements over time.

1.Introduction.

1.1 Relevance.

Future European labour markets are expected to become subject to multiple, different transformations of skill demands. The European Union identifies two major sources of changes in skill demands: the digital and the ecological transition, together referred to as the twin transition. Both entail the creation of new and the disappearance of old skills, activities, jobs and possibly even entire occupations. Labour market transformations of such magnitude affect a broad set of economic outcomes and raise the need for effectively guiding skills policy. In fact, a forward-looking understanding of such transformations in skill demands – skill foresight - is a declared policy objective of the European Union since the Maastricht communiqué was agreed upon 20 years ago (European VET ministers, 2004). Yet, uncertainty regarding future skill demands remains high. Recently, the twin transition incited the European Union to adopt a new industrial strategy (European Commission, 2020). Multiple independent expert policy briefs, published by the European Commission, emphasised the relevance of adopting Industry5.0 to reach the goals of the new EU industrial strategy (i.e. European Commission, 2021; European Commission, 2022). While still in conceptual development, Industry 5.0 introduces three new dimensions to Klaus Schwab's highly digitalised and autonomous Industry 4.0 (2017): human-centricity, sustainability and resilience (Oeij et al., 2023).

The HORIZON EUROPE project Bridges 5.0, which this report is funded by and a part of, serves to understand what skill demand and supply changes European labour markets need to address when transitioning to Industry 5.0 (Oeij et al., 2023). While various econometric models (e.g. Autor et al., 2004; Brynjolffson et al., 2018; Cedefop Skills Forecast) and machine-learning (ML) approaches (e.g. Mühlbauer and Weber, 2022; Macedo et al., 2022) have been developed to analyse skill trends, these methods currently lack the predictive power needed to proactively design skill policies as envisioned by the policy debate of the last decade (ILO, 2016; Colombo et al., 2018; Cedefop, 2024). ML-based approaches, which have emerged more recently and depart from traditional methods of economic modelling, remain underexplored and hold significant theoretical promise. Reasons for this are technological progress: the digitalisation of the labour market, which yielded large amounts of new data, vacancy- and employee-connected, progresses in the ability to process unstructured data, but also general methodological considerations: ML is capable to process and identify connections between unforeseen amounts of variables and potentially connected factors (such as for instance the development of demand for other, similar skills) (Giabelli et al., 2022). This does not imply superiority of one approach over another as ML-approaches still come with severe drawbacks (i.e. transparency and interpretability), rather, a well-composed approach to skills foresight should entail reasonably balanced and diverse tools (Cedefop; 2024).

This report focuses on the development of methodologies for ML-based predictions of skills demand and how such demand relates to alternating degrees in which Industry 5.0 company practices are implemented. As such, this report exclusively considers generic ML-based, theory-agnostic prediction models. We utilise two datasets: (1) the Online Job Advertisements (OJA) dataset based on Eurostat's and Cedefop's Skills-OVATE database (Cedefop, 2021); (2) the Bridges UK Vacancies dataset developed within the Bridges 5.0 project.

1.2 Research Questions and Structure of This Report.

Emerging from the Bridges 5.0 grant agreement, this report focuses on the following research questions:

- **RQ1.** How can an AI-based methodology be designed and implemented to forecast future skill demands and to analyse their relationship with practices characteristic of Industry 5.0 organisations?
- **RQ2.** To what extent can the predictions of future skill demand produced by the AI model be explained?
- **RQ3.** What are the methodological and data-related limitations of the proposed approaches, and how can these be addressed or mitigated in future applications?

This report first outlines relevant prior work leading to the current state-of-the-art methodologies of skills forecasting. Afterwards, we propose and describe two ML-based methods for skills forecasting that seem, in our view, most promising. These are then implemented on both datasets as a proof-of-concept (POC) after which the main insights in the potential of these models, as well as their requirements in terms of necessary data, are discussed.

2. Prior Research.

2.1 Skills Foresight through econometric modelling.

The first attempts to predictively prepare the workforce for transitioning labour markets date back to efforts of the OECD in the 1960s and were developed in the following by econometricians (Parnes, 1962; Willems, 1996). The models used were built not on the basis of individual tasks or skills, but on aggregated educational or skill groups: high-, mid- and low-skilled or -educated. This remained the most prominent approach for the rest of the 20th century, also when assessing the impact of digital technologies through which skill demand changes overarch educational groups (see Katz and Murphy, 1992). A reason for this is a constraint in granularity and availability of data on skills, as researchers were restricted to household surveys and administrative records (Sparreboom, 2013; Mezzanzanica and Mercorio, 2018). Within the context of technological change, Autor, et al. (2003), were the first to make use of the then freshly expanded O*NET taxonomy to broadly categorise activities based on the type of task instead of hierarchy or educational affiliation. The authors then assumed that a historic trend of declining demand in one category, routine tasks, will continue and that higher emphasis should be put on equipping workers with non-routine skills (Autor et al., 2003). Brynjolfsson et al. (2018) calculated suitability for ML scores per occupation, based on priorly assessed scores for each Detailed Work Activity (DWA) in O*NET, to assess which tasks and occupations will see declines in demand due to being theoretically executable by ML. Since then, many others have calculated automation exposure scores based on the automatability of each individual skill contained in an occupation (e.g. Felten et al. 2018 using O*NET and technology data; Nedelkoska and Quintini, 2018 using PIAAC; Webb, 2020 using patent data and O*NET). Next to the methodical act of generating insights on future skill demands, these studies generally aimed to enrich the policy debate and were not of pure methodological nature.

2.2 Methodological advancements beyond econometric models paving the way for novel approaches.

Beyond econometric modelling, novel methods have been developed or made feasible for economic research in the last decade. In some cases, new technologies have been implemented into existing methodologies for existing questions, replacing other techniques. One example is the usage of LLMs for the evaluation of automatability: while prior research calculated automation indicators on economic theory and expert assessments (i.e. Autor et al., 2003; Brynjolffson et al, 2018), researchers prompted GPT-4 to come up with automation scores (Eloundou et al., 2024; Gmyrek et al., 2024). Other methodological advances enable researchers to ask new questions or existing questions in greater depth, due to enhanced processing capabilities: recent advances and popularity gains in NLP allow for large-scale web-scraping and qualitative exploration of unstructured data, such as vacancy texts. Economists adopted web-scraping and NLP as Python libraries such as *NLTK* or *beautifulsoup4* gained popularity in the mid-2010s with courses and ‘cookbooks’ made widely available to non-adept users (Perkins, 2010; Hajba, 2018). Whether a skill is demanded in a certain occupation could now be assessed in real time and without costly employer surveys with the help of TF-IDF (term-frequency inverted-document-frequency) or RCA (relative comparative advantages)-based analyses. The increases in breadth and depth of available data simultaneously paved the way for the adoption of methods which existed in other fields but were priorly less suited for labour market research: Operations Research (OR)-based approaches like graph modelling and optimisation algorithms, could now be applied to labour market research. Colace et al. (2019) used topic modelling to classify online vacancies to ISCO codes enhanced by explainable AI methods to generate explanations for classifications, adding a layer of theoretical validity as compared to prior ML-based classification attempts, which often are critiqued due to their opaqueness and thus no prediction interpretation (Boselli et al. 2017a; Biran and Cotton, 2017; Colace et al. 2019).

2.3 Labour Market Intelligence, LMI-systems and ML-based approaches for skills foresight.

With novel methods and more granular, more up-to-date data for labour markets gaining popularity, the field of labour market intelligence or labour market information (LMI) emerged in the mid-2010s (Mezzanzanica and Mercorio, 2018). LMI refers to the process and outcome of AI-generated labour market data (Boselli et al., 2017b). LMI systems (LMIS) have no strict definition but can generally be understood as broad and recent labour market data collected and administered by a dedicated institution and made available in a comprehensive and easily accessible fashion to anyone. LMIS were coined as a concept by an ILO economist, originally for developing economies, but has established itself successfully mainly in advanced economies (Sparreboom, 2013; Barnes et al., 2023). In the US, the Bureau of Labour Statistics provides federal information, but the administration of LMIS is the responsibility of the respective state's departments of labour (US Congress, 2014). The responsibility for a shared EU-LMIS with a focus on skills has been given to Eurofound and Cedefop, which launched the "Skills Online Vacancy Analysis Tool for Europe", skills-OVATE, in 2018 (Cedefop, 2021). Skills-OVATE makes use of multiple non-traditional methods, such as web-scraping and NLP. The OJA database is part of skills-OVATE and covers job advertisements on many paid platforms across the EU (Beręsewicz and Pater, 2021). Other European LMIS are the European Labour Market Barometer, administered by the German IAB, and a newly established labour market intelligence tool by the European employment services EURES, Cedefop and Eurostat, which contains more general labour market information for EU member states.

2.4 Current attempts to skill demand prediction and known gaps.

Direct attempts at ML-based skills demand forecasting are rare at this point. This report builds on the work of two working papers of a research group from IBM and MIT and applies it to European data (Das et al., 2020; Macedo et al., 2022). With similar motivation, to make use of an LMI-system to equip policymakers and labour market stakeholders with the knowledge necessary to quickly react to skill demand changes, Das et al. (2020) built and deployed an ARIMA forecasting model, and Macedo et al. (2022) use convolutional neural networks (CNN) and long-short term memory (LSTM) models to predict skill demand using historical skill demand (skill share) data from online job advertisements. This report contributes to the policy debate by applying the goal to the European LMI infrastructure and outlining needs for further research. Another contribution is in the application to Industry 5.0, which aids subsequent work packages of the Bridges 5.0 project in forecasting further needs for an Industry 5.0 transition and the design of training interventions.

3. Data.

In this section, we describe our source data, namely skills-OVATE and the Bridges UK Vacancies dataset, which were used to apply the ML-based prediction methods described in this report.

3.1 Skills-OVATE OJA Data.

The skills-focused Labour Market Intelligence System (LMIS) skills-OVATE comprises frequently updated data on European skill demands from multiple sources. Online Job Advertisements (OJA) of open vacancies are the main pillar of skills-OVATE. OJA are not the totality of vacancies on the labour market, as in certain positions and regions, vacancies are preferably filled using recruiting agencies or other non-public or offline means. Yet, for the majority of EU labour markets, OJA have become the most common practice to advertise a vacancy (Podjanin et al., 2020, Chapter 1.1). Through this, in combination with a broad coverage of online job platforms, skills-OVATE OJA data provides the most comprehensive overview of skill demands in European labour markets to date (Barnes et al., 2023; Beręsewicz and Pater, 2021).

3.1.1 OJA Data Collection – What is Known.

While, as for many big data LMISs, no official documentation exists for skills-OVATE, a range of sideline publications exists, to which we refer the curious reader: the tender for the gathering of OJA data was held in 2014 and awarded to an Italian research group based in Milan (Cedefop, 2014). Together with Cedefop, this group has continued to perform groundbreaking work on skills-OVATE and advanced the field of LMI significantly with papers published within the context of the Cedefop projects, which give an indication towards the collection efforts in OJA (i.e. Boselli et al., 2017a; Boselli et al. 2017b; Mezzanzanica and Mercorio, 2018; Colace et al., 2019; Giabelli et al., 2022). Next to these publications, Cedefop regularly provided news and briefs regarding the most recent progress on skill-OVATE, which by themselves do not document the data collection in-depth but provide an image of the development process. Moreover, Cedefop and other organisations such as the ILO or ITU provided regular reports on the usage of OJA data for skills forecasting, in which members of the Italian research group, but also skills forecast coordinators, shared their methodological insights (i.e. Podjanin et al. 2020, Chapters 1.1, 1.2, 1.3). Lastly, methodological aspects for scientific use, such as representativeness and veracity of gathered data, were discussed by a stakeholder report written by skills-OVATE collaborators from Eurostat (Beręsewicz and Pater, 2021).
We accessed Skills-OVATE OJA data over the Web Intelligence Hub (WIH) hosted by the European Commission. Considerations regarding the methodological aspects of the data collection are gathered below.

3.1.2 Data Structure.

The data is stored within a frequently updated relational SQL database and can be accessed by querying the Web Intelligence Hub (WIH). As such, no one tabular source exists or is made available to researchers; skills-OVATE OJA generates tabular data according to the user's query. The accessible OJA Data contains up to 84 variables for each skill mentioned in an OJA and can thus contain as many observations per OJA as skills have been identified. We put the variables into 14 categories for simplicity:

- **Identification Variables:** Enables identifying an OJA (useful as one and the same OJA normally is referred to in multiple rows).
- **Date Variables:** Contain the first and last active date of an OJA, as a total date and individual variables for day, month and year.
- **Language Variables:** Names the language and OJA originally was posted in, classified according to ISO 639-1.
- **Occupation Variables:** Contain the ISCO codes for occupations, on 1- to 4-digit level
- **Skill Variables:** Contain ESCO skill codes and relate them to other ESCO skills within the ESCO hierarchy.
- **Location Variables:** Contain information on locations of the OJA's corresponding vacancy (country, city and NUTS-regions).
- **Contract Variables:** Contains information on what type of contract is mentioned in OJA.
- **Education Variables:** Contains information on the required education level.
- **Economic Activity Variables:** NACE Rev2 code for economic activities
- **Salary Variables:** Contains the Category of salary range and salary range in increments of 10,000€ per year.
- **Working Time Variables:** Contains distinction into full-time, part-time and unknown, plus category variable.
- **Experience Variables:** Contains information on the required working experience in years.
- **Source Variables:** Contains information on job platform OJA was found on (name, country, stability).
- **ESCO Variables:** Contains connection of identified skill to ESCO concepts, such as digital competencies or green skills.

Next to the categorical information a column contains (i.e. between 20,000€ to 30,000€ per year for salary), oftentimes an additional column with the suffix _id is provided, referring to just the category (i.e. category 3 for 20,000€ to 30,000€ per year). This means that an observation always has fewer than 84 information points. Even though vacancies usually do contain company and unit-level information, in skill-OVATE OJA no direct information about the company is provided, except for the economic sector in which a company is active. There is no explicit explanation for this removal of data, but likely explanations include privacy considerations and exclusion of usage for commercial purposes. The availability of data on the economic sector in which a company operates allows for between-sector comparisons, but does not allow for comparisons of different types of companies within a sector.

The exact variables and affiliation to categories are in Table x in Appendix A.

3.1.3 Limitations.

While skills-Ovate OJA data constitutes a leading LMIS and offers significant and novel opportunities to researchers, it contains limitations which directly affected our efforts to predict Industry 5.0 skill demands. Some of the limitations listed below might be of lesser magnitude than we assume, since we decided to pre-emptively expect certain limitations due to missing documentation. Other limitations were mentioned by researchers affiliated with skills-OVATE OJA in prior publications and can be understood as factual.

The first set of limitations refers directly to the opaqueness induced by the absence of official documentation. While Skills-OVATE OJA data is comparatively well documented in sideline publications regarding releases, updates and technical features, researchers must nonetheless expect induced statistical biases as well as problems with the veracity of data and reproducibility. One example is the mapping of job titles in OJAs and ESCO. Job titles are entered into the job platform by a representative of an employer who, in all likelihood, is neither familiar with ESCO codes nor skill profiles but aims to find a suitable fit to their implicit assumptions about the position. Further in the pipeline, a job platform also does not ask for or provide an ESCO classification. We assume the ESCO classification is obtained through mapping the online job title to an ESCO occupation title, regardless of whether a match would hold up to the ESCO understanding of said occupation. If such a mapping is performed over a LLM or classic sentence-BERT, similarity scores make the veracity of the mapping dependent on the training data of the tool used (Lipton, 2018), inducing further opaqueness.

Another set of considerations emerges from the nature of OJAs, which are not directly intended to list an accurate skills profile but to attract good candidates. While skills expectations are partially named explicitly and partially implicitly communicated through the job description, the extracted skills, however, are only those which are mentioned explicitly. An OJA also does not represent the entire communication on skills an employer might have with a candidate. Some employers might also hope for candidates to list their own skills extensively and then pick the ones they find to be best-suited, which would delay the mentioning of skills to a phase which OJA analysis does not capture. Moreover, an OJA is not to be confused with a vacancy; a vacancy is the actual job to be awarded, whereas an OJA is just the online listing. Some vacancies have multiple OJAS on different platforms, while others might have none (Beresewicz and Pater, 2021). The data, however, are not directly linkable to a vacancy, but only to an OJA.

The limitation affecting our work most profoundly is that hardly any company-level information is provided. We identified two strategies to address this issue, the first being the use of economic sector information. This option is ultimately unsuited for our purposes because the distribution of Industry 5.0 practices is heterogeneous within sectors and depends more on individual company practices. The second option was to “translate” some of the available information contained in the dataset into expected Industry 5.0-related practices by the company behind the OJA. To this end, we created a “sustainable company practice” dictionary that identifies several ESCO skills (that are available in the dataset) that may be regarded as a proxy of sustainable company practices. The definition of this dictionary is described in the next section.

3.1.4 Sustainability Dictionary.

In order to attempt to retrieve some insight into the practices of the companies behind the OJAs, we created a dictionary that relates ESCO skills to sustainable company practices. This dictionary contains a list of ESCO skills that might indicate that the company posting the vacancy and demanding these skills is more likely to have adopted sustainable company practices.

The dictionary is based on the “green skill” label developed by ESCO. This green skill label is an annotation of skills that indicates that the labelled skill is relevant to the green transition. In total, ESCO labelled 591 ESCO skills as being “green”. Some of the labelled skills, however, were in our view clearly not related to company-level sustainable practices. An example is “develop management plans”. Such a skill might be relevant to the green transition, but when it is encountered in a vacancy, you cannot use it to infer anything about sustainable company practices, as it is not green-specific. In total, we selected 489 skills that we included in our sustainable company practice dictionary.

It should be noted that this dictionary provides, at best, a weak proxy of sustainable company practices. The ESCO skills on which it is based are not designed as indicators of company practices, nor was the extraction of these skills the purpose of the skills-OVATE dataset. Still, it is the only way to attempt to retrieve company-level information on sustainable practices from the skills-OVATE OJA dataset that was provided for this project.

3.1.5 Data Pre-Processing.

In order to prepare the data for the AI models, the data is restructured such that a datapoint (a row) transfers from being an observed skill with variables (the columns) into an observed OJA with variables. To do this, all datapoints corresponding to the same OJA are collected and stored as variables for that OJA. For example, all skills that were observed in the same OJA are collected and stored as a variable containing the skills for that OJA. This results in a new data structure where each row corresponds to a unique OJA and the columns contain variables about that OJA.

The variables we use for the models are the following:

- **Date of OJA posting** – this is necessary to know when the OJA was posted.
- **Occupation category** (ESCO 3-digit level) – this allows us to select occupations.
- **Skills** (ESCO 1-digit level) – this allows us to see to what extent different ESCO skills categories are demanded.
- **Sustainable company practice (I5.0)**[1] – whether the OJA is classified as a sustainable company practice OJA (an “I5.0 OJA”).
- **Salary level** – Salary might have a relationship with skills or company practices.
- **Required experience level** – experience levels might have a relation with skills or company practices.
- **Economic activity division** (2-digit level) **–** the NACE rev economic activity division might have a relationship with skills or company practices.

If any of these variables is missing for an OJA, we exclude the OJA from the dataset. To create time-series data, we compute aggregates of these features per month. That is, for every month we compute:

[1] An OJA is classified as Industry 5.0 if at least one of the elements in the sustainable company practice dictionary is encountered in the OJA.

- **Proportion skill demand:** for different skill categories of interest, we compute the proportion of OJAs that mention skills in that category.
- **Proportion Industry 5.0**: the proportion of OJAs that are classified as Indsutry5.0 based on the sustainable company practice dictionary.
- **Proportion salary levels:** for every salary range level, we compute the proportion of OJAs that fall within that salary range.
- **Proportion experience levels:** for every experience level, we compute the proportion of OJAs that fall within that experience level.
- **Proportion economic activity division**: for every economic activity division, we compute the proportion of OJAs that fall within that division.

3.2 IER Vacancy Data for the UK.

Unlike the EU's OVATE dataset, the UK job advertisements database includes rich, free-text vacancy descriptions, which offer a more detailed window into the actual practices and technologies used by companies. Since we are interested in identifying Industry 5.0-related practices, having access to richer textual data improves the model's ability to detect and quantify those practices.

3.2.1 Data Collection.

The data analysed in this section is drawn from a bespoke, automated web scraping system developed by the Institute for Employment Research (IER) to support high-frequency labour market monitoring in the UK. This system was first implemented in 2019 as part of the Labour Market Information for All (LMI for All) initiative, in response to the growing need for more timely, granular, and cost-effective data on employer demand. It systematically collects job advertisements from major UK online job portals, capturing vacancies across all economic sectors on a monthly basis. As of mid-2024, the resulting database includes over 8 million job postings. This is the same database used in WP3.

Each job advert in the dataset includes information such as job title, description, date of posting, job portal, and regional location. To enable structured occupational analysis, all job postings are automatically classified into the Standard Occupational Classification (SOC 2020 UK) at the four-digit level using CASCOT, a tool designed to assign occupational codes based on vacancy text. This process ensures consistent classification across time and supports robust segmentation by occupation.

By eliminating reliance on external sources, the IER system ensures greater flexibility and transparency in the collection process. Its regular, real-time updates also provide a significant advantage over traditional survey-based methods, which typically suffer from longer lags and limited frequency.

As a result, this data collection approach makes it possible to observe short-term changes and long-term trends in employer behaviour with a level of detail not available in conventional labour market surveys. It provides a valuable foundation for scenario-based forecasting, policy analysis, and workforce planning.

3.2.2 Data Structure.

The data comprises online vacancies from the UK labour market, covering the period from 2019 to the first half of 2024, with approximately 8 million job vacancies. Each vacancy comprises essential fields[2] such as:

- **Occupational code**: enabling the segmentation of the dataset by occupation based on the Standard Occupational Classification for the UK at 4-digit level. These codes were automatically assigned using CASCOT that supports the consistent coding of job titles and descriptions into SOC categories.
- **Timestamp attributes**: used for time-series aggregation or for creating features indicating temporal patterns in the vacancy data based on month and year of publication.
- **Skills**: a list of ESCO-defined skills explicitly demanded on job vacancies. A list of ESCO-defined skills demanded in job vacancies. These skills were extracted and mapped directly from the job descriptions of the vacancies, allowing for a detailed understanding of the specific competencies employers were seeking.
- **Skills groups**: Clustering of skills to organise specific skills into broader analysis groups. These groupings help simplify the analysis and make it easier to observe patterns across different types of competencies, such as digital, environmental, or management skills.
- **Industry 5.0 practices**: appearing as a list of related to Industry 5.0 practices within company vacancies.
- **Covid-related attributes**: variables indicating whether a job vacancy was posted during the Covid-19 lockdown periods in the UK, allowing for analysis of hiring trends and labour market fluctuations during those months.

3.2.3 Data Cleaning.

To prepare the dataset for training the AI models, the dataset undergoes several filtering steps to ensure consistency and relevance. It is important to note that the nearly 8 million job vacancies represent individual job postings, each linked to a specific employer and job role. While the dataset does not include detailed firm-level information, this vacancy-level granularity remains a key strength, as it enables a more precise understanding of labour market trends based on the characteristics and content of each job advert.

However, to conduct robust modelling, this granular data is subsequently aggregated on a monthly basis by occupation. Not all occupations are included in the final analysis—only those that show a stable presence over time. Specifically, an occupation must appear in at least six different months per year from 2019 to 2023, and in at least three months during the first half of 2024. Moreover, for each of those months, there must be a minimum of 100 job postings to avoid generating insights based on sparse or highly volatile data. As a result of this filtering process, 130 occupations at the 4-digit level were retained and used in the skill prediction exercise reported later in this report.

This filtering process ensures that only occupations with consistent and representative volumes of data are included in the time-series analysis. By doing so, the model avoids making predictions based on irregular patterns and instead focuses on roles with sufficient historical data to support consistent forecasting. Once occupations meeting

[2] There are additional variables in the database but they are not mentioned as they were not part of the analysis.

these criteria are selected, their vacancy data is then aggregated into monthly time-series data at the occupational 4-digit level. For each occupation, this includes:

1. The proportion of vacancies classified as I5.0 in that month[3].
2. The proportion of vacancies mentioning the types of skills categories[4].
3. Additional control variables such as Covid-related economic shocks.

[3] A vacancy is classified as I5.0 if its description includes one or more phrases associated with the Industry 5.0 dictionary developed and validated in the previous Work Package. This classification relies on a detailed analysis of the vacancy text to detect explicit references to I5.0-related practices such as human-centricity, resilience, or sustainability.

[4] The full list of labelled clusters is as follows: Manufacturing and Material Processing Skills, Information Technology and Engineering Skills, Construction, Repair, and Safety Management Skills, Management and Advisory Skills, Healthcare, Veterinary, and Therapeutic Skills, Creative Arts and Performance Skills, Analytical and Research Skills, Transportation, Maritime, and Aviation Operations Skills, Environmental Sustainability and Resource Management Skills, Safety and Security Management Skills, and Machinery Operation and Maintenance Skills. The clustering process involved grouping skills based on their semantic and functional similarity using the K-means algorithm. In a second stage, a Large Language Model (LLM), specifically OpenAI's '4o-mini', was employed to analyse the skill clusters and generate descriptive labels that capture the core characteristics of each group.

4.Methodology.

This Chapter describes a generically applicable method for skills prediction, which we refer to as “Time series-Based Prediction”. We first describe the method in a generic way: the type of ML models that are used, how the models can be used to observe effects of different scenarios, how the models should be trained and evaluated, and the limitations of the method. We then describe how the method is implemented on the OVATE dataset and the Bridges UK dataset. Here, we specify the choices made in the implementation, such as the skills we focused on, which occupation groups were targeted, and which time frame was used in the prediction.

4.1 Generic Description.

4.1.1 Introduction to Method.

In order to achieve insights into future skill demands and requirements related to Industry 5.0 company practices, we utilise a method we refer to as “Time series-Based Prediction”. This method uses AI models that learn how various phenomena in the labour market statistically relate to each other over time. For example, the requirements (or “demand”) of a specific skill may vary over time, but so do other variables, such as the degree to which sustainable company practices are implemented in the market. Other variables may be, for example, salary levels, seniority requirements, specific company-level information or market-level information. By looking at the covariations of all these variables over time, as well as the temporal patterns that exist in these time trends, the AI models may learn how these variables will extend into the future. Furthermore, when the AI models learn how various variables relate to skill demand, the impact of hypothetical future “scenarios” on skill demand can be examined. For example, when the models have learned how the extent to which sustainable company practices are adopted by organisations relates to the demand for a particular skill type, it can be calculated how, according to the models, a specific future adoption of sustainable practices may affect demands for that type of skill. It is important to stress that this method is generic: it may be applied on data from any particular country (or group of countries), for any skill (group) of interest and any market segment or occupation (group) of interest. As with any data-driven method, the primary question is whether the data is available and of sufficient quality.

4.1.2 Models.

The method is comprised of two models that complement each other. A first model, a Long Short-Term Memory network (LSTM), learns to expand multiple time series of different features into the future. A second model, the tree-based model XGBoost, learns to use the LSTM-based future feature values to make a prediction of the future demand of a particular skill of interest.

Features are the variables that are used by the AI models to predict the outcome variable. The outcome variable is the skill category of interest. Hence, the features can consist of any available variable that may somehow be related to the skill of interest. Depending on the interests of the researcher and the availability of data, various features can be used. In the context of Industry 5.0 policy making, these can be, for example, the degree to which organisations have adopted specific Industry 5.0-related company practices in the labour market, amongst others. Figure 1 contains several time series of hypothetical features that might be measured in the labour market for a particular occupation category. The “evolution of green strategy” feature might be a percentage of relevant companies that have explicitly defined and committed to a green strategy. The “evolution of green investments” feature might be the proportion of corporations that invest in sustainable initiatives. The “evolution of salary” could be the evolution of salaries in that industry, and “evolution of seniority” could be the average seniority level as observed in vacancies for that occupation.

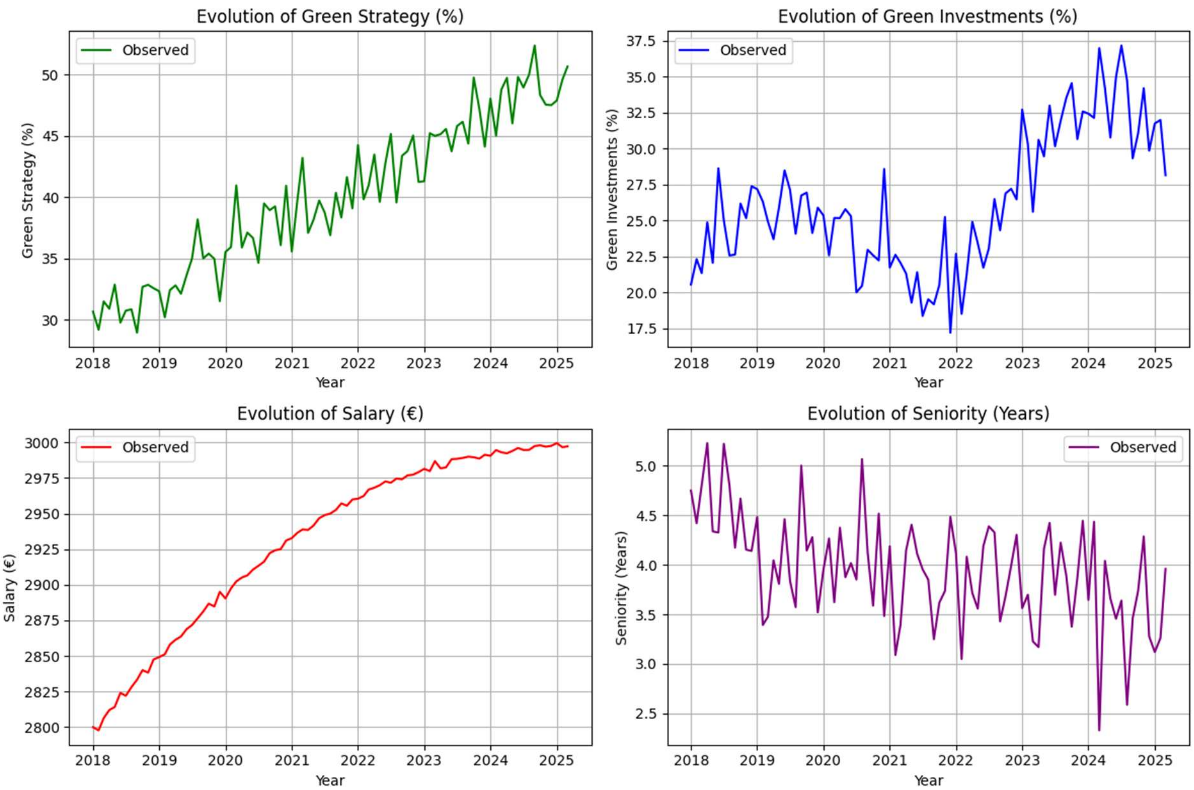

Figure 1. An example of four features that evolve over time. Note: These plots are generated for illustration purposes using mock-up data.

These (and many other) features and time trends can be provided to an LSTM model that learns how these variables behave and how their time trends might be predicted into the future. LSTM models are a class of recurrent neural network (RNN) models specialised in temporal dependencies in data. LSTM models were also suggested in the study by Flemming (2020) for future research in the context of predicting task and skill requirements over time. Figure 2 illustrates how the LSTM might predict how the features mentioned above would behave in the future. Note that an LSTM can capitalise on covariations between these features. For example, in principle, the LSTM is able to learn how green strategies might affect green investments in a later time window, provided such a relationship exists and has occurred in the past (the historic time trend). When having multiple features in the dataset, such covariations become increasingly important.

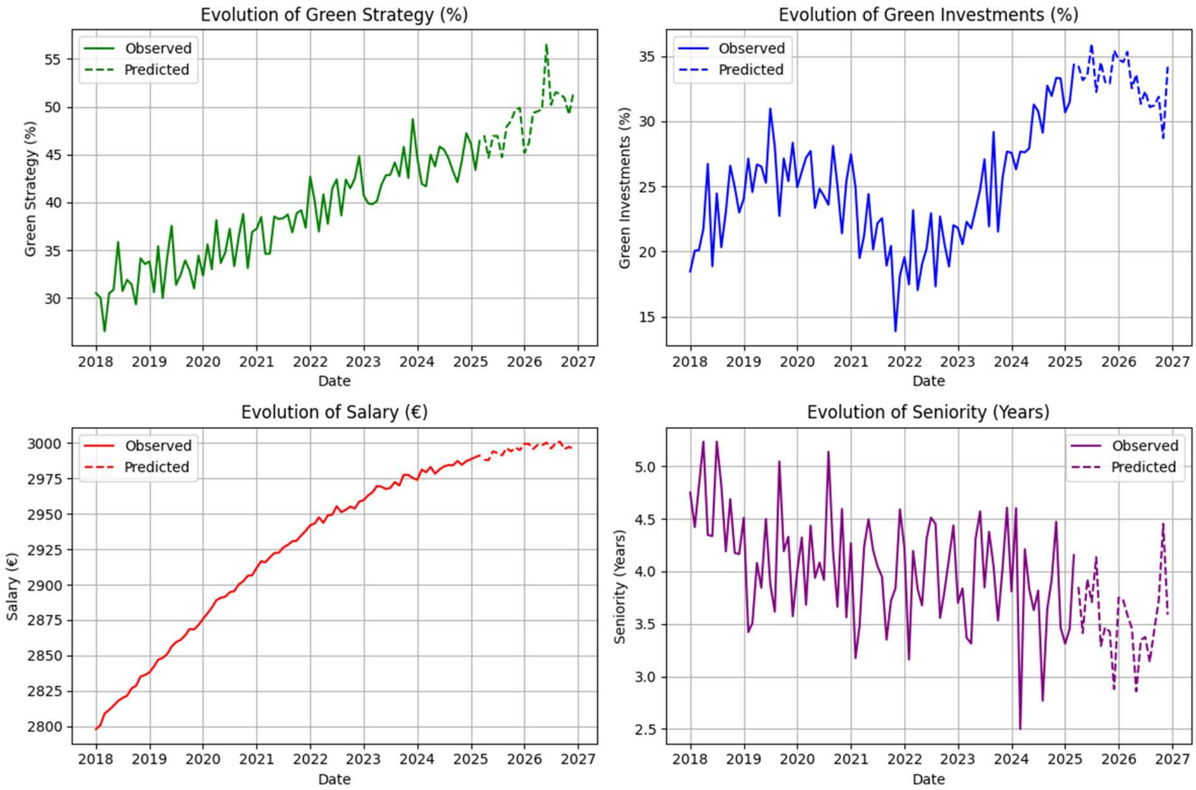

Figure 2. Example of how the features might be extended over time. Note: These plots are generated for illustration purposes using mock-up data.

The output of the LSTM predictions of all the features is then provided to an XGBoost model that uses these outputs and, based on these, makes a prediction of the demand for a particular skill category of interest. XGBoost is a tree-based model that is very well equipped for making predictions based on tabular data. The way in which XGBoost builds on the output of the LSTM to make a skills prediction is illustrated in the Figure below. XGBoost takes the last value of the LSTM's output (in this case the predicted values for beginning 2027) and, based on how the model was trained on historic data of the same features and skill of interest, infers the most probable skill demand for that time point (in this case "50%"). This percentage may, for example, indicate the proportion of job vacancies for which the skill is desired as observed in vacancy texts.

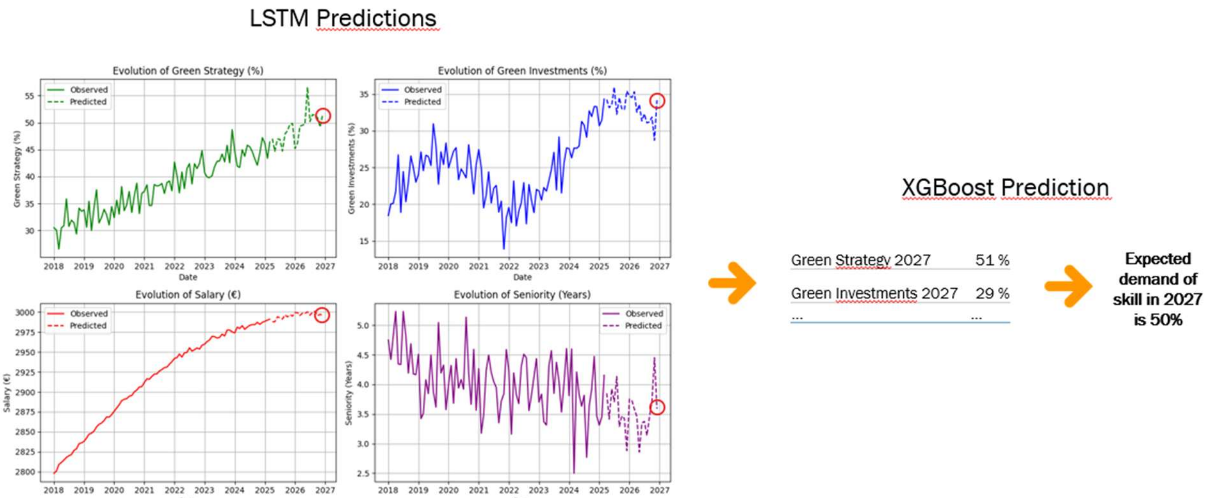

Figure 3. Illustration of how XGBoost uses the output of the LSTM to generate a prediction for the

demand of a specific skill of interest. Note: the data is generated for illustration purposes using mock-up data.

Again, note that the XGBoost algorithm can learn how the *combination* of features relates to skill demand. Utilising the patterns between multiple features becomes increasingly important when an extensive set of features is used in the prediction. Another benefit of XGBoost is that it is straightforward to compute the extent to which the different features contribute to the prediction, which is called “feature importance”. If more extensive explanations are desired, such as the internal “reasoning” of the model regarding a specific prediction, an even more explainable model might be considered instead of XGBoost. An example would a Decision Tree (DT) algorithm. Generally, however, such explainability-by-design comes at a cost in predictive power. Finally, using a tree-based model like XGBoost makes it relatively straightforward to implement “alternative scenarios” and observe how these impact the model’s output. Such alternative scenarios are also a type of explanation of the model, as they show how the model’s output is affected by different inputs. Importantly, these alternative scenarios are especially interesting as a means for insight for Industry 5.0 stakeholders and are discussed in the next section.

4.1.3 How to Use the Models: Different Scenarios.

In case there is data available of an extended period of time containing relevant features (such as Industry 5.0-related company practices) for skills predictions, the Time series-based Prediction method described in this chapter can be used to derive insights in the way these factors, according to the models, evolve over time and how this affects future skill demands. This can be considered the “current scenario” as the models look at the recent past, project the patterns that are observed into the future, and statistically infer how that impacts skill demand. However, as a means for insight, it can be useful to see how skill demand would be impacted by alternative future scenarios. For example, an informative question could be: “How would skill demand be impacted if Industry 5.0-related company practices were to further increase with respect to the current projections?”. Answers to such questions could be helpful in learning the impact of various future scenarios on skill demand. An example is illustrated in the Figure below.

Figure 4. Illustration of how a user can use the models to make inferences about alternative future situations. In this case: How does demand for a specific skill category change according to the models

when certain green policies would further increase in the future? Note: the data and numbers in this figure are mock-up and are solely meant for illustration purposes.

In this example, the initial projections of “green strategy” in 2027 are 51% and “green investments” in 2027 are 29%. The corresponding skill demand, according to XGBoost, is then 50%. However, in case the future projections of green strategy and green investments are higher, in this case 70% and 40%, respectively, the expected skill demand increases to 60%. This can be interesting when stakeholders want to know how the models estimate the effect of a future increase in green policies.
It is important to realise that XGBoost can learn non-linear relationships between features and skill demand. The model may learn that an initial increase of one of the features may statistically relate to the demand for a skill, but only up to a certain point. A further increase of the feature may not have such an impact. In addition, XGBoost may learn how the effect of one feature depends on the effect of another. Imagine, for example, that decreased required seniority levels as observed in vacancy texts are positively associated with the demand for a certain skill category. This relationship, however, may depend on a combination of certain company practices.

4.1.4 How to Train and Evaluate the Models.

Both models should be trained and evaluated on observed historical data. The LSTM model can be fitted to the historical time trends of the different features. XGBoost can be trained using the feature values and target value (i.e., demand of the skill category of interest) at various time points. In order to evaluate the models, they should be applied to the recent past. For example, the LSTM may be provided data up until 2022 and project the time trends up to 2025 (i.e., a three-year prediction). XGBoost is then applied to that projected data between 2022 and 2025 to make predictions of the demand for the skill category of interest. These are then compared to the actual observed demand for that skill category, resulting in an error, such as the mean absolute error (MAE). Furthermore, in order to say anything about whether this error is “high” or “low”, the error should be compared to a simple baseline prediction. A straightforward approach for a baseline prediction is to assume that the future does not change. Hence, this baseline would take the last observed (known) skill demand (in our example, 2022) and assume that it will be the same in the future (in our example, in 2025). Logically, if the model does not improve on such a simple baseline, there is no use in applying it to the given data.

4.1.5 Limitations.

The time series–based prediction method described in this chapter has several limitations that potential users and researchers should take into account. We describe three limitations that are, in our view, most prominent. These are: 1) correlational inferencing; 2) timepoints as learning units instead of vacancies or companies; 3) data requirements.

- **Correlational inferencing.** The first limitation is that all associations that are found in the data are based on correlations. As we know, correlational relationships do not imply causal relationships. Strictly speaking, the only way to observe and learn causal relationships is by adopting controlled scientific experiments. This means that you cannot use these models to see how one variable literally *impacts* another variable. For example, if the model expects that the demand for a specific skill category increases when certain company practices are abandoned, this does not necessarily mean that abandoning such company practices actually leads to such an increase in skill demand. It only suggests that in historical data, the model has found such a statistical relationship. This limitation is true for any data-driven approach, but it is important to bear this in mind. This stresses, again, the need for adopting multiple paradigms, some data-driven, some theoretical.

- **Timepoints as a learning unit instead of vacancies or companies.** Another limitation is that the Time series-Based method does not utilise co-occurrences of features within market instances, such as within vacancies or within companies. By taking *time-points* as a unit of inference (i.e., as "learning unit"), the models capitalise on industrial aggregations (such as averages or totals) over time. For example, the *average* skill requirement at a specific point in time is associated with the average level at which sustainable company practices are implemented at another point in time. The number of datapoints from which the model can learn is therefore constrained by the number of time points at which these averages can be compared to each other. At time point *A, the model receives certain averages, and at time point B,* the model receives other averages. Time-points are not unlimited and also not independent from each other, so the amount of information that can be provided to train the model is considerably constrained. Another approach would be to take *companies* as a learning unit for the AI models. This allows for the incorporation of co-occurrences of features that exist within these units. For example, *within* a specific company, certain features may co-occur (i.e., be observed within the same company). By taking the companies as a learning unit, the number of data points increases up to the number of relevant companies, as opposed to the number of time points. This significantly elevates the ability of the models to learn relationships between the different features. However, given that we do not have company-level information in the skills-OVATE OJA dataset (e.g., it is unknown to which company a specific OJA belongs), it is not possible to use companies as a learning unit in the current project. The Time Series-Based Method we currently adopt is specifically meant for handling time series data and projecting these into the future, and although it does not utilise co-occurrences within company-units, it is an interesting approach to utilise information that is captured in such time trends.
- **Data requirements.** The Time series-Based Prediction approach requires data over an extended period of time. If there is only data available for one or two years, the LSTM model will not be able to learn how the different features relate to each other over time. In addition, projecting time trends into the future assumes there exists periodicity in the data. This means that there should exist some degree of recurring patterns over time in the historical data that allow the model to hypothesise how the features might look in the future. This is a hard requirement on the data, as the labour market is always evolving and therefore always "new" in a way. Certain events, however, might recur, which could induce periodicity elements in the data. It should be noted that periodicity is always a requirement for data-driven projections into the future. It is therefore a feature of the data-driven paradigm and not only of the way it is implemented in the current methodology. In addition, the relationships between the different features at any given time point, as learned by the models, can still be valuable and informative. Experimenting with different future scenarios and observing how the model reasons with these scenarios in terms of how these may affect skill demand, can therefore still be useful even though the future as it will play out now will be hard to predict.

4.2 Implementation on skill-OVATE OJA Dataset.

This section describes how the Time series-Based Prediction method is implemented on the skills-OVATE OJA dataset.

4.2.1 Selected occupations.

Skills predictions can be performed for various occupation categories. To perform a proof of concept and see whether we can utilise the Time series-Based Prediction method on the skills-OVATE dataset we received, we select an occupation group at 1-digit level from each of the nine major ESCO occupation classes and perform predictions for these occupations. For every occupation group, we need sufficient relevant OJAS, so the occupation categories should not be too narrow. On the other hand, to generate interpretable insights, the occupation categories should not be too broad. We therefore choose ESCO occupation 3-digit code level, which provides a balance between both objectives. We focused on the following 3-digit occupation groups for which we perform skills predictions, namely: "Sales marketing and development managers", “Physical and engineering science technicians”, “General office clerks”, “Cashiers and ticket clerks”, “Market gardeners and crop growers”, "Transport and storage labourers", "Machinery mechanics and repairers", "Engineering professionals", "Heavy truck and bus drivers". Further, we initially focus on OJAs from the UK. If our evaluation shows we can generate models with predictive capacities, we will extend to other European countries.

4.2.2 Selected skills.

The skills categories of interest are the six major ESCO transversal skills categories, namely: “core skills and competences”, “thinking skills and competences”, “self-management skills and competences”, “social and communication skills and competences”, “physical and manual skills and competences”, and “life skills and competences”. These are major ESCO skills that are present in the skills-OVATE data and can therefore be used as a target for predictions. ESCO designed the transversal skills classification to define skills that are valuable and relevant across a wide range of industries and occupational classes. It is interesting to observe whether the demand for different types of transversal skills relates to sustainable company practices.

4.2.3 Training and evaluation.

An LSTM model is used to extend the time series features one year into the future. The features are Proportion Industry 5.0, Proportion salary levels, Proportion experience levels, and Proportion economic activity division. Based on these predicted time series, an XGBoost model predicts the demand for the skill category of interest.

As the relationship between features such as company practices and skill demand may vary across occupation categories and across different types of skills, we train and evaluate separate prediction models for each of the specified occupation groups and skill categories of interest.

Models are trained based on monthly data between January 2018 and June 2023 and tested on data between Jul 2023 and Jun 2024. Performance testing occurs by computing the mean absolute error (MAE) between the predicted skill demand and the actual skill demand at each time point (month). To evaluate the resulting MAE, we compare the models to a simple baseline method which assumes that the skill demand remains constant in the future. If MAE decreases with respect to this baseline, we know that the method has predictive capacity.

4.3 Implementation on Bridges UK Vacancies Dataset.

This section describes how the Time series-Based Prediction method is implemented on the Bridges UK Vacancies dataset.

4.3.1 Modelling Industry 5.0 adoption in job vacancies.

Each vacancy is labelled as “I5.0” if it includes specific Industry 5.0 practices mentioned in job vacancies. These labels are aggregated each month to determine what fraction of vacancies in a given occupation relate to I5.0. By converting everything into monthly proportions, the model reduces day-to-day noise and focuses on overall trends.

An LSTM network is trained on monthly data from 2019 to March 2023, learning how I5.0 adoption rates have evolved over time. The model is then tested on a holdout period from April 2023 to June 2024, during which the predicted adoption rates are compared against observed data. Results from this test phase show that, despite relatively low I5.0 adoption levels overall, the model captures the key upward trends with acceptable error margins. Based on this, a one-year forecast is produced for July 2024 to June 2025. The decision to limit the forecast horizon to a single year is deliberate: references to Industry 5.0 practices in job advertisements remain relatively recent, as many of these concepts—such as human-centric design, resilience, or sustainable automation—are still emerging within the UK labour market. As a result, historical data offers only a short window from which to learn stable patterns. Extending forecasts beyond one year under these conditions could introduce considerable uncertainty, as the trends are not yet sufficiently established to support reliable long-term projections. As I5.0 adoption becomes more widespread and consistently recorded in job advertisements, future versions of the model may incorporate longer forecast windows with greater confidence.

4.3.2 Estimating skill demand in response to Industry 5.0 trends.

Alongside Industry 5.0 adoption, the model also tracks how often different skill categories appear in vacancies. Each skill mentioned is grouped into one of twelve broad categories, such as “Information Technology and Engineering Skills”, “Manufacturing and Material Processing Skills,” and “Healthcare, Veterinary, and Therapeutic Skills”. By looking at the fraction of vacancies requiring each category every month, the model captures broad changes in skill demand over time.

To see how skill demand relates to changes in I5.0 by occupation, an XGBoost regression is trained to predict the monthly share of its vacancies mentioning each skill category, using:

1. The LSTM-based I5.0 forecasts (i.e. how much I5.0 adoption is expected each month for the occupation).
2. Additional time-based features, including simple indicators for lockdowns and ongoing Covid effects (although there are none in later months).

By linking I5.0 adoption directly to these skill proportions, the model helps us understand which skill areas might expand or contract as I5.0 becomes more widespread.

The XGBoost model is trained to estimate the monthly proportion of vacancies that mention each skill category, at the 4-digit occupation level. It does so by using the output of the LSTM model—namely, the predicted monthly share of vacancies related to Industry 5.0—as a key input. This setup allows the model to assess how expected changes in I5.0 adoption may influence the demand for different types of skills over time, such as whether an increase in I5.0 practices is associated with a greater share of vacancies mentioning “Management and Advisory Skills” or “Information Technology and Engineering Skills.” Time-specific contextual variables, such as lockdown periods, are also included to help isolate the effects of I5.0 from other external factors.

At this stage, the analysis focuses on the overall presence of I5.0 practices in vacancy texts, rather than distinguishing between their subgroups (e.g., human-centricity, sustainability, resilience). This decision reflects the exploratory nature and proof of concept of the model and the relatively recent appearance of these practices in job advertisements. When tested, time-series analysis at the level of individual I5.0 dimensions proved too volatile to produce consistent results, making it difficult to assess the robustness of the modelling approach.

4.3.3 Scenario simulations.

After training the LSTM and XGBoost models, the model creates two “what if” scenarios where I5.0 adoption from April 2023 onwards is increased by 10% (1.1 times) or 30% (1.3 times) compared to the model’s original projections. After, the model uses these boosted I5.0 forecasts into the XGBoost model to see how skill demand might change if I5.0 grows more quickly than expected.

For example, suppose that in a given month, 20% of vacancies in an occupation are projected to be related to I5.0. In the scenario where I5.0 adoption is increased by 10%, this proportion becomes 22% (i.e. 10% more than 20%). In the 30% scenario, it rises to 26%. These adjusted values are then used by the XGBoost model to estimate how such increases in I5.0 adoption could affect the demand for different skill groups. By running these simulations, the model provides insights into how the labour market might respond if Industry 5.0 spreads faster than expected.

4.3.4 Changes in skill demand and model robustness.

Once the simulated scenarios of increased I5.0 adoption are defined, the next step involves evaluating their potential impact on skill demand. This is done by comparing the projected share of vacancies referencing each skill category under the baseline and the two simulated scenarios. The goal is to determine whether accelerated uptake of I5.0 practices is associated with measurable changes in the types of skills employers are likely to seek. To ensure that these results are meaningful and not driven by noise, the analysis also considers the model’s prediction error margins. This allows for a more confident interpretation of which skill shifts may reflect genuine underlying trends and which may fall within normal variation.

To understand the model’s outputs, it is important to compare the predicted percentage of vacancies by occupation, mentioning each skill category in:

1. A baseline scenario (using the LSTM’s original I5.0 forecasts).
2. The 10% and 30% scenarios (where we artificially increase future I5.0 adoption).

If, under the baseline, 10% of vacancies are expected to require a particular skill category, but that figure rises to 13% when I5.0 adoption is boosted by 30%, the model reports a +3 percentage point difference. Because both the LSTM and XGBoost predictions have error margins, the model also tracks performance measures like mean squared error (MSE) for the LSTM and mean absolute error (MAE) for the

XGBoost. If these errors are small, there is more confidence that a 3-point change is meaningful rather than random variation.
As a result, this combined LSTM + XGBoost approach seeks to estimate how I5.0 practices might develop and how these developments could shift the skills that employers demand. By testing future scenarios of faster I5.0 expansion, the model highlights which skill areas may become increasingly important, helping inform targeted interventions for training, education, or policy.

5 Results.

This Chapter describes the results of applying the models on the skills-OVATE OJA dataset and the Bridges UK OJA dataset.

Take-aways
The results show that the models are able to predict skill demand when applied to the Bridges UK OJA dataset, but not when applied to the skills-OVATE OJA dataset.
According to the predictions based on the Bridges UK dataset, the models expect for specific occupation categories that certain skill types increase or decrease during the next year.
Further, different future scenario's in which the proportion of Industry 5.0-related OJAs increase with 10% or 30% show that the models sometimes expect that an increase in Industry 5.0 practices affects skill demand for certain occupations.

5.1 Skills-OVATE OJA Dataset.

The results from applying the models to the skills-OVATE OJA dataset show that they do not significantly outperform the baseline approach, which assumes that skill demand does not change in the future. This is evinced by the average MAE of the predictions (+- 3.5%) being close to the average MAE of the baseline (+- 3.6%). This shows that the models are not able to generate meaningful predictions of skill demand when applied to the skills-OVATE OJA dataset. Given that the models do not have predictive power, it is not meaningful to use them to experiment with alternative scenarios, such as “what if more organisations would adopt sustainable I5.0 company practices?”. Such alternative scenarios are therefore not implemented based on the skills-OVATE OJA dataset.

5.2 Bridges UK OJA Dataset.

This section presents the outcomes of the forecasting exercise on the Bridges UK OJA dataset. In short, the results suggest that the models do have predictive power and can generate useful insights when applied to this dataset. Based on the predictions, the models are used to see how simulated increases in Industry 5.0 (I5.0) adoption may influence skill demand across a wide range of occupations at 4-digit level in the UK. Using scenario simulations based on the combined LSTM and XGBoost model, the analysis explores potential changes in the proportion of job vacancies on occupations referencing different skill categories under varying levels of I5.0 uptake. The goal is to identify which occupations and skill areas are most responsive to hypothetical shifts in I5.0 practices, and to what extent these changes exceed normal variation. The results offer an evidence base for anticipating skill needs in a more technology-driven, human-centric, and resilient labour market context.

5.2.1 Does more I5.0 mean more skills? Initial results and stable occupational changes.

Understanding whether increases in Industry 5.0 adoption lead to measurable changes in skill demand is central to anticipating how occupations might evolve. Forecasts are generated under three I5.0 scenarios to explore how different levels of adoption affect the share of vacancies referencing key skill categories. The goal is to

identify patterns of change that are consistent, occupation-specific, and significant enough to inform workforce or training strategies.
This analysis covers 130 different occupations under three Industry 5.0 (I5.0) scenarios:
- Baseline
- 1.1 (10% increase in I5.0)
- 1.3 (30% increase in I5.0)

By adopting the scenario-based perspective, the analysis offers decision makers a clearer sense of which skill groups are more likely to respond—sometimes modestly—to enhanced I5.0 activity. While this approach provides numeric estimates of future demands, it also underscores inherent uncertainties. The underlying correlations may not capture disruptive policy or technological shifts outside the scope of the historical dataset. Nonetheless, these scenarios function as realistic "what if" exercises, demonstrating how skill usage might vary across plausible degrees of I5.0 expansion.
The analysis includes 130 occupations. For each occupation, twelve skill categories are evaluated under three scenarios: a baseline and two hypothetical increases in Industry 5.0 adoption (by 10% and 30%). This results in 4,680 scenario combinations. Of these, 3,120 involve changes compared to the baseline. In many cases, the model does not predict major shifts in skill demand when Industry 5.0 adoption is increased. While it may be tempting to attribute this to the short one-year forecasting window, it is important to note that the model simulates relatively large increases in I5.0 adoption—by 10% and 30%—within that period for each occupation. These simulated jumps are designed to stress-test the relationship between I5.0 practices and skill demand. The fact that most occupation–skill combinations show minimal or no increase suggests that, at present, the link between I5.0 and certain skill categories is either weak or delayed. The table below summarises the distribution of projected changes (in percentage points) for the 3,120 scenario outcomes beyond the baseline derived from the combined LSTM+XGBoost model.

Table 1. Model's estimated percentage points growth for occupations at the 4-digit level

Describe statistics	Values
Number of scenarios	3,120
Mean of Estimated Percentual Points Growth (%)	0.351
Standard Deviation (std)	2.039
Minimum Estimated Percentual Points Growth (%)	-21.138
25th Percentile (Q1) (%)	-0.012
Median Estimated Percentual Points Growth (%)	0.034
75th Percentile (Q3) (%)	0.389
Maximum Estimated Percentual Points Growth (%)	26.38

Source: own elaboration based on the model's results

These figures indicate that the average projected change in skill demand across all occupation–skill pairs is slightly positive (0.35 points). Some scenarios do suggest moderate or even larger increases, with the maximum reaching over 26 percentage points. However, many shifts remain near zero, and certain occupations exhibit negative changes of over -20 percentage points. This indicates that a higher prevalence of I5.0 practices does not automatically raise demand for all types of skill categories.

Even among occupations where the model initially estimates a skill-demand increase of at least one percentage point, further checks show that many of these projected gains fall within the model's forecast error range, making them statistically uncertain. To provide a clearer picture of the most meaningful changes, the next table highlights a subset of occupation–skill combinations that stand out for their relatively strong and stable increases under higher I5.0 adoption scenarios.

Table 2 shows the 10 most important positive and stable changes in skill group demand over a simulation of increasing I5.0 practices between July 2024 and June 2025. Each row corresponds to an occupation–skill pair under a simulated scenario where the baseline proportion of I5.0 (scenario factor 1.0) is compared against a 10% or 30% boost (scenario factors 1.1 or 1.3). These simulated increases are not based on observed trends but are introduced deliberately to test how sensitive skill demand is to higher levels of I5.0 adoption. Since this is an exploratory model and references to I5.0 remain relatively new in vacancy texts, the simulation focuses on overall adoption rather than separating out individual dimensions like sustainability or human-centricity.

The first column, Occupation, identifies the four-digit occupational group based on the UK classification. Next, Skill Category shows the broad skill group (e.g. "Management and Advisory Skills") that the model tracks within job postings. The Increase Scenario indicates whether the forecast is multiplied by 1.1 (10% increase) or 1.3 (30% increase) for mentions of Industry 5.0 practices within job advertisements for occupations. Under Predicted Future Demand, the table presents the average percentage of vacancies expected to require that skill category over the next 12 months when the I5.0 scenario is applied. The Additional Demand vs. Baseline column is the key indicator. It shows the number of additional percentage points by which demand for each skill group is expected to increase if Industry 5.0 practices become more widespread. A positive figure means that a 10% or 30% increase in I5.0 adoption would lead to a higher share of job vacancies requiring that specific skill group. The Model's LSTM Error reports how much the Industry 5.0 forecasting module (LSTM) might deviate from observed trends, in mean squared error (MSE). Meanwhile, the XGBoost MAE column captures how far the model's skill-demand predictions typically are from real data, converted to percentage points for easier interpretation.

After applying a stricter reliability threshold on the results from both LSTM and XGBoost models, only a small set of about 15 occupations shows a clearly positive and dependable response to higher I5.0 adoption. In these cases, the changes tend to concentrate in specific skill categories such as management and advisory skills, multilingual and community engagement abilities, and competencies linked to sustainability or safety. These shifts suggest that, as I5.0 practices expand, certain roles may increasingly require broader organisational, interpersonal, or environmentally oriented skills. Nevertheless, the results should be interpreted alongside other considerations, particularly each model's error margins, to determine which shifts are both large enough and consistent enough to guide policy decisions. While most occupations appear relatively unaffected, certain roles do stand out for notable, consistent gains in specific skill categories, signalling where future growth or training investments might be most impactful.

Table 2. Predicted Skill Demand Increases by Occupation Under Industry 5.0 Adoption Scenarios

Occupation at 4-digit level	Skill Category	Increase Scenario of I5.0	Predicted Future Demand	Demand Above Baseline	Model's LSTM Error	XGBoost MAE (%)
Engineering professionals N.E.C.	Multilingual and Community Engagement Skills	+30%	14.60%	5.31%	0.005	1.95%
Teaching professionals N.E.C.	Safety and Security Management Skills	+30%	15.95%	7.86%	0.0012	1.35%
Actuaries, economists and statisticians	Healthcare, Veterinary, and Therapeutic Skills	+30%	9.84%	5.76%	0.0041	3.06%
Business sales executives	Management and Advisory Skills	+30%	36.42%	9.04%	0.0049	6.81%
Other vocational and industrial trainers	Management and Advisory Skills	+30%	26.33%	16.86%	0.0006	3.25%
Care Workers And Home Carers	Multilingual and Community Engagement Skills	+10%	15.49%	9.45%	0.0024	2.75%
Care Workers And Home Carers	Multilingual and Community Engagement Skills	+30%	17.15%	11.11%	0.0024	2.75%
Routine inspectors and testers	Information Technology and Engineering Skills	+30%	19.81%	6.82%	0.0034	5.18%
Cleaners and domestics	Safety and Security Management Skills	+30%	15.99%	12.18%	0.0015	8.08%
Sales-related occupations N.E.C.	Environmental Sustainability and Resource Management Skills	+30%	6.18%	3.20%	0.0049	2.07%
Financial administrative occupations N.E.C.	Management and Advisory Skills	+30%	30.73%	4.91%	0.0048	1.57%

Source: own elaboration based on the model's results

Between the forecasting for July 2024 and June 2025, Business Sales Executives focusing on Management and Advisory Skills could see their demand climb by about nine percentage points if I5.0 practices increase by 30%. This jump exceeds the typical error margins from the LSTM and XGBoost models, suggesting a relatively consistent prediction that sales roles may require stronger leadership and organisational expertise when higher levels of I5.0 are in play.

In a similar vein, Other Vocational and Industrial Trainers might experience nearly a 17-percentage-point rise in Management and Advisory Skills under a 30% boost in I5.0. Within the July 2024 to June 2025 forecast window, this increase reflects the potential impact of I5.0-driven changes in manufacturing. As factories and production lines integrate advanced robotics or data-driven processes, trainers may need to develop new curricula and guidance, and this sizeable shift also stands well above the model's error thresholds, providing confidence that such a scenario is plausible.

Another example appears in Sales Related Occupations N.E.C., where a 30% I5.0 scenario is linked to more than a three-percentage-point gain in Environmental Sustainability and Resource Management Skills. This result, which also sits beyond the model's margin of error, indicates that if I5.0 becomes more common during the July 2024 to June 2025 period, sales-related roles might increasingly emphasise eco-conscious practices, supply-chain transparency, or sustainable resource planning.

Care Workers and Home Carers show moderate but still meaningful improvements of around nine to eleven percentage points in Multilingual and Community Engagement Skills when I5.0 adoption rises by 10% or 30%. By surpassing typical forecast uncertainty, these increments suggest that more inclusive, technology-oriented care models may benefit from stronger communication and cultural understanding skills.

These results suggest that the impact of Industry 5.0 adoption on skill demand is not limited to traditionally high-skilled or technical roles. While some of the most pronounced increases are observed in professional occupations—such as engineers, educators, or analysts—other roles at medium and lower skill levels, including care workers, cleaners, and sales assistants, also show consistent growth in specific skill areas. This indicates that the influence of I5.0 extends across the occupational spectrum, particularly where tasks involve human interaction, procedural adaptation, or shifts in service delivery models.

The types of skills that increase most prominently under these scenarios reflect the multidimensional nature of I5.0. Beyond technical or digital skills, the simulations point to greater demand for leadership and advisory capabilities, socioemotional and intercultural communication, and competencies linked to safety and sustainability. This reinforces the idea that I5.0 is not solely about technological transformation, but also about elevating the human, inclusive, and environmental dimensions of work. As such, policy interventions and training efforts aimed at preparing for I5.0 should consider both technical upskilling and the development of complex human capabilities across all levels of the labour market.

As the methodology is further refined, integrating more detailed employer-level data and additional indicators of I5.0 implementation could enhance its ability to capture nuanced variations in skill demand across different occupational contexts.

6 Discussion.

Take-aways

- The results indicate a proof-of-concept that AI-based modeling can generate insights into future skill demand and how this relates to Industry 5.0 practices.
- The results, however, also highlight a need for more detailed LMI that contains data on the organisational context in which skills are demanded.
- Further, the Time series-Based Prediction method applied in this report has several limitations that can be circumvented by changing the learning unit of the model fitting algorithms from timepoints to OJAs or companies.
- As soon as more detailed data on organisations is available, future work should therefore focus on deriving further insights by applying AI-based methods incorporating organisations as learning unit on extended datasets.

Applying the Time series-Based prediction method on this dataset, we do find that the models have predictive power for future skill demand. This is particularly evident in a range of occupations where the model identifies shifts in specific skill categories under increased I5.0 adoption. For example, greater emphasis on management and advisory skills is projected for Business Sales Executives, while sustainability-related competencies emerge more strongly in sales-related roles. These patterns suggest that predictive modelling can capture how different occupations may respond to evolving organisational practices, even when those changes are unevenly distributed across the labour market. This is in line with findings by Macedo et al. (2022) and Das et al. (2020), who also adopted LSTM models to generate skills predictions.

Our results, however, indicate a clear need of company-level information. Although we attempted to derive some information on the company level in the skills-OVATE dataset by translating some of the identified ESCO skills into potential indicators of sustainable practices of the company behind the OJA, this translation is at best a weak proxy for that purpose. The results show that the models were not able to learn how the extracted company-level information would behave over time, let alone how it could be used to predict skill demand over time. The Bridges UK OJA dataset, on the other hand, provides richer information on organisational contexts, primarily because it includes full job descriptions—a key variable not available in the Skills-OVATE OJA dataset. This allowed for a more accurate identification of Industry 5.0 practices as explicitly referenced by employers in their vacancy advertisements. Applying the Time series-Based prediction method on this dataset, we do find that the models have predictive power of future skill demand. In addition, the experiments with alternative scenarios, namely an increased future realisation of Industry 5.0 company practices, show that, for some occupation and skill categories, the models expect increased or decreased skill demands. On the one hand, these alternative scenarios provide explanations by example of the AI models as they highlight how the model's outputs are affected by different inputs. On the other hand, they also show how future skill demand is, according to the models, affected by different degrees in which companies adopt Industry 5.0 practices. This indicates that predictive modelling can provide useful insights for researchers and other stakeholders interested in how alterations in future company practices might relate to skill demand changes.

As detailed in section 4.1.5, the Time series-Based Prediction method described and applied in this current report has several limitations. An important limitation is that timepoints are used as a learning unit for the AI algorithms, which significantly impacts the number of datapoints the models can learn from as well as the ability to learn associations between variables. An alternative approach is to take vacancies, or even better, *companies*, as a learning unit. This makes it easier for the model to learn from co-occurrences of variables that exist within vacancy or company instances. Namely, any given company has specific characteristics. By taking companies as a learning unit, as opposed to industrial aggregates of multiple companies over time, the idiosyncrasies of companies can be better recognised by the algorithms, and the number of data points increases up to the number of companies (relevant), as opposed to the number of time points. This significantly elevates the ability of the models to learn relationships between the different features. Furthermore, it is interesting to note that when taking companies as a learning unit, the algorithms might still learn time effects. This can be achieved by defining time-related features and providing these to the ML model. For example, it is possible to define features that represent how a company may change over time. These can be included as features about the company and can then be used by the model to learn how alterations over time affect (future) skill demand. In the current project, however, we did not have company-level information in the skills-OVATE OJA dataset (e.g., it is unknown to which company a specific OJA belongs, and other organisational information was not provided), and it was therefore not possible to use companies as a learning unit. Taking companies as a learning unit would provide a promising approach, and future work should therefore be undertaken to improve datasets by incorporating company-level information and taking companies as a learning unit for the AI algorithms.

While the findings from this forecasting exercise provide initial evidence on how skill demand may evolve under Industry 5.0 conditions, they represent only part of the broader effort to understand how these transformations may unfold. The current analysis relies on vacancy data and simulated scenarios to estimate potential future trends; however, certain limitations, such as the novelty of I5.0-related practices, the lack of detailed company-level indicators, and the challenges in capturing complex organisational behaviours, highlight the need for complementary approaches.

To address this, further foresight and indicator-based research will be carried out under WP4. This next phase will integrate expert interviews with a diverse set of stakeholders across technologies, sectors, and countries, aiming to identify key obstacles to digital and automation integration, as well as the human and institutional challenges associated with the transition to Industry 5.0. The combination of quantitative modelling and qualitative foresight will support the development of realistic future scenarios and policy indicators. Together, these insights will help guide skilling and educational strategies designed to ensure that labour markets remain inclusive, adaptive, and aligned with the evolving demands of Industry 5.0.

7 Conclusion and Future Work.

In line with prior research, this report shows that ML-based methods can be used to predict future skill demand and how this is affected by contextual information; in our case, company practices related to Industry 5.0. Based on our findings, we draw three main conclusions corresponding to our research questions.

How can an AI-based methodology be designed and implemented to forecast future skill demands and to analyse their relationship with practices characteristic of Industry 5.0 organisations?

We built a proof-of-concept of a time series-based prediction method to do this forecast. Our results show that the time series–based prediction method presented in this report is capable of generating insights into future skill needs and their relationship with organisational practices tied to Industry 5.0. The trained models can signal which skill categories are likely to become more (or less) important over time. Furthermore, by implementing alternative scenarios in which the adoption of Industry 5.0 practices increases, the models can simulate how such shifts might influence skill demands across different occupational groups. This approach offers a scenario-based lens for anticipating labour market dynamics. The one-year prediction time horizon that was adopted, however, is considerably narrow. Once better datasets are available, multi-year analyses should be carried out.

To what extent can the predictions of future skill demand produced by the AI model be explained?

The alternative future scenarios provide a type of “what-if” explanation of the models that indicate how the output is affected by alternating inputs. If further explainability is desired, researchers can look at feature importance, indicating how much effect different features have on the output. Additionally, in cases where the number of input features is limited, more interpretable algorithms, such as Decision Trees, can be used instead of XGBoost. Such models offer the benefit of fully traceable internal reasoning, thereby enhancing explainability.

What are the methodological and data-related limitations of the proposed approaches, and how can these be addressed or mitigated in future applications?

Constraints exist both on a methodological and data level. On a **methodological level**, any data-driven approach needs to be complemented with theoretical and qualitative methodologies. In addition, as discussed in this report, the Time series-Based Prediction method has several methodological limitations that we expect can be largely mitigated by altering the learning unit from timepoints to companies. In the current skills-OVATE OJA dataset, this was not possible as there was no company-level information available. Future work, however, should focus on extending datasets with company-level information and adopting an ML-based approach in which the learning unit is changed to individual companies, as opposed to aggregates of multiple companies over time. This brings us to the primary data-related limitation that needs to be addressed in the future.

On a **data level**, the major constraint of currently available datasets is the lack of company-level information. In order to learn how skill demands depend on or relate to Industry 5.0, we need data on Industry 5.0 practices, which exist at the company level. The features extracted in the skills-OVATE OJA dataset do not capture such information. It contains information on which skills are observed in which vacancy, but the company behind the vacancy and the characteristics it has are unknown. It therefore remains unknown *why* skills are required in the vacancy. The Bridges UK OJA dataset described and used in this report was specifically developed with the purpose of containing more contextual information. Currently, however, the Bridges UK OJA dataset only contains information extracted from vacancies, but it may very well be extended with other information sources, such as company descriptions, evaluations, sustainability ratings or the type of technology that is adopted. All such information may be relevant to the skills that are required by the employees.

Based on these conclusions, we suggest extending skills-OVATE OJA data to capture company-level information next to information on skill demand. In order to fully meet researchers' needs, the way in which these datasets are generated should be well documented. Once these datasets are available, the ML-based methods discussed in this report can be applied to show how these factors relate to each other and provide insight into various future scenarios: what happens to skill demand if more companies adopt Industry 5.0 policies in the future?
Here, we suggest implementing ML algorithms that use companies as a learning unit, as we expect these are well-equipped for capturing the various relationships between organisational context and skill demand. If desired, time aspects can be included as features and provided to these models, potentially allowing for insights into how evolving organisational practices influence skill demand.

References.

Autor, D. H., Levy, F., & Murnane, R. J. (2003). The skill content of recent technological change: An empirical exploration. The Quarterly Journal of Economics, 118(4), 1279–1333. https://doi.org/10.1162/003355303322552801

Barnes, S. A., Hogarth, T., Wright, S. A., & Cárdenas-Rubio, J. (2023). Labour market information and an assessment of its applications: a series of international case studies.

Beręsewicz, M., & Pater, R. (2021). *Inferring job vacancies from online job advertisements*. Publications Office of the European Union.

Biran, O., & Cotton, C. (2017, August). Explanation and justification in machine learning: A survey. In *IJCAI-17 workshop on explainable AI (XAI)* (Vol. 8, No. 1, pp. 8-13).

Boselli, R., Cesarini, M., Mercorio, F., & Mezzanzanica, M. (2017a). An AI planning system for data cleaning. In *Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2017, Skopje, Macedonia, September 18–22, 2017, Proceedings, Part III 10* (pp. 349-353). Springer International Publishing.

Boselli, R., Cesarini, M., Mercorio, F., & Mezzanzanica, M. (2017b). Labour Market Intelligence for Supporting Decision Making. In *SEBD* (p. 74).

Brynjolfsson, E., Mitchell, T., & Rock, D. (2018). What can machines learn, and what does it mean for occupations and the economy? AEA Papers and Proceedings, 108, 43–47. https://doi.org/10.1257/pandp.20181019

Cedefop. (2014). Tender: *Real-time labour market information on skill requirements: Feasibility study*. European Centre for the Development of Vocational Training. Retrieved March 27, 2025, from https://www.cedefop.europa.eu/en/about-cedefop/public-procurement/real-time-labour-market-information-skill-requirements-feasibilityreate

Cedefop. (2021). *Skills-OVATE: Skills online vacancy analysis tool for Europe*. https://www.cedefop.europa.eu/en/tools/skills-online-vacancies

Cedefop. (2024). *Next generation skills intelligence for more learning and better matching: Skills anticipation trends, opportunities and challenges in EU Member States*. Luxembourg: Publications Office. https://doi.org/10.2801/180485 ISBN: 978-92-896-3686-5

Cedefop; European Commission; ETF et al. (2021). *Perspectives on policy and practice: tapping into the potential of big data for skills policy.* Luxembourg: Publications Office. https://data.europa.eu/doi/10.2801/25160

Colace, F., De Santo, M., Lombardi, M., Mercorio, F., Mezzanzanica, M., & Pascale, F. (2019). Towards labour market intelligence through topic modelling.

Colombo, E., Mercorio, F., & Mezzanzanica, M. (2018, February). *Applying machine learning tools on web vacancies for labour market and skill analysis* [Unpublished manuscript]. Università Cattolica del Sacro Cuore and CRISP

Eloundou, T., Manning, S., Mishkin, P., & Rock, D. (2024). GPTs are GPTs: Labor market impact potential of LLMs. *Science*, *384*(6702), 1306-1308.

European Commission. (2020). *A new industrial strategy for Europe* (COM(2020) 102 final). https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52020DC0102

European Commission. DG RES (2021). *Industry 5.0: Towards a sustainable, human-centric and resilient European industry* (R&I Paper Series Policy Brief). Publications Office of the European Union. https://doi.org/10.2777/308407

European Commission. DG RES (2022). *Industry 5.0: A transformative vision for Europe* (ESIR Policy Brief No. 3). Publications Office of the European Union. https://doi.org/10.2777/17322

European Ministers of Vocational Education and Training, European Social Partners, & European Commission. (2004, December 14). *Maastricht Communiqué on the future* *priorities of enhanced European cooperation in vocational education and training (VET).* https://www.cedefop.europa.eu/files/communique_maastricht_priorities_vet.pdf

Felten, E. W., Raj, M., & Seamans, R. (2018, May). A method to link advances in artificial intelligence to occupational abilities. In *AEA Papers and proceedings* (Vol. 108, pp. 54-57). 2014 Broadway, Suite 305, Nashville, TN 37203: American Economic Association.

Gmyrek, P., Winkler, H., & Garganta, S. (2024). *Buffer or bottleneck? Employment exposure to generative AI and the digital divide in Latin America* (No. 121). ILO Working Paper.

Hajba, G. L. (2018). Website Scraping with Python. *Berkeley: Apress*.

International Labour Organization, ILO. (2016). Anticipating and matching skills and jobs: Guidance note. https://www.ilo.org/sites/default/files/wcmsp5/groups/public/@ed_emp/@ifp_skills/documents/publication/wcms_534307.pdf

Giabelli, A., Mercorio, F., Mezzanzanica, M., & Trentini, F. (2022, September). Capturing the Skill Transformation Using Big Data and Artificial Intelligence: a Novel Change Index to Observe how the Italian Labour Market has been Changing over Time. In *The Relevance of Artificial Intelligence in the Digital and Green Transformation of Regional and Local Labour Markets Across Europe* (pp. 107-128). Nomos Verlagsgesellschaft mbH & Co. KG.

Katz, L. F., & Murphy, K. M. (1992). Changes in relative wages, 1963–1987: Supply and demand factors. The Quarterly Journal of Economics, 107(1), 35–78. https://doi.org/10.2307/2118323

Lipton, Z. C. (2018). "The Mythos of Model Interpretability." *Communications of the ACM***,** 61(10), 36–43. https://doi.org/10.1145/3233231

Macedo, M. M. G. de, Clarke, W., Lucherini, E., Baldwin, T., Queiroz Neto, D., Abreu de Paula, R., & Das, S. (2022). Practical skills demand forecasting via representation learning of temporal dynamics. In *Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society (AIES '22)*, August 1–3, 2022, Oxford, United Kingdom. ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3514094.3534183

Mezzanzanica, M., & Mercorio, F. (2018). Big data enables labor market intelligence. In S. Sakr & A. Zomaya (Eds.), Encyclopedia of Big Data Technologies (pp. 1–11). Springer. https://doi.org/10.1007/978-3-319-63962-8_276-1

Mühlbauer, S., & Weber, E. (2022). Machine learning for labour market matching (IAB-Discussion Paper No. 3/2022). Institute for Employment Research (IAB). https://doi.org/10.48720/IAB.DP.2203

Nedelkoska, L., & Quintini, G. (2018). *Automation, skills use and training* (Vol. 202). Paris: OECD Publishing.

Oeij et al., P.R.A. (2023). Conceptual framework of Industry 5.0 to study workforce skills (BRIDGES 5.0 deliverable D1.1/ version 2 – January 2024). Leiden: BRIDGES 5.0. (Retrieved from: https://bridges5-0.eu/publications/)\

Parnes, H.S. (1962), Forecasting Educational Needs for Economic and Social Development, OECD, Paris.

Perkins, J. (2010). *Python text processing with NLTK 2.0 cookbook*. PACKT publishing.

Podjanin, A., Pouliakas, K., & Van Loo, J. (2020). The feasibility of using big data in anticipating and matching skills needs. *International Labour Organization*.

United States Congress. (2014). Workforce Innovation and Opportunity Act, Public Law 113-128. https://www.congress.gov/113/bills/hr803/BILLS-113hr803enr.pdf

Schwab, K. (2017). *The fourth industrial revolution*. Crown Publishing Group. New York https://dl.acm.org/doi/book/10.5555/3137529

Sparreboom, T. (2013). Labour market information and analysis systems. In: Cazes, S., & Verick, S. (eds). Perspectives on Labour Economics for Development (pp. 255-280). Geneva International Labour Office

Webb, M. (2019). The impact of artificial intelligence on the labor market. *Available at SSRN 3482150*.
Willems, E. (1996). Manpower forecasting and modelling replacement demand: An overview (ROA Working Paper No. 4E). Research Centre for Education and the Labour Market (ROA), Maastricht University. https://doi.org/10.26481/umarow.199604E

BRIDGES 5.0 PROJECT IDENTITY

Project name	BRIDGES 5.0 **B**ridging **R**isks to an **I**nclusive **D**igital and **G**reen future by **E**nhancing workforce **S**kills for industry 5.0
Coordinator	**Prof. Dr Steven Dhondt (scientific coordinator). Nederlandse Organisatie Voor Toegepast Natuurwetenschappelijk Onderzoek (TNO), (Netherlands)**
Consortium	Katholieke Universiteit Leuven Austrian Institute of Technology Panepistimio Patron (Patras University) Conservatoire National des Arts et Métiers, Centre d'Études de l'Emploi et du Travail-Lirsa Departamento de Educacion del Gobierno Vasco The University of Warwick Technische Universität Dortmund Stichting Platform Beta en Techniek Mondragon Goi Eskola Politeknikoa, Jose Maria Arizmendiarrieta S Coop Lietuvos Pramonininku Konfederacija Universita degli Studi di Bari Aldo Moro Universitetet I Agder Workplace Innovation Europe CLG Comau SPA Infineon Technologies Austria AG UAB Kitron Industrie 4.0 Plattform Osterreich Kriziu tyrimo centras (Hybridlab) FH Joanneum Gesellshaft MBH Kauno Technologijos Universitetas
Funding	Horizon Europe Programme, Grant Agreement Nr. 101069651
Duration	2023-2027

Funded by the European Union under grant agreement No 101069651. The contents of this publication are however the sole responsibility of the BRIDGES 5.0 project consortium only and do not necessarily reflect those of the European Union or HADEA. Neither the European Union nor HADEA can be held responsible for them.