Technological Forecasting & Social Change 173 (2021) 121177

Contents lists available at ScienceDirect

Technological Forecasting & Social Change

journal homepage: www.elsevier.com/locate/techfore

Towards ESCO 4.0 – Is the European classification of skills in line with Industry 4.0? A text mining approach

Filippo Chiarello [*,a], Gualtiero Fantoni [a], Terence Hogarth [b,c], Vito Giordano [a], Liga Baltina [b], Irene Spada [a]

[a] University of Pisa, School of Engineering, Italy
[b] Fondazione Giacomo Brodolini, Italy
[c] University of Warwick, Institute for Employment Research, UK

ARTICLE INFO

Keywords:
Industry 4.0
Technological change
Employment
Skill analysis
Text mining

ABSTRACT

ESCO is a multilingual classification of Skills, Competences, Qualifications, and Occupations created by the European Commission to improve the supply of information on skills demand in the labour market. It is designed to assist individuals, employers, universities and training providers by giving them up to date and standardized information on skills. Rapid technological change means that ESCO needs to be updated in a timely manner. Evidence is presented here of how text-mining techniques can be applied to the analysis of data on emerging skill needs arising from Industry 4.0 to ensure that ESCO provides information which is current. The alignment between ESCO and Industry 4.0 technological trends is analysed. Using text mining techniques, information is extracted on Industry 4.0 technologies from: (i) two versions of ESCO (v1.0 - v1.1.); and (ii) from the 4.0 related scientific literature. These are then compared to identify potential data gaps in ESCO. The findings demonstrate that text mining applied on scientific literature to extract technology trends, can help policy makers to provide more up-to-date labour market intelligence.

1. Introduction

Production processes have been transformed by the integration of technologies that fall under the rubric of Industry 4.0 (I4.0). Increased use of robotics and artificial intelligence routines linked to rapid improvements in communication systems have facilitated both the automation and globalisation of industrial systems. With the passage of time, interest in I4.0 has broadened from an initial focus on technological matters to encompass the economic and social consequences of its assimilation into many areas of everyday life (Giordano et al., 2021). From an economic standpoint, I4.0 is regarded as productivity enhancing. But from a social perspective, the outcomes appear more moot (Wang et al., 2020). In the current manufacturing environment, Industry 4.0 technologies are fundamental for a sustainable production, ensuring lesser weight on the environment, thanks to the utilization of available resources (Khanzode et al., 2021). Over the past fifty years technological change, has brought about real wage growth for those in employment, especially those educated and skilled to a relatively high level (i.e., college educated individuals). The response by policy makers in many countries was to invest heavily, across the board, in the provision of university level education (Reischauer, 2018).

More recently the monotonic relationship between increasing levels of educational attainment, real wage growth, and technological change no longer appears to hold. Rather the relationship between real wage growth and technological change is increasingly dependent upon the specificities of the technologies being introduced and the particular types of skills required to use them (Acemoglu and Autor, 2011; Acemoglu and Restrepo, 2020; Shakina et al., 2021). I4.0 would appear to have the capacity to allow a range of routine tasks to be done by automation. Policy makers have, accordingly, become interested in identifying those specific skills for which technological change creates a demand (Nam, 2019).

Skill, while a difficult concept to define, is usually meant to mean the qualities workers possess and the capabilities required for accomplishing certain tasks. Skills are divided into cognitive skills, such as literacy and numeracy, and non-cognitive ones, such as teamwork and various behavioural traits, required to undertake the variety of tasks which comprise a job (Heckman, 2000). Whereas once upon a time the concept

* Corresponding author at: University of Pisa, School of Engineering, Italy
E-mail address: filippo.chiarello@unipi.it (F. Chiarello).

https://doi.org/10.1016/j.techfore.2021.121177
Received 26 February 2021; Received in revised form 24 August 2021; Accepted 25 August 2021
Available online 9 September 2021
0040-1625/

of skill would have been proxied by occupation or level of education, policy makers now seek much more granular information on the specific skills required to carry out the tasks in a particular type of job and, importantly, want to know how this is changing as a consequence of technological change (Bag et al., 2021).

Traditional social science techniques used to collect information on skills - such as questionnaire surveys of workers and employers respectively - often struggle to provide the level of detail required: information is required at a high level of disaggregation. Typically, analysis is conducted at a high level of aggregation where jobs are grouped into 10 or 20 categories and the types of tasks associated with those 10 or 20 groups of jobs are empirically identified (Acemoglu and Autor, 2011; Bonacini et al., 2021). The second issue is timeliness. Occupational classifications currently in use were often developed ten or so years ago meaning that the skills and tasks which they list in relation to an occupation or job may not be up to date, especially so where jobs have been subject to relatively rapid technological change of a type associated with I4.0.

Just as the nature of technological change ushered in by I4.0 complicates the analysis of skills demand, it also provides a solution through the application of natural language processing to datasets which allow information on technologies to be linked to that on skills to reveal how the content of existing jobs is changing alongside. This is what we propose in the present paper. An example is provided of how text mining techniques can be used to identify and classify the skills emerging as a result of technological change - associated with I4.0 - and incorporated within an existing classificatory system, in this case ESCO. Natural language processing is used to collect information from databases of scientific papers which make reference to specific technologies and the associated skill needs these give rise to make them operational.

To date, the ESCO occupations pillar is a manually built taxonomy based on input from labour markets experts. Few papers exist that shows how ESCO can be used as a source for data analysis (Fareri et al, 2021). More recently, the European Commission has been exploring the use of artificial intelligence to gain insights on the evolution of the labour market, with particular focus on occupations of the ESCO taxonomy. Also, companies are experimenting in the use of ESCO as a Knowledge base. One example is JobsIreland (the Irish Public Employment Service of the Department of Social Protection) that is offering a free job advertising service to employers and job seekers using the ESCO classification as a common language. Another company, Skilllab B.V., has developed a skills assessment tool based on ESCO that helps refugees to rapidly identify their professional skills and explore the professional career pathways they could pursue. Finally, ESCO has listed many cases that demonstrate further examples of the impact of digitalisation on ESCO[1].

In this regard, we are motivated to give a further evidence of how Artificial Intelligence (and in particular Text Mining) can help to manage a complex skills framework. Two key groups of research questions (the answers of the latter depend on the first) are addressed:

A.1 - *Is ESCO in line with Industry 4.0 research?*

A.2 - Is the updating process of ESCO able to follow Industry 4.0 technological developments? If yes, by how much?

If the two sources are not consistent with one another:

B.1 - Which trending (considering scientific literature) 4.0 technologies are contained / excluded from ESCO?

B.2 - Does ESCO have the appropriate level of technological detail?

B.3 - Why are some technologies not aligned between the two sources?

To answer the questions listed above, first the technological content - i.e., the technologies which drive I4.0 skills change - of the last two versions of ESCO (v1.0 and v1.1.) are assessed. Second, changes in the content of ESCO are suggested based on the collation of the data via text mining. The overall approach is based on application of a lexical search approach using a lexicon of 4.0 technologies previously developed by Chiarello et al (2018).

The evidence is presented as follows. Section 2 presents an overview of the related literature, focusing on the ontology we will use (i.e., ESCO), on the current approaches to manage skills and job taxonomies, and on text driven tools used for skills and jobs analysis. Section 3 provides details of the method we adopted. Section 4 shows and discusses the results, highlighting the misalignment between ESCO and scientific literature, and proposing solutions. Finally, section 5 briefly recaps the main evidence of the work and summarises advantages (for researchers and practitioners) and limitations of the proposed approach.

2. Literature review

2.1. ESCO: European multilingual classification of Skills, Competences and Occupations

ESCO[2] is a multilingual classification of skills, competences, qualifications, and occupations which was first mooted in 2008 in the European Commission (EC)'s New Skills for New Jobs Communication (De Smedt et al., 2015). After a decade of development, it currently contains information on the skills of 2,942 occupations. Separately it provides a description of 13,485 skills/knowledge and further provides links to qualifications that certify the acquisition of the various skills/knowledge listed in ESCO. It is a vast information resource designed to assist with the process of ensuring that skills supply better meets demand. It provides employers with useful information on skills/knowledge they expect from their employees and thereby find the best person for a job. It also helps individuals with a better idea of the skills employers are looking for and assists them with their career development. It is also an information resource for training providers and careers guidance counsellors.

ESCO is a live entity that is undergoing a constant process of updating and enriching. There have been several versions of ESCO since 2013 when ESCO v0 (a demo version) was first launched, with the first full ESCO version (ESCO v1.0) released in mid 2017. The next release will be ESCO v1.1 at the end of 2021. The work on ESCO v1.1 started in 2018 in close collaboration with sectoral experts and ESCO stakeholders. An important input is expected to be the results from a series of studies commissioned by the EC to identify skill needs in various sectors (resulting from the EU's blueprints on sectoral cooperation on skills initiative), alongside other research activities which have sought to identify emerging skill needs.

In the present work, consideration is given to two versions of ESCO (see Table 1). These are 1.0 (more precisely the last version of 1.0, 1.0.8) and a pre-release version of 1.1. ESCO v1.1. The pre-release version includes a content update with the inclusion of occupations, skills and knowledge not currently part of ESCO v1.0. This explains the different number of occupations between the two versions shown in Table 1. The pre-release version is already available in a dedicated portal[3] that was used for data collection.

Few works exist on the analysis and use of ESCO data. Fernández-Sanz et al. (2017) worked on the development of common language in the area of ICT in order to prove a consistent model covering the different views of the area by different standards (ESCO and the European ICT Body of Knowledge). Also, le Vrang et al. (2014) worked in a

[1] For the list, visit https://www.esco-projects.eu/esco/portal/howtouse/dc9a812c-8135-4f46-92f6-7364a1714ae0 (opened in June 2021)

[2] https://ec.europa.eu/esco/portal/home

[3] https://ec.europa.eu/esco/portal/news/686d8b99-fe02-40c7-a374-57a37c72bca6

Table 1
Comparison between 1.0 and 1.1 versions of ESCO.

Version	Date of release	Number of Skills	Number of Knowledge	Number of Occupations	Qualification	Number of Languages
v 1.0	2017 (Last update v 1.0.8 in August 2020 - Corrections in the translation of descriptions, labels, ISCO groups and NPTs in various languages)	10.583	2.902	2.942	Framework of qualification pillars (17191 qualification organized per thematic area, location and EQF level)	27
v 1.1	2020	10.874	3.052	3.008	-	1 (English)

similar direction, working on the development of a semantic interoperability between the labour market and ESCO, to help address the problem of skills mismatch. Colombo et al. (2019) developed a labour market intelligence tool using machine learning techniques and focusing on web vacancies on the Italian labour market. Their goal was to calculate the skills requirement for each occupation. Also, in this case, information about the job market is mapped onto the ESCO classification. More recently, González et al. (2021) reported a first experimentation on the potential of matching knowledge elements in ESCO to Wikidata using Named Entity Recognition and document similarity. Similarly, Fareri et al. (2021), has used ESCO to test their Natural Language Processing system, able to extract soft skills from any document. Finally, Mirski et al. (2017) using the ESCO database triangulates skills, learning items and job offers to simulate career paths through visualization.

As it is evident, many attempts already exist to map ESCO to external data sources (for a full view of the literature, see Fareri et al. (2021)). The present paper is novel, since, to the best of our knowledge, we present a first attempt to map ESCO to scientific literature. Given the peculiarities of scientific jargon, we rely on Natural Language Processing techniques to reach our goal. This solution has proven to be effective in a part of the literature dealing with ESCO (Colombo et al., 2019; González et al., 2021; Fareri et al., 2021). The second characteristic of novelty, with respect to the revised literature on ESCO analysis, is in the purpose of the present paper. Our goal is to measure if (and how much) ESCO is in line with the technological trends reported in the scientific literature (with a focus on Industry 4.0). Other works focus on the alignment between ESCO and the labour market (Fernández-Sanz et al., 2017; Colombo et al., 2019; González et al. 2021). In this sense, our contribution,given the positive evidence that emerges from our results, may further stimulate the usage of ESCO to analyse the labour market, which is typically time-aliened (if not delayed) with respect to scientific literature (Zeidmane and Cernajeva, 2011).

2.2. *Current approaches for skills and jobs classification*

As the foregoing has indicated, the interest of policy makers has become increasingly focused on being able to anticipate the skills requirements (Felten et al., 2021). Skills and jobs are codified to occupational classifications, to group them together if considered similar. Data which allows something to be inferred about the skill content of particular jobs or occupations in order to design occupational classifications has, until relatively recently, been mainly derived from surveys of individual's skills (PIAAC, 2009), surveys of employers (Massarelli and Wozowczyk, 2009), skills and technology foresight exercises (Kamprath and Mietzner, 2015; Vallet et al., 2020), expert panels (Srivastava and Patil, 2020), or mixed methods (Paliokaitė et al, 2015)

More recently through the application of natural language processing techniques it has been possible to obtain data on skills from sources which previously would have been impractical to search, such as job bulletin boards/recruitment websites (cf. Cedefop's OVATE database) and, potentially individuals' CVs (cf. Europass). Accordingly, a fifth means of collecting data on skills can be added to the list: text mining. This has allowed more detailed information to be collated on the specific skills required in specific jobs to fulfil particular tasks. We will review this approach in greater detail in section 2.3.

The collation of data about skills from a variety of sources has been linked to the creation of skills databases typically structured around occupational classifications. Data from a variety of sources can be combined to potentially produce something which is greater than the sum of their parts. Important here is the Dictionary of Occupational Titles (DOT), first developed in the USA in 1938, which was subsequently superseded by Occupational Information Network (O*NET) which provides detailed information about the tasks and skills required in an occupation, wage level, qualifications required for entry, etc. (Handel, 2016). It is the most comprehensive database of its kind. In the European Union, ESCO is a more recent development in this direction. Under the pressure of COVID-19, many studies have highlighted the high value of these frameworks as knowledge bases for other studies (Baker et al., 2020; Dingel and Neiman, 2020; Bochtis et al., 2020)

2.3. *Text mining tools relevant to skills and jobs analysis*

Before describing the methodology in more detail, it is worth considering some of the main developments in the use of data science to collate and analyse data on skills and jobs. The automatic analysis of text is called Natural Language Processing (NLP) (Manning and Schutze, 1999). The NLP approach usually involves the execution of a software pipeline with the aim of extracting information from text.

These approaches have been used to bring about revelatory findings in the analysis of, as an example, customer reviews (Ali et al., 2020), social media usage (Liu et al, 2021), and patent analysis (Trappey et al, 2021; Yuan and Cai, 2021; Song et al, 2018). More recently they have been used in the analysis of skill demand and changes in the job profiles of people whose jobs have been affected by I4.0 (Malandri et al 2021; Fareri et al., 2020). In the field of human resource management (HRM), the main application has been in relation to recruitment. For example, research in HRM field has been used to monitor the changing skill needs mentioned in job adverts (Giabelli et al, 2021; Giabelli et al, 2020; Lumauag, 2019; Boselli et al, 2018).

The main methods used to analyse HRM related textual data are as follows:

- *Decision Tree Algorithm*, that is a widely spread algorithm used to classify datasets, following human-readable classification rules (Ben-Haim and Tom-Tov, 2010);
- *keywords Extraction/Knowledge Extraction*, that are, respectively, the extraction of relevant words of the text in order to summarize it and their ontological formalization in a machine-readable format (Manning and Schutze, 1999) using techniques such as word embeddings (Devlin et al., 2018);
- *association Rules*, whose identification leads to the discovery of interesting relations between variables in databases (Piatetsky--Shapiro, 1991);
- *named Entity Recognition* (NER), that is a technique used to recognize information units (such as the names of people, organizations, locations) and numeric expressions (such as time, date, money and percentages) in unstructured texts (Nadeau and Sekine, 2007).

In the described methods, the most interesting one for the present work is NER. These techniques can be used to detect technologically related textual entities (Chiarello et al., 2021; Chiarello et al., 2020; Melluso et al., 2020; Bonaccorsi et al., 2020). The present work uses NER techniques to identify technologies in scientific papers and job profiles taxonomies (as contained in ESCO).

Before moving to the description of the adopted methodology, Table 2 summarizes the relevant literature for current (qualitative) approaches for skills and job classifications and quantitative and text mining tools relevant to skills and jobs analysis.

3. Methodology

This section presents the methodology to collate data on I4.0 technologies from both ESCO and scientific publications. The method is composed of the following steps: (1) search and retrieve a set of scientific papers on I4.0; (2) tag technologies contained in both ESCO and the scientific papers; (3) revise and clean the data produced by the automatic extraction; and finally (4) measure the gap between the two sources. These methodological phases are described in detail in the next subsections.

3.1. Paper retrieval

The goal of the first phase was to retrieve a set of papers on I4.0. The papers were retrieved from Scopus, a database of scientific publications launched in 2004 by Elsevier. The Scopus database was chosen because of the high volume of papers it contains, the high frequency with which it is updated, and its comprehensiveness with respect to engineering and innovation papers. Scopus contains abstracts and citation data from all scholarly journals indexed. The query used to retrieve the papers set of Industry 4.0 was:

TITLE-ABS-KEY("Industry 4.0" OR "Industrie 4.0")

The query was performed on 12/02/2021 and retrieved 13,712 papers along with the title, abstract, keywords and year of publication of the documents. Not considering the whole document for the analysis, does not represent a limitation to our approach since we are searching for relevant mentions of I4.0 technologies. Mining the whole text could retrieve also spurious mentions (e.g., technologies mentioned in the state of the art of the paper), thus papers that are out-of-scope or weakly linked to I4.0. This methodological choice (limiting the analysis to the abstract) is also supported by other works that are focusing on detecting emerging technologies using text mining techniques on the scientific literature (Ranaei et al., 2020; Xu et al., 2021).

3.2. Technology extraction with named entity recognition

Named Entity Recognition (NER) allowed us to automatically identify the technologies in our collection of the textual data. The NER method used for this purpose was a lexicon-based approach, that automatically identifies I4.0 technologies using the lexicon developed by Chiarello et al., 2018. A sample of this lexicon is shown in Appendix A. Using this approach, we were able to extract 3,566 different entities. In particular, we retrieved 3,322 technologies from the title, abstract and keywords of Industry 4.0 papers and 663 technologies from the skill labels, skill alternative labels and skill description of the different ESCO versions. The overall method was implemented in RStudio (RStudio Team, 2020) using the *tidytext* packages for textual analysis (Silge and Robinson, 2016). The running time for processing the industry 4.0 papers was equal to 600 seconds, while for the ESCO skills/knowledge data was around 680 seconds on a HP ProBook 450 G5 machine running an Intel(R) Core(TM) i7-8550U processor.

3.3. Data pre-processing pipeline

We then processed the list of 3,566 technologies with a data preprocessing pipeline for removing non-technology terms and for clustering together synonyms and similar technologies, following the approach used in Mazzei et al. (2021). The process consisted of the following steps.

A. Data cleaning

In this first step, we automatically removed:

- entities composed by a number of words greater than 4 (e.g. *real-time not only the monitoring, the world is accelerating the digitisation of many branches of the; the complex way radio waves propagate; gps aided geo augmented navigation; v. with mass customized production becoming the; modules and types of equipment*);
- entities composed by a number of characters lower than 3 (e.g. *nm, mm, v, v., ie*);
- entities starting by a punctuation (e.g. *.net, .networks, .com, .org, -system*).

Table 2

Summary of the relevant literature. For each work we show the main Authors, year, the Journal, the purpose and approach, and if the work is a qualitative, quantitative or mixed contribution. The references can be found in the reference section.

Authors	Year	Journal	Purpose	Approach	Type
Kamprath & Mietzner	2015	*Technological Forecasting and Social Change*	Integrate individuals' competences in scenario studies	Scenario Study	Qualitative
Dingel & Neiman	2020	*Journal of Public Economics*	Estimate how many jobs can be done at home	Survey	
Srivastava & Patil	2020	*Journal of Critical Reviews*	Study the relation of education, knowledge and skills	Expert panels	
Vallet et al.	2020	*Futures*	Provide a basis for socially responsible planning	Scenario Study	
Paliokaitė et al.	2015	*Technological Forecasting and Social Change*	Select smart specialisation priorities	Expert panels, statistical analysis	Mixed
Felten et al.	2021	*Strategic Management Journal*	Measure of an occupation's exposure to AI	Crowd-sourced analysis	
Boselli	2018	*Journal of Intelligent Information Systems*	Collect and classify multilingual job vacancies	Statistical analysis	Quantitative
Colombo et al.	2019	*Information Economics and Policy*	Explore the link between automation and skills	NLP	
Lumauag	2019	*International Journal of Recent Technology*	Provide decision support for personnel selection	Decision Trees	
Baker et al.	2020	*PLoS One*	Estimate the number of workers exposed to infection	Statistical analysis	
Bochtis et al.	2020	*Sustainability*	Asses the pandemic impacts on agricultural labor	Statistical analysis	
Fareri et al.	2020	*Computers in industry*	Assess the impact of Industry 4.0 on jobs	Named Entity Recognition	
Giabelli et al.	2020	*Multimedia Tools and Applications*	Explore labour market information	Graph databases	
Giabelli et al.	2021	*Applied Soft Computing*	Recommend suitable occupations to a user	Word Embeddings	
González et al.	2021	*Metadata and Semantic Research*	Populate skills ontologies	Named Entity Recognition	
Malandri et al.	2021	*Computers in Industry*	Link different labour taxonomies	Word Embeddings	

Furthermore, we manually revised the excluded words, in order to avoid the loss of important information. After this step, we replaced all hyphen characters ("-") with an empty space for identifying similar technologies (e.g., *cyber-physical-system* was replaced by *cyber physical system*). At this point, the entities list comprised 2,435 items.

B. Data screening
Data Screening was a manual task to clean the corpus of entities extracted by the machine. The entities list was manually screened by three of the authors (two professors at the School of Engineering of the University of Pisa, and a Ph.D. Student of the course in Smart Manufacturing of the University of Pisa). Each author was provided with a table containing the list of extracted entities. The assignment was the following:

"Read each extracted entity and decide whether the entity is a technology or not".

The authors were allowed to use external sources for achieving their goal. After the individual screening, the authors discussed together the cases were there was not a full agreement and took a final decision on each one. These cases were 19 on the total of 2,435, showing a strong agreement. After this step, the list of technologies comprised 1,431 elements.

C. Identifying synonyms
This final step identified the groups of technologies in our set that were synonyms, using semantic similarity methods and a pair-wise comparison. Each technology was mapped to a vector in a 1024-dimensional semantic space using the Bidirectional Encoder Representations from Transformers (BERT) model (Devlin et al., 2018). The BERT model aimed at structuring a technology as a vector composed of 1024-dimensions (for example the *Artificial Intelligence* technology represented as the following mathematical vector: -0.6113, -0.4964, -1.9548, …, 0.9718, 0.3563, 0.2444). Structuring technologies as vectors has proven to be effective for identifying similarities between them, as recently shown by Azimi et al., (2020). We then measured the cosine similarity, as the cosine of the angle between the two semantic vectors (Han et al., 2011). We clustered technologies with a cosine similarity greater than 0.97, following the approach described by Hu et al (2018).

In Table 3, we provide a sample of the technologies and the information about relative groups extracted by the proposed methodology. We obtained a final set of 752 different groups of technologies (we recall that a group is composed of synonyms of technologies). In the table, the first column contains the entities extracted using the lexicon of I4.0 developed by Chiarello et al. (2018); the second column is the group label, decided using the approach by Hu et al (2018). For facilitating the reproducibility of our work, the complete list of the collected technologies is also available online (Giordano et al., 2021).

Table 3
A sample of the technologies extracted using the lexicon of Chiarello et al. (2018) with the group label assigned using the semantic similarity method, i.e., Bidirectional Encoder Representations from Transformers (BERT). The column "Technology" contains the entities extracted, while the column "Tag" is the cluster label.

Technology (Synonyms)	Group Label
3d display	*3d display*
3d displays	
3d printers	*3d printing*
3d printer	
3d printed	
printed on 3d	
printed in 3d	
computer aided design	*computer aided design*
computer aided designs	
computer aided designing	

3.4. Measures

To measure if and how much ESCO is updated with technological trends in scientific publications (research questions A.1 and A.2), we evaluated the growth of the number of papers containing I4.0 technologies. The main assumption we make is that rapid growth in paper numbers indicates a technology with high growth potential (Lee and Lee, 2013). There exists a large literature that describes method to measure the growth of a technology, but the most used method is based on the growth curves (also known as S-curve) (Daim et al., 2006). The S-curves represent the growth in technological performance over time and they are operationalize using various indicators such the number of patents or scientific publications produced (Adamuthe and Thampi, 2019). Anyway, the growth curves are not able to provide a synthetic index of a technology growth (as we need for the present method) but gives a diffusion stage of the technology under the analysis (emerging, growth, maturity or saturation phase). Technology trend analysis (TTA) is another methodology used for understanding the growth pattern of a technology in literature. Mun et al. (2021) provide a comprehensive literature review on the technique used for TTA. Among other, we use a synthetic index for measuring the papers activity: Relative Development of Growth Rates (RDGR). RDGR is widely used to assess the growth of a technology in literature. The methodology was firstly introduced by the seminal paper of Ernst H. (1998), and Huang (2016) used the RDGR for assessing the technological attractiveness of cloud computing technologies. Also, Baumann et al. (2021) employ the RDGR to evaluate the patent growth of different countries for battery storage, hydrogen and bioenergy technologies.

For each technology in our final list, we calculate the RDGR to investigate the growth rate of technologies in scientific papers. We measure the RDGR as proposed by Huang et al. (Huang and Hsu, 2017). It is measured by calculating the number of papers published during the past three years (i.e., 2018–2020) divided by the number of papers published n a 9-year span (i.e., 2012–2020):

$$RDGR_i = \frac{(N.\ of\ Papers\ containing\ tech_i)_{2018-2020}}{(N.\ of\ Papers\ containing\ tech_i)_{2012-2020}}$$

The formula considers the portion of scientific literature about the i-th technology 4.0 produced in the last 3 years on the total number of articles. The starting year is where the first paper on Industry 4.0 appeared (Ahrens, 2012).

4. Results and discussion

In this section we present and discuss the results we obtained applying the method we described in section 4 to scientific papers and to the two versions of ESCO. The goal is to answer the two groups of research questions introduced in section 1.

4.1. Measures of alignment

In the present section, we aim to answer to the first two research questions:

A.1 - Is ESCO in line with Industry 4.0 research?

A.2 - Is the updating process of ESCO able to follow Industry 4.0 technological developments? If yes, how much?

Table 4 shows a summary of our results. As it is evident from columns 2 to 4, the results can be divided in five cases, considering if the technologies are contained in ESCO (v1.0 or v1.1) or in I4.0 scientific papers. Columns 5 and 6 shows respectively the number of technologies in the top-80 (in terms of papers containing the technology) and in the whole extracted list. Column 7 contains an example of the technologies. Considering the top-80 technologies in terms of scientific papers, allowed us to identify a set of more relevant technologies. In the whole

Table 4
The number of technologies for each possible case. The **yes** indicates that the technology under analysis is in a given class (in ESCO version 1.0, in ESCO version 1.1, in the set of Industry 4.0 papers). The **no** indicates that the technology under analysis is not in the given class. The column "Example" contains three technologies that have fallen in the case under the analysis.

Cases	in ESCO v1.0	in ESCO v1.1	In Industry 4.0 papers	Total number of technologies (in the top-80)	Total number of technologies	Example
Column:	**2**	**3**	**4**	**5**	**6**	**7**
1	**yes**	**yes**	**Yes**	47	135	*Big Data, Machine Learning, 3D Modelling*
2	**no**	**yes**	**Yes**	9	18	*Human-Machine Interaction, Deep Learning, Wearable Device*
3	**no**	**no**	**Yes**	24	563	*Digital Twin, Edge Computing, Manufacturing Execution System*
4	**yes**	**yes**	**No**	/	32	*3D Footwear Prototype, Apache Tomcat, Haskell*
5	**no**	**yes**	**No**	/	4	*Document Management Procedure, Intelligent Safety System, Manufacture Metal Additive*
			Total:	80	752	

set of 716 technologies contained in papers, many are in the long tail of the distribution, i.e., contained in only on scientific papers. Even if these technologies can be interesting for a qualitative study, they are nor relevant for the purposes of the present statistical analysis.

First of all, the set of technologies (47 in the top-80 and 135 in the whole set) is contained both in ESCO (and have been contained since version 1.0) and 4.0 scientific papers (case 1). This is a first evidence of the good quality of the alignment between ESCO and I4.0. Then, case 2 shows the technologies contained in papers but not in ESCO v1.0, while case 3 shows technologies contained in papers but not in any versions of ESCO. To this last case, belongs about the 30% (24/80) among the top-80 technologies. This will be discussed in section 5.2.1, together with the technologies that have been added to the new version of ESCO (1.1), that are 10 among the top-80. Also, as it is evident from column 6, on the total extracted technologies (752), 563 (almost 80%) the 75% are not contained in ESCO (563/752).

On the opposite, from cases 4 and 5 it is evident that there also exist technologies that are contained in ESCO but not mentioned in scientific papers. We will discuss in further details these cases in section 5.2.2.

To summarise the numbers, we have from the 752 technologies in the whole dataset (union of ESCO versions and papers), 716 are contained in papers (563 only in papers), and 189 are contained in ESCO versions (36 only in ESCO). From this description, it is possible to notice that ESCO is partially aligned with scientific production (research question A.1). In fact, considering the whole list of technologies only 20% are contained in papers and one of the versions of ESCO. If we focus on the top-80 (the set of more relevant technologies) the situation changes: we have that 70% of technologies are contained in both papers and one of the two versions of ESCO. In other terms, this is a first evidence that ESCO contains a vast majority of the I4.0 technologies that are relevant (from the perspective of research) for Industry 4.0.

Thus, we can move to a more precise answer for the second research question, that is if and how much is ESCO updated in line with what is happening in scientific production related to I4.0 (A1.2). Given the increase of the number of technologies among the top-80 in terms of scientific papers from ESCO 1.0 to 1.1 (a delta of 9 technologies, see Table 3) a preliminary answer to the first part of research question A1.2 is yes. We see in fact, an increase in the number of technologies considered relevant by the scientific literature and included in ESCO. To understand if this is statistically relevant in terms of RDGR (the measure of the growth of the technologies in scientific production) we compare the first three cases of Table 4. In Figure 1, we report the statistical distribution of RDGR for the top-80 technologies. The figure shows the box plots and in particular the median (black horizontal line), the Interquartile range (dimension of the box) for each class discussed above. As it is evident from Figure 1, the RDGR is high on average (the means of the distribution are all above 80). This is due to the fact that there has been an explosion of 4.0 related papers (Chiarello et al, 2018). This effect does not change the effectiveness of the statistical tests

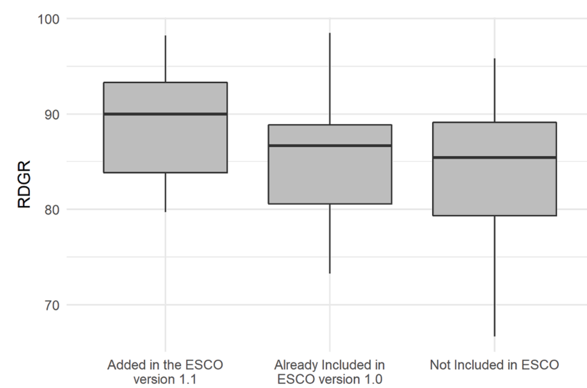

Fig. 1. Distribution of RDGR for the cases of technologies (Added in ESCO version 1.1, Already Included in ESCO version 1.0, Not Included in ESCO). Note: The box plots show the mean (cross), median (horizontal line), upper and lower 25% quartiles (top and bottom of the bar), maximum and minimum values as well as the outliers (dots) for each of the factors within the group.

provided in the present section. The figure also shows as the class "Added in ESCO version 1.1" has the highest values of RDGR in the technologies sample (i.e., these are the most trending technologies, considering scientific literature). This means that the newly added 4.0 technologies are new also for the scientific community, a remarkable result.

The one-way Analysis of Variance (Anova) test was performed on the top-80 technologies sorted by the number of papers (see Table 5) for analysing if this difference is statistically relevant. Anova was recognised as appropriate for this test because it compares means in a situation where there are more than two groups (Howell, 2012). The Anova test is popular in this field of research, as affirmed by Scuotto et al. (2020). The assumption of the one-way Anova is that the distributions of the three cases ("Added in the ESCO version 1.1", "Already included in ESCO version 1.0" and "Not Included in ESCO") are normal. For each case, we tested if the observations are normally distributed with the Shapiro-Wilkinson test. If the p-value of the Shapiro-Wilkinson test is

Table 5
Significance of the mean differences among the groups of technologies (Added in ESCO version 1.1, Already Included in ESCO version 1.0, Not Included in ESCO). The One-way ANOVA test; p-values: "." 0.05 confidence level; "*" 0.01 confidence level; "**" 0.001 confidence level; "***" 0 confidence level. The p-value of the test is equal to 0.08.

	Estimate	Relevance
Added in ESCO v1.1	89.252	***
Included in ESCO v1.0	85.511	
Not Included in ESCO	83.480	*

greater than the chosen alpha level of .05, then the null hypothesis cannot be rejected and there is evidence that the data tested are normally distributed. In our case, the p-value of the test for the “Added in the ESCO version 1.1” case is 0.704, for “Already included in ESCO version 1.0” one is 0.467. Finally, in the case of “Not Included in ESCO” the p-value is 0.118. For these reasons, the null hypothesis of Shapiro-Wilkinson test cannot be rejected and the Anova assumption of normality is verified.

The Shapiro-Wilkinson test and the one-way Anova are performed with the R base package (R Core Team, 2020).

The results of the Anova test are shown in Table 5. In the experiment, all assumptions for Anova are verified: dependent variables are measured at interval level, independent variables consist of independent groups, observations are independent, there are no significant outliers, dependent variables are approximately normally distributed and there is homogeneity of variances.

Table 5 shows as the technologies added in ESCO v1.1 (case 2) have a higher Estimate value of RDGR than other cases (p-value = 0.08***). As expected, no statistical differences can be drawn also for the comparison between case 1 and case 3.

To sum up, ESCO versions 1.1 is a statistically relevant step forward being in line with the fourth industrial revolution, giving a positive response also to our second research question (research question A.2). Anyway, it is also interesting to notice that ESCO is not fully in line with technological trends in I4.0. Given this result, it is reasonable to move to the investigation of the second group of research questions, looking in

Table 6
The top-80 technologies extracted using the Industry 4.0 lexicon (Chiarello, et al., 2018) with the number of papers related to Industry 4.0 and the amount of skills/knowledge in ESCO (version 1.0 and 1.1, respectively) that are referred to the technology under the analysis. In **bold** the technologies contained in the papers and not in ESCO, in *italic* the ones that have been added in version 1.1.

Technology	N. of papers	RDGR (%)	N. of skills v1.0	N. of skills v1.1	Technology	N. of papers	RDGR (%)	N. of skills v1.0	N. of skills v1.1
Internet of Things	3176	83.7	2	2	Mechatronic	101	73.3	10	11
Cyber Physical System	**1786**	**74.2**	**0**	**0**	**Real Time Monitoring**	**99**	**88.9**	**0**	**0**
Big Data	1179	81.7	2	3	Laser	95	87.4	24	26
Cloud Technology	1151	82.9	9	22	**Collaborative Robot**	**94**	**90.4**	**0**	**0**
Simulation Technology	1109	87.7	20	28	Smart Device	91	87.9	2	3
Artificial Intelligence	1045	91.9	5	11	Global Positioning System	88	93.2	12	12
Machine Learning	772	94.2	3	7	Human Machine Interaction	84	82.1	1	2
Robot	677	86.0	8	9	Cryptography	80	88.7	2	2
Cloud Computing	619	83.4	1	1	**Anomaly Detection**	**79**	**84.8**	**0**	**0**
Digital Twin	**544**	**93.7**	**0**	**0**	Middleware	78	79.5	1	1
Augmented Reality	*377*	*85.9*	*0*	*3*	**Scada**	**76**	**85.5**	**0**	**0**
Virtual Reality	346	79.8	1	4	**Smart Manufacturing System**	**75**	**85.3**	**0**	**0**
Neural Network	316	93.3	1	3	Wide Area Network	73	93.1	1	2
Led	274	86.5	8	10	**Service Oriented Architecture**	**70**	**75.7**	**0**	**0**
Blockchain	*272*	*97.4*	*0*	*6*	**Manipulator**	**69**	**85.5**	**0**	**0**
Deep Learning	*254*	*93.3*	*0*	*1*	Drone	67	98.5	3	5
Data Mining	251	82.5	7	7	Enterprise Resource Planning	65	80.0	2	2
Cybersecurity	214	91.6	3	3	**Product Life Cycle Management**	**65**	**69.2**	**0**	**0**
ICT Communication Protocols	207	80.2	3	3	Ethernet	64	76.5	2	2
Human Machine Interface	*201*	*83.1*	*0*	*1*	**Mobile Robot**	**63**	**87.3**	**0**	**0**
3D Printing	186	91.4	5	5	User Interface	63	88.9	9	9
Manufacturing Execution System	**171**	**71.9**	**0**	**0**	Microcontroller	62	91.9	1	1
Wireless Sensor Network	*167*	*83.8*	*0*	*1*	Access Control	57	87.7	5	6
Computer Aided Manufacturing	156	75.6	5	5	*Wearable Device*	*57*	*98.2*	*0*	*1*
Data Processing	148	83.9	8	8	**Arduino**	**55**	**90.9**	**0**	**0**
Programmable Logic Controller	147	83.0	2	2	3D Modelling	51	92.2	9	9
Radio-Frequency Identification	**144**	**80.6**	**0**	**0**	**Intelligent Manufacturing System**	**51**	**66.7**	**0**	**0**
Edge Computing	**144**	**95.8**	**0**	**0**	**Collaborative Robotics**	**49**	**89.8**	**0**	**0**
Computer Aided Design (CAD)	137	81.0	17	18	**Raspberry Pi**	**49**	**87.7**	**0**	**0**
Computer Vision	129	86.8	1	4	Semantic Web	47	78.7	1	1
Smart Grid	*128*	*79.7*	*0*	*3*	Business Intelligence Software	46	80.4	2	2
Decision Support System	125	91.2	2	2	**Bluetooth**	**43**	**83.7**	**0**	**0**
Agent Based Model	**121**	**85.9**	**0**	**0**	Field Programmable Gate Array	43	88.4	1	1
iOS	121	76.9	3	3	**Flexible Manufacturing System**	**43**	**83.7**	**0**	**0**
Automatic Control	120	88.3	2	2	Representational State Transfer	43	88.4	6	7
Wireless Network	**113**	**82.3**	**0**	**0**	**Machine Vision**	**42**	**90.5**	**0**	**0**
Virtualization	112	88.4	1	1	Tablet	42	64.3	4	5
Autonomous Vehicle	*109*	*91.7*	*0*	*1*	**Internet of Service**	**41**	**73.2**	**0**	**0**
Embedded Systems	107	79.4	2	3	Software as A Service	41	87.8	5	5
Computer Numerical Control	103	77.7	14	14	*Remote Monitoring*	*40*	*90.0*	*0*	*2*

greater details to the differences between research and ESCO, also discussing the reasons why these I4.0 technologies are/are not contained in ESCO for proposing improvements.

4.2. *Specific technological differences*

In this subsection, we dive into a more specific view, answering to the second group of research questions:

B.1 - Which trending (considering scientific literature) 4.0 technologies are contained / excluded from ESCO?

B.2 - Does ESCO have the appropriate level of technological detail?

B.3 - Why are some technologies not aligned between the two sources?

As it is evident from Table 4 and from the discussions of section 5.1, we can have two types of misalignments: 4.0 technologies that are trending in papers and not contained in ESCO, and 4.0 technologies that are contained in ESCO but not mentioned in scientific literature. The next two subsections respectively face these two dimensions. When identifying gaps between the two sources, we also evaluate the reasons and offer possible solutions, proposing answers to research question B.3.

4.2.1. *Gaps of ESCO*

Table 6 ranks the top-80 technologies according to the number of 4.0 related scientific papers (second column). The RDGR (third column) shows the *Relative Development of Growth Rates* (see section 4.4), used to compute the analysis in section 5.1. The fourth and fifth columns are the numbers of skills/knowledge in ESCO version 1.0 and 1.1 respectively, that mentioned the technology under analysis. Technologies in **bold** are the ones that are not contained in ESCO; technologies in *italic* are the ones introduced in ESCO 1.1. The complete table of the 716 technologies is reported in the Appendix B.

As it was evident from Table 4, a majority of 4.0 technologies is taken into consideration by ESCO (47 on 80). This is the case, among the others, of some pillars of I4.0 such as *Internet of Things, Big data, Cloud Technology and Artificial Intelligence* (Vaidya et al., 2018)*;* but also, of more specific and more technical such as *Embedded Systems, Microcontroller, Semantic Web* and *Field Programmable Gate Array* (Tubaishat and Madria, 2003). ESCO, in the new version, contains technologies that are growing fast, such as *Blockchain* (Chiarello et al., 2021; Mangla et al., 2021) (RDGR of 97.4)*, Deep Learning* (Blazquez and Domenech, 2018) (RDGR of 93.3), *Autonomous Vehicle* (Penmetsa et al., 2019)(RDGR of 91.7), *Wearable Devices* (RDGR of 98.2) and *Remote Monitoring* (RDGR of 90.0). ESCO is thus both timely/ up to date (research question B.1) and at the right level of detail (research question B.2).

From the ranking in Table 6, the first technologies that are missing in ESCO are *Cyber Physical Systems, Digital twin* (Stark and Damerau, 2019) and *Manufacturing Execution Systems (MES)*. This is a limitation to a timeliness update of ESCO related to I4.0, because these concepts are fundamental and strongly established. As evidence, these technologies are part of the pyramid of automation (Cimino et al., 2019) and are also cited in the RAMI framework (Zezulka et al., 2016). If these technologies are not as fundamental as skills, they are particularly important as knowledge and should be inserted to assure an alignment to the I4.0 technological framework.

Similarly, technologies that are rarer in scientific literature, but still important as a baseline for I4.0 such as *Smart/Intelligent/Flexible Manufacturing Systems* and *SCADA* are also missing. These concepts can be included in further versions of ESCO as knowledge, to ensure that workers at different levels of the hierarchy can properly work with the 4.0 paradigm. The absence of these concepts in ESCO risks to not help enough companies (especially the less aware, such as SMEs) to encompass the three main dimensions of I4.0, namely high-grade digitization of processes, smart manufacturing, and inter-company connectivity (Müller et al., 2018).

For what concerns the absence of the technology *wireless network* in ESCO, we can find similar entries in skills such as “design computer network” and “test wireless device”; on the other hand, technologies as *Radio-Frequency Identification* and *Bluetooth* are totally missing. Going further in the table, also the group related to robotics technologies is missing: *Collaborative Robot/Robotics* and *Agent-based model*. Networks and robots (especially networks of robots) are another pillar of I4.0, especially after the COVID-19 breakout (Abdel-Basset et al., 2021). The fact that these technologies are missing from ESCO is not a drawback of the ontology per se. It is fact reasonable to include more generic terms in the ontology, able to cover all the possible shades of skills and knowledge. Anyway, this goes in the opposite direction of the fact that our algorithm found in ESCO several trade names of technologies, such as *Android* and *iOS* (mobile operating systems), *ASP.NET* (an open-source server-side web-application framework for web development) (see Appendix B). If we understand that this may be due to the wide diffusion that these names have, this creates a strong limitation to a positive answer to our research question B.2. ESCO shows a problem in terms of granularity in the taxonomy, since more generic words are missing (i.e., robots and networks), while more specific are present (i.e., trade names). Also, this is an evidence that ESCO favour the inclusion of skills with respect to knowledges.

Furthermore, other trade names that are fundamental in I4.0 (and are contained in papers) are missing, such as *Arduino* and *Raspberry Pi* (microcontrollers). Since trade names are technologies at the lowest level of granularity (i.e., very specific technologies), this inconsistency should be managed in future versions of ESCO, avoiding the usage of trade names or trying to include all the trade names that are relevant for a specific paradigm. We understand that this is not a trivial task: the methodology presented can be used to support this process by an automatic search of trade names in scientific literature or in patents.

Another omitted technology is *Edge Computing* (Shi et al., 2016), an example of a strong I4.0 research trend (RDGR of 95.8). *Edge computing* is important because it is more than a new computing paradigm (Han et al., 2015) and because it is going to have an impact in different sectors such as healthcare (Tortorella et al., 2021) and banking (Garg et al., 2021). Thanks to the proliferation of the Internet of Things and the success of Cloud Computing the new paradigm calls for processing the data at the edge of the network. This, which seems a technical change, brings a strong impact on the way that Industrial systems will be designed, managed and used, thus having an important impact on skills (Catal and Tekinerdogan, 2019). Our approach of using the RDGR to spot growing technologies, can help ESCO in the future to timely spot this kind of technological trends. Furthermore, ESCO already contains technologies that are seeds of *Edge Computing*, such as *Embedded Systems,* and can leverage on these to introduce the concept.

On the other hand, a point in favour of a positive answer to the research question A.1 is the fact that ESCO contains technologies related to *Artificial Intelligence,* such as *Machine Learning, Big Data* and *Deep Learning,* that are strongly impacting modern digital factories (Lu, 2019; Raut et al., 2021). What are surprisingly left out are some industrial applications of these technologies, such as *Real Time Monitoring, Machine Vision* and *Anomaly Detection.* This has an important impact on skills: leaving the application (but not the technologies) outside, can cause the ontology to under-represent some workers, especially the ones who operate directly in the design and the maintenance of these systems. We also point out that it is particularly important to include these applications, since they are tasks that can be carried out by machines. Their presence in ESCO can clarify the interaction that workers will have with Artificial Intelligence also lowering some barriers and fears raised by recent research on the topic (Frey and Osborne, 2017). In this direction goes the introduction in ESCO 1.1 of Artificial Intelligence related applications such as *Autonomous Vehicles.* ESCO can take great advantage in filling this gap, by using recent systems able to extract skills and knowledge from different kinds of textual sources (such as Fareri et al., 2020).

Another missing I4.0 pillar is *Product Lifecycle Management*, which is important for two reasons: it is considered as a baseline in green transitions, and it makes use of innovative digital technologies and methods (Benzidia et al., 2021; Jabbour et al., 2019). It is evident that ESCO's attention on the green transition is growing considering the fact that the technology *Smart Grid* has been included in ESCO 1.1 and also by the fact that a similar concept is in the skill "assess life cycle of resources". Anyway, given the fact that the proliferation of green related semantics can disorientate workers and HR specialists, in the future development of ESCO, improved searchability of the tool may be needed. At this time in fact, the tools use alternative labels in order to make the navigation easier for the user. ESCO can consider the use of advanced search techniques such as vectorised words representation to avoid misleading results, to be less error-prone, and to stay updated with the semantics of exploding technological fields. (Dai and Callan, 2019).

Finally, the aim of I4.0 is to change not only the technologies but also the business model, especially for manufacturing SMEs (Müller et al., 2018; (Frank et al., 2019)). Therefore, all the elements supporting these transitions, such as *Service Oriented Architecture* and *Internet of Service* (but also *Platform as a Service* and *Data as a Service*, see Appendix B), should be included in ESCO too. The fact that these are missing, probably means that ESCO is focusing on the technological part of the fourth industrial revolution. This can be a drawback for representing competences that are related to the job profiles that are working on the business side of I4.0. Also, given the importance of the adoption of new digitally driven business models for SMEs, leaving these concepts outside is a missed opportunity to guide European smaller companies towards a more precise focus on business related skills. Here, recent methods able to extract transversal and soft skills from documents, can help ESCO to update the Ontology also considering this kind of concepts (Fareri et al., 2021).

To sum up the insights gained in this section we have shown that ESCO is in general in line with the 4.0 paradigm, containing many of the fundamental technologies of the fourth industrial revolution. Anyway, the framework is missing some important technologies such as *Cyber Physical Systems, Digital twin* and *Manufacturing Execution Systems*. Our system can help to spot such strong signals from the literature (but also weaker ones like *edge computing*), to foster a timely update of the frameworks, which is hard considering the ever-evolving characteristics of socio-technical revolutions like Industry 4.0.

4.2.2. Gaps of scientific literature

Table 7 shows the technologies that are found in the ESCO database but are not mentioned in the set of I4.0 papers under analysis. For each technology, we show the amount of skills/knowledge in ESCO versions 1.0 and 1.1.

First of all, this is an interesting and unexpected result. Our assumption was that all the I4.0 related technologies which are contained in ESCO should be mentioned also in scientific papers (at least in a few ones), given the different technological granularity to which the two types of documents can arrive (strong specificity of research, more high level of jobs frameworks). The only fact that this table is not empty is another signal of the possibility to positively answer both research questions on the timelines (B.1) and granularity (B.2) of ESCO, considering the limitations described in section 5.2.1. Anyway, it is interesting to analyse with a more focused lens these technologies, discussing the reasons why these are missing from scientific literature.

A first group of expressions in Table 7 can be reconducted to terminological issues in ESCO, already discussed in section 5.2.1. Terms such as *Automated Teller Machine, Automated Vehicle Locator, Chipset, Cracker Which Tests Security, Information Storage Technology*, are unusual in the typical I4.0 jargon and in the scientific community. We can explain this first group of missing technologies by the fact that ESCO sometimes uses a very specific language. This can create a problem, because using terms that are not "the state of the art" regarding technological concepts, can be a barrier to the consultation of the ontology.

Table 7
The list of technologies extracted using the Industry 4.0 lexicon (Chiarello, et al., 2018) from ESCO but not found in the set of papers related to Industry 4.0. For each technology, we show the amount of skills/knowledge in ESCO (version 1.0 and 1.1, respectively) that are referred to the technology under the analysis.

Technology	N. of skills v1.0	N. of skills v1.1	Technology	N. of skills v1.0	N. of skills v1.1
3D Computer Graphics	1	1	Linux Distribution	3	3
3D Display	3	3	Mac Os	2	2
3D Footwear Modelling	1	1	Microsoft Windows	1	1
3D Footwear Prototype	1	1	Optical Assembly Equipment	1	7
3D Imaging	1	1	Optical Equipment	7	1
Apache Tomcat	1	1	Palmar	1	1
Automated Teller Machine	2	2	Perl	1	1
Automated Vehicle Locator	1	1	Pyrometer	1	1
C Sharp	1	1	SMED	0	1
Cascading Style Sheets	2	2	Spread Network	1	1
Chipset	1	1	Unix	2	2
Cracker Which Tests Security	1	1	Visual Alignment Equipment	1	1
Document Management Procedure	0	1	Visual Appliance	0	1
Drupal	1	1	Visual Equipment	9	9
Haskell	2	2	Visual Peripheral Equipment	1	1
Information Storage Technology	1	1	Water Jet Cutter	1	1
Intelligent Safety System	0	1	Water Jet Cutting	2	2
iPhone	1	1	Windows Phone	2	2

A second group is composed of very specific technologies. These are, *Intelligent Safety System, Optical Assembly Equipment, Visual Alignment Equipment* and *Visual Peripheral Equipment*. Also, it is interesting to find technologies that are typical of a sector (e.g. Footwear) such as *3D Footwear Modelling* and *3D Footwear Prototype* or typical of a job profiles (e.g. Water Jet Operator) such as *Water Jet Cutter* and *Water Jet Cutting*. Considering the fact that some works have very specific skills and knowledge to manage, it is reasonable to have these technologies in ESCO not cited in scientific papers. In this case ESCO proved to be able to answer research question B.2, also for highly specific technologies.

A third group can be related to problems in the technologies' retrieval phase (see section 4.2). These are entries such as *C Sharp* (C#), *Cascading Style Sheets* (CSS) and *SMED* (Single Minute Exchange of Die). The lexicon proposed by Chiarello et (2018) did not contain also the acronyms or the long versions of these technologies, avoiding finding them also in scientific articles. We have an evidence of the appropriateness of the method for the proposed application thanks to the fact that there are only 3 entries (on a total list of technologies of 716).

Finally, we have a group of technologies related to another issue already discussed in section 5.2.1, that are specific product names and trade names. These are *Apache Tomcat, Drupal, Haskell, Iphone, Linux Distribution, Mac OS, Microsoft Windows, Perl* and *Windows Phone.* It is to be expected that these technologies will not be mentioned in scientific papers (they can be used only in specific case studies) but it is an open question why such trade names are contained in ESCO. For this group, the solution can be to decide policies on the inclusion/exclusion of a trade name in ESCO. Since this can advantage some companies, the criteria for using a product name in a public ontology should be clearly stated. This in turn, can help ESCO in having a more homogeneous granularity level regarding technologies.

To sum up the *insights* gained in this section, we have shown that not all of the I4.0 related technologies that are contained in ESCO are mentioned also in scientific papers. This evidence show how ESCO is in line with the technological advancement, and it has a high level of detail in some case. Anyway, this is also due to the fact that ESCO uses a very peculiar language, that sometimes speaks differently from the established scientific and technical language, typical of academic and industrial research. This can create a barrier to the consultation of the ontology both from researchers and practitioners. Again, a Natural Language Processing analysis such as the one proposed in the present paper can help the European Commission (and other competences frameworks' maintainers) to be in line with the evolving language of technology and science.

5. Conclusion

The analysis provided here demonstrates the extent to which ESCO captures the technologies and related skill needs associated with I4.0. If ESCO is to meet its ambition of providing a one-stop resource of information on the skills required to work in a given occupation, then it will need to comprehensively capture information on emerging technologies and how they affect the tasks undertaken in jobs. To achieve this aim, ESCO must have the most up to date information available. To date the evidence suggests that ESCO has had a degree of success here: it is partially in line with technology trends related to I4.0 (research question A.1) and there is a statistically significant improvement in alignment between ESCO v1.0 and v1.1 (research question A.2). So far, so good. But the analysis also reveals some information gaps that need to be filled (research question B.1 and B.2).

In the present section, we conclude the paper discussing some practical and theoretical implications of our work. Furthermore, we discuss the limitations of our method which must be acknowledged as well, especially by scholars who intend to use it.

5.1. Theoretical implications

The present paper has two main theoretical implications, linked to the novelties introduced with the proposed methodology and its application. First of all, we are presenting a first attempt to map ESCO to scientific literature. From a theoretical point of view, we are open to different experiments on other skills frameworks (such as O*Net, the United Stated Framework of skills and jobs) and other technological sources (such as patents). For these future studies, the use of Natural Language Processing has proven to be suitable. In fact, in the present paper, we give other evidence that thanks to the use of ad-hoc knowledge bases and Natural Language Processing tools, it is possible to properly analyse also complex text, as the case of scientific articles that have a jargon different from the every-day-life text, and skills that are written in a peculiarity way respect to usually analysed elements with NLP. Second, the present work opens to researchers in the fields of innovation management the possibility to study in deeper details the characteristics of 4.0 technologies, that makes them hard-to-be-mapped. From our results, it is evident that some pillars of I4.0 are missing in ESCO (e.g., *Cyber Phisical Systems, Digital Twin, Manufacturing Execution Systems*), limiting the timeliness update of ESCO to the I4.0 paradigm. Also, the current version of ESCO seems to manage correctly skills, lacking some of the important knowledges required for I4.0 (e.g., *Edge computing, Wireless Network, Radio Frequency Identification*). We hypnotize that some characteristics of 4.0 technologies like the velocity growth, or their convergence can be the cause of such lack in skills taxonomies.

5.2. Practical implications

The method can have practical applications for potentially every organisation that uses ESCO, but in particular the European Commission and companies that are using ESCO. Our work shows that scientific papers, a source typically far from data used for skill intelligence (e.g., job vacancies, curricula), is suitable to increase the quality of skills frameworks. Better skills frameworks (i.e., frameworks that reflect the reality of skills and jobs) can help policy makers address the *social issues* that are arising from the fast evolution of new technological paradigms. Given that the European Commission is exploring the use of artificial intelligence to gain insights on the evolution of the labour market, the present paper can shed light on the usage of alternative data to feed Artificial Intelligence systems. Also, considering that scientific literature is a validated and peer-reviewed source of knowledge, its usage can help mitigate biases that are typically introduced by other textual sources (e.g., the well-known phenomenon of gender biases in job vacancies).

Also, many companies are experimenting on the use of ESCO as a Knowledge base (as discussed in section 1), in order to face different business issues. In the industrial context, having rapidly updated information, is even more important than in the policy context. Our approach can help in this direction, since it gives an output where skill requirements are linked very much to specific technological developments, and the level of detail, or granularity, at which skill requirements can be stated is much greater than that obtained from other sources (e.g., data derived from probabilistic surveys of workers or employers). Furthermore, the use of scientific papers as source of data on technology provides an indication of likely emerging skill needs, another feature of particular interest for companies. Scientific papers are sometimes submitted for publication in advance of a technology becoming a mainstream one, also giving information that is up to date given the fact that the analysis can be easily repeated periodically. Furthermore, the replicability of our method makes it suitable to be applied in other domains with respect to I4.0 (e.g., green technologies, blue technologies, precision agriculture), opening to different industrial applications.

Decision makers may benefit of our results for enhancing the decision process related to the R&D strategies and adoption of future technologies, thanks to the link between the skill requirements and the specific technological developments. Moreover, they may use the results of this paper as a guide for defining the re-skilling or up-skilling strategies. In fact, knowing the emerging skill needs well in advance helps companies and policy makers to better allocate their resources based on the company needs.

5.3. Limitations

First, we give no indication of the scale of demand: it is impossible for our method to predict how many people will need to acquire a specific skill. Second, the representativeness of the findings is largely dependent upon the comprehensiveness of the sources analysed and quality of query to collect the documents to analyses. Finally, the relationships between technologies and skill needs are statistically rather than behaviourally defined. Employers may well have a degree of strategic choice about how the organise work around a given technology.

While the limitations of the approach are not trivial, they need to be regarded in the wider context of the process of skills anticipation. Increasingly there is no single source of data that can address the requirements of a wide variety of users of labour market skills intelligence, especially in the growing field of I4.0. Potentially these users include policy makers concerned with planning education and training provision, individual learners/students who want to identify the skills and qualifications required to enter an occupation or job, training providers looking to update their courses, etc.. In the present paper we shown that Natural Language Processing applied on different sources, can be used to help this wide range of users and meet their information needs.

CRediT authorship contribution statement

Filippo Chiarello: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. **Gualtiero Fantoni:** Conceptualization, Formal analysis, Funding acquisition, Investigation, Methodology, Supervision, Validation, Writing – review & editing. **Terence Hogarth:** Investigation, Validation, Writing – original draft, Writing – review & editing. **Vito Giordano:** Data curation, Formal analysis, Investigation, Methodology, Software, Visualization, Writing – original draft, Writing – review & editing. **Liga Baltina:** Investigation, Validation, Writing – original draft, Writing – review & editing. **Irene Spada:** Data curation, Investigation, Writing – original draft, Writing – review & editing.

Acknowledgments

This study was partly founded by the EU project ASSETs+ Project (Alliance for Strategic Skills addressing Emerging Technologies in Defence) EAC/A03/2018 - Erasmus+ programme, Sector Skills Alliances, Lot 3: Sector Skills Alliance for implementing a new strategic approach (Blueprint) to sectoral cooperation on skills G.A. NUMBER: 612678-EPP-1-2019-1-IT-EPPKA2-SSA-B.

We also thank the European Commission and the ESCO team for their work in making ESCO an accessible and valuable source of information.

Appendix

Appendix A

In the present paper we used a lexicon of Industry 4.0 technologies developed by Chiarello et al. (2018) composed of 1,211 technologies. Even if the authors in Chiarello et al. (2018) mapped also the relations between the technologies, for our method we used only the list of technologies, sticking to a flat structure.

In the present Appendix we list a sample of the technologies contained in the lexicon, to show its content. The 716 technologies listed in appendix B are extracted using this list (i.e., they are the subset of the 1,211 technologies that we retrieved from the papers and the ESCO versions). All the data is open access (Giordano et al., 2021).

1 **Big Data Technologies**: Virtual machine, Data mining, User interface, Computer vision, Virtual reality, Human–computer interaction, Supervised learning.
2 **Internet of Things Technologies**: Wireless sensor network, Arduino, OPNET, Telemetry, TinyOS, MiWi, LinuxMCE.
3 **Computing Technologies**: MacOS, Cloud computing, Microcomputer, Personal digital assistant, Tablet computer, ASCII.
4 **Programming Languages Technologies**: Python (programming language), Ruby (programming language), HTML, Hypertext Transfer Protocol, XML, Java (software platform), .NET Framework.
5 **Communication Network and Infrastructures Technologies**: Wi-Fi, Cellular network, Router (computing), Digital subscriber line, General Packet Radio Service, Global Positioning System, Evolution-Data Optimized.
6 **Protocols & Architectures Technologies**: 1-Wire, Profibus, Smart meter, Local Interconnect Network, Fleet Management System, Keyword Protocol 2000, RAPIEnet.
7 **Intel Technologies**: 3D XPoint, Intel SHA extensions, Intel Cluster Ready, Intel Mobile Communications, Intel Modular Server System, Intel PRO/Wireless, Intel Quick Sync Video.
8 **Identification Technologies**: Barcode, RFID, QR code, Mobile tagging, Code 128, High Capacity Color Barcode, Aztec Code.
9 **Transactions, Digital Certification, Digital Currency Technologies**: Bitcoin, Cryptocurrency, Digital currency exchanger, Ethereum, Monero (cryptocurrency), Namecoin, Blockchain, Ripple (payment protocol).
10 **Production Technologies**: 3D printing, Home automation, Agricultural robot, Nanorobotics, Smart grid, OLED, Computer-generated holography.
11 **Embedded Systems Technologies**: Programmable logic controller, Zilog Z80, CMOS, Zilog Z8, NEC μPD780C, MOS Technology 6502, Hitachi HD64180.

Appendix B

Table B1 ranks technologies according to the number of papers that mentioned it in the title, abstract or keywords from our set of papers related to Industry 4.0. The *Relative Development of Growth Rates* (RDGR) (Ernst H., 1998) investigates the growth rate of technologies in scientific papers. It is measured by calculating the number of papers published during the past three years (i.e. 2018–2020) divided by the number of papers published in a 9 years span (i.e. 2012–2020). Finally, the v1.0 and v1.1 columns are the numbers of skills/knowledge in ESCO version 1.0 and 1.1 respectively, that mentioned the technology under the analysis. All the data is open access (Giordano et al., 2021b).

Table B1
The list of technologies extracted using the Industry 4.0 lexicon (Chiarello, et al., 2018) with the number of papers related to Industry 4.0 and the amount of skills/knowledge in ESCO (version 1.0 and 1.1, respectively) that are referred to the technology under the analysis.

Technology	N. of papers	RDGR (%)	v1.0	v1.1	Technology	N. of papers	RDGR (%)	v1.0	v1.1
Internet Of Things	3176	83.69	2	2	Intelligent Multi Agent System	2	50	0	0
Cyber Physical System	1786	74.24	0	0	Intelligent Negotiation Mechanism	2	100	0	0
Big Data	1179	81.68	2	3	Intelligent Network System	2	100	0	0
Cloud Technology	1151	82.88	9	22	Intelligent Planning System	2	50	0	0
Simulation Technology	1109	87.74	20	28	Intelligent Prognostics Tool	2	100	0	0
Artificial Intelligence	1045	91.87	5	11	Intelligent Quality Control Systems	2	100	0	0
Machine Learning	772	94.17	3	7	Intelligent Service System	2	50	0	0
Robot	677	85.97	8	9	Laser Metal Deposition	2	50	0	0
Cloud Computing	619	83.36	1	1	Mobile Sensor Network	2	100	0	0
Digital Twin	544	93.75	0	0	Modular Assembly System	2	100	0	0
Augmented Reality	377	85.94	0	3	Modular Construction	2	100	0	0
Virtual Reality	346	79.77	1	4	Modular Manufacturing System	2	100	0	0
Neural Network	316	93.35	1	3	Multicasting	2	100	0	0
Led	274	86.5	8	10	Network Emulation	2	100	0	0
Blockchain	272	97.43	0	6	Network Gateway	2	50	0	0
Deep Learning	254	93.31	0	1	One Time Password	2	100	0	0
Data Mining	251	82.47	7	7	Predictive Model Markup Language	2	50	0	0
Cybersecurity	214	91.59	3	3	Quick Response Code	2	100	0	0
ICT Communication Protocols	207	80.19	3	3	Real Time Remote Monitoring	2	50	0	0
Human Machine Interface	201	83.08	0	1	Recording System	2	100	1	1
3D Printing	186	91.4	5	5	Reinforcement Learning Agent	2	100	0	0
Manufacturing Execution System	171	71.93	0	0	Remote Control	2	100	0	0
Wireless Sensor Network	167	83.83	0	1	Remote Terminal Unit	2	50	0	0
Computer Aided Manufacturing (CAM)	156	75.64	5	5	Restricted Boltzmann Machine	2	100	0	0
Data Processing	148	83.78	8	8	Semi Supervised Learning	2	100	0	0
Programmable Logic Controller	147	82.99	2	2	Smart Card	2	100	0	0
Radio-Frequency Identification	144	80.56	0	0	Smart Logistic Management System	2	100	0	0
Edge Computing	144	95.83	0	0	Smart Manufacture System	2	50	0	0
Computer Aided Design (CAD)	137	81.02	17	18	Smart Manufacturing Assembly System	2	100	0	0
Computer Vision	129	86.82	1	4	Smart Mechatronic System	2	100	0	0
Smart Grid	128	79.69	0	3	Smart Mobile Device	2	100	0	0
Decision Support System	125	91.2	2	2	Smart Parking System	2	50	0	0
Agent Based	121	85.95	0	0	Smart Personalized Product	2	100	0	0
iOS	121	76.86	3	3	Smart Predictive Maintenance System	2	100	0	0
Automatic Control	120	88.33	2	2	Smart Sensor System	2	100	0	0
Wireless Network	113	82.3	0	0	Smart Transport System	2	50	0	0
Virtualization	112	88.39	1	1	Super Capacitor	2	100	0	1
Autonomous Vehicle	109	91.74	0	1	Technology Forecasting	2	100	0	0
Embedded Systems	107	79.44	2	3	Transponder	2	100	0	0
Computer Numerical Control	103	77.67	14	14	Virtual Private Network	2	100	1	1
Mechatronic	101	73.27	10	11	Visual Sensor	2	50	0	0
Real Time Monitoring	99	88.89	0	0	Visual System	2	50	3	3
Laser	95	87.37	24	26	Widgets	2	100	0	0
Collaborative Robot	94	90.43	0	0	Wired Communication Network	2	100	0	0
Smart Device	91	87.91	2	3	Wireless Body Sensor Networks	2	100	0	0
Global Positioning System	88	93.18	12	12	Content Management System	1	100	3	3
Human Machine Interaction	84	82.14	1	2	3D Modeling	1	100	0	0
Cryptography	80	88.75	2	2	3D Prototype	1	100	0	1
Anomaly Detection	79	84.81	0	0	3D System	1	100	0	0
Middleware	78	79.49	1	1	Advanced Risc Machine	1	100	0	0
Scada	76	85.53	0	0	Agent Based System	1	100	0	0
Smart Manufacturing System	75	85.33	0	0	Agricultural Robot	1	100	0	0
Wide Area Network	73	93.15	1	2	Angularjs	1	100	0	0
Service Oriented Architecture	70	75.71	0	0	Apache Hadoop	1	100	0	0
Manipulator	69	85.51	0	0	Automated Robotic Arm	1	100	0	0
Drone	67	98.51	3	5	Automatedground	1	100	0	0
Enterprise Resource Planning	65	80	2	2	Automaton	1	100	0	0
Product Life Cycle Management	65	69.23	0	0	Capacitive Level Sensor	1	100	0	0
Ethernet	64	76.56	2	2	Capacitive Proximity Sensor	1	100	0	0
Mobile Robot	63	87.3	0	0	Central Processing Unit	1	100	2	2
User Interface	63	88.89	9	9	Cloud Robotic	1	100	0	0
Microcontroller	62	91.94	1	1	Communication Middleware Component	1	100	0	0
Access Control	57	87.72	5	6	Communication Process Components	1	100	0	0
Wearable Device	57	98.25	0	1	Communication Sensory System	1	100	0	0
Arduino	55	90.91	0	0	Computer Aided Architectural Design	1	100	0	0
3D Modelling	51	92.16	9	9	Dead Reckoning	1	100	0	0
Intelligent Manufacturing System	51	66.67	0	0	Dialogue System	1	100	0	0
Collaborative Robotic	49	89.8	0	0	Direct Deposition	1	100	0	0
Raspberry Pi	49	87.76	0	0	Direct Energy Deposition	1	100	0	0
Semantic Web	47	78.72	1	1	Direct Fabrication	1	100	0	0
Business Intelligence Software	46	80.43	2	2	Display Control System	1	100	0	0
Bluetooth	43	83.72	0	0	Distributed Agent Based Manufacturing	1	100	0	0

(continued on next page)

Table B1 (*continued*)

Technology	N. of papers	RDGR (%)	v1.0	v1.1	Technology	N. of papers	RDGR (%)	v1.0	v1.1
Field Programmable Gate Array	43	88.37	1	1	Distributed Database Network	1	100	0	0
Flexible Manufacturing System	43	83.72	0	0	Distributed Optical Network	1	100	0	0
Representational State Transfer	43	88.37	6	7	Distributed System	1	100	0	0
Machine Vision	42	90.48	0	0	Eeprom	1	100	0	0
Tablet	42	64.29	4	5	Embedded Microcontroller	1	100	0	0
Internet Of Service	41	73.17	0	0	Energy Router	1	100	0	0
Software As A Service	41	87.8	5	5	Ensemble Learning	1	100	0	0
Remote Monitoring	40	90	0	2	Executable Programs	1	100	0	0
Robotic Arm	39	89.74	4	4	Extranet	1	100	0	0
Distributed Manufacturing	33	60.61	0	0	Federated Access Control	1	100	0	0
Autonomous Robot	32	93.75	0	0	Flexible Manufacturing Assembly	1	100	0	0
Nanotechnology	32	93.75	3	3	Flexible Manufacturing Machines	1	100	0	0
Graphics Processing Unit (GPU)	31	96.77	1	1	Fortran	1	100	0	0
Machine To Machine (M2M)	30	83.33	2	2	Fused Filament Fabrication	1	100	0	0
Android Phone	29	96.55	3	3	Genetic Marker	1	100	0	0
Modbus	29	82.76	0	0	Gesture Controller	1	100	0	0
Gamification	28	89.29	0	0	Host Server	1	100	0	0
Application Programming Interface (API)	26	80.77	3	3	Inclinometer	1	100	0	0
Fuzzy Logic	26	92.31	0	0	Indian Regional Navigational Satellite System	1	100	0	0
Intelligent Machines	26	84.62	0	0	Industrial Automation Controller	1	100	0	0
Pattern Recognition	26	80.77	1	1	Industrial Standard Controller	1	100	0	0
Simulink	26	84.62	0	0	Industrial Utility Controller	1	100	0	0
Cognitive Computing	25	100	0	0	Information Modelling Technology	1	100	0	0
Cryptocurrency	25	92	0	0	Injection Modelling	1	100	0	0
Intrusion Detection System	25	92	1	1	Intelligent Adaptive Control System	1	100	0	0
Multisensor	25	92	0	0	Intelligent Agent System	1	100	0	0
Semantic Technology	25	64	1	1	Intelligent Alarm Management System	1	100	0	0
Storage System	25	88	1	2	Intelligent Asset Management System	1	100	0	0
Expert System	24	83.33	1	1	Intelligent Attack Detection System	1	100	0	0
Zigbee	24	87.5	0	0	Intelligent Autonomous Product	1	100	0	0
Graphical User Interface	23	86.96	1	1	Intelligent Bin System	1	100	0	0
Smart Meter	23	95.65	1	1	Intelligent Bothouse System	1	100	0	0
Smart Production System	23	95.65	0	0	Intelligent Building Management System	1	100	0	0
Ultra Wideband	23	91.3	0	0	Intelligent Computerised System	1	100	0	0
Profinet	22	77.27	0	0	Intelligent Computing System	1	100	0	0
System Dynamic	22	90.91	0	0	Intelligent Connected System	1	100	0	0
Intelligent Agent	21	76.19	1	1	Intelligent Container	1	100	0	0
Modularization	21	76.19	0	0	Intelligent Control Unit	1	100	0	0
Statistical Package For Social Science	21	100	0	0	Intelligent Cyber System	1	100	0	0
Cellular Network	20	95	0	0	Intelligent Distribution System	1	100	0	0
Ethereum	20	100	0	0	Intelligent Educational Management System	1	100	0	0
K Nearest Neighbor	19	94.74	0	0	Intelligent Energy Management System	1	100	0	0
Multiuser	19	89.47	0	0	Intelligent Energy System	1	100	0	0
Supervised Learning	19	94.74	0	0	Intelligent Factory Training System	1	100	0	0
Wirelesshart	19	84.21	0	0	Intelligent Field Device	1	100	0	0
3D Scanning	18	88.89	3	3	Intelligent Glove System	1	100	0	0
Biotechnology	18	83.33	5	5	Intelligent Hci Device	1	100	0	0
Broadcasting	18	88.89	33	33	Intelligent Industrial Architectural Component	1	100	0	0
Intelligent Product	18	61.11	0	0	Intelligent Industrial Iot System	1	100	0	0
Anova	17	88.24	0	0	Intelligent Information Management System	1	100	0	0
Barcode	17	88.24	1	1	Intelligent Inspection Terminal	1	100	0	0
Gate Array	17	88.24	0	0	Intelligent Instrument System	1	100	0	0
Microelectronic	17	88.24	9	13	Intelligent Integrated Management System	1	100	0	0
Sensing System	17	82.35	1	2	Intelligent Interconnection Between Product	1	100	0	0
Broadband	16	93.75	1	1	Intelligent Knowledge Based System	1	100	0	0
Crowdsourcing	16	87.5	1	1	Intelligent Learning Management System	1	100	0	0
Distributed Computing	16	68.75	1	1	Intelligent Logical Unit	1	100	0	0
Integrated Circuit	16	93.75	14	16	Intelligent Logistics Module	1	100	0	0
Intelligent Network	16	75	0	0	Intelligent Management System	1	100	0	0
Java	16	87.5	8	8	Intelligent Mantainance System	1	100	0	0
Sintering	16	100	1	1	Intelligent Manufacturing Control System	1	100	0	0
Ubiquitous Computing	16	75	1	1	Intelligent Manufacturing Device	1	100	0	0
Unsupervised Learning	16	87.5	0	0	Intelligent Manufacturing Information System	1	100	0	0
Bitcoin	15	93.33	0	0	Intelligent Manufacturing Product	1	100	0	0
Blueprint	15	80	6	6	Intelligent Mechatronic Component	1	100	0	0
Computer Integrated Manufacturing	15	86.67	0	0	Intelligent Mechatronic System	1	100	0	0
Contactless	15	86.67	0	0	Intelligent Medical Monitoring Device	1	100	0	0
Gripper	15	86.67	2	2	Intelligent Multi Sensory System	1	100	0	0

(*continued on next page*)

Table B1 (*continued*)

Technology	N. of papers	RDGR (%)	v1.0	v1.1	Technology	N. of papers	RDGR (%)	v1.0	v1.1
High Performance Computing	15	86.67	0	0	Intelligent Network Component	1	100	0	0
Mobile Phone	15	93.33	5	6	Intelligent Networking System	1	100	0	0
Predictive Analysis	15	100	0	0	Intelligent Networks Of Machine	1	100	0	0
Smart Glass	15	93.33	0	0	Intelligent Operation Planning System	1	100	0	0
Smart Home	15	93.33	0	2	Intelligent Patent Summarization System	1	100	0	0
System On Chip	15	86.67	0	0	Intelligent Patent Summary System	1	100	0	0
Beacon	14	85.71	2	2	Intelligent Plant Experimental System	1	100	0	0
Computer Network	14	78.57	5	5	Intelligent Polymeric Product	1	100	0	0
Gesture Recognition	14	78.57	0	0	Intelligent Predictive Maintenance System	1	100	0	0
Life Cycle Assessment	14	92.86	0	0	Intelligent Process System	1	100	0	0
Load Balancing	14	71.43	0	0	Intelligent Quality Management System	1	100	0	0
Qr Code	14	100	0	0	Intelligent Quality System	1	100	0	0
Virtual Machine	14	78.57	1	1	Intelligent Reinforcement System	1	100	0	0
Cloud Storage	13	76.92	1	1	Intelligent Searching Object	1	100	0	0
Edge Device	13	100	0	0	Intelligent Semiconductor Manufacturing System	1	100	0	0
Fieldbus	13	76.92	0	0	Intelligent Sensing System	1	100	0	0
Intelligent Transportation Systems	13	92.31	0	0	Intelligent Smart Process System	1	100	0	0
Malware	13	92.31	3	3	Intelligent Software Tool	1	100	0	0
Microprocessor	13	84.62	4	4	Intelligent Spindle	1	100	0	0
Multi Robot	13	76.92	0	0	Intelligent Storage System	1	100	0	0
Network	13	76.92	0	0	Intelligent Strengthening System	1	100	0	0
Optical Fiber	13	100	1	1	Intelligent Transfer Module	1	100	0	0
Sensing Devices	13	61.54	3	3	Intelligent Tutoring System	1	100	0	0
Structured Query Language (SQL)	13	92.31	6	6	Intelligent User Interface System	1	100	0	0
Data Warehouse	12	83.33	2	2	Intelligent Vehicle System	1	100	0	0
Encoder	12	100	1	2	Intelligent Vessel Support System	1	100	0	0
Life Cycle Cost	12	83.33	0	0	Intelligent Warehouse Management System	1	100	0	0
Virtual Assistant	12	50	0	0	Intelligent Welding System	1	100	0	0
Catalyst	11	100	2	2	Intelligent Workshop Products	1	100	0	0
Firmware	11	90.91	4	4	Interconnect Production Modules	1	100	0	0
Laser Sintering	11	100	0	0	Interconnector	1	100	0	0
Mobile Robotic	11	72.73	0	0	Internet Of Aircraft Things	1	100	0	0
Modular Production System	11	72.73	0	0	Internet Of Battlefield Things	1	100	0	0
Mtconnect	11	81.82	0	0	Internet Of Health Things	1	100	0	0
Protocol Stack	11	90.91	0	0	Internet Of Industrial Things	1	100	0	0
Teleoperation	11	90.91	0	0	Internet Of Manufacturing Things	1	100	0	0
Arima	10	100	0	0	Internet Of Measurement Things	1	100	0	0
Biosensor	10	80	0	0	Iot Protocol	1	100	0	0
Fused Deposition	10	100	0	0	K Nearest Neighbors Algorithm	1	100	0	0
Global System For Mobile Communications (GSM)	10	80	0	0	Lab On A Chip	1	100	0	0
Intelligent Control System	10	80	0	0	Laboratory Automation	1	100	0	0
Minicomputer	10	50	0	0	Laser Additive Manufacturing	1	100	0	0
Near Field Communication (NFC)	10	70	0	0	Laser Printing	1	100	2	2
Real-Time Locating System	10	100	0	0	Laser Scanner	1	100	0	0
Third Generation Partnership Project	10	100	0	0	Laser Welding	1	100	0	0
Warehouse Management System	10	100	3	4	Layer Chemical Vapor Deposition	1	100	0	0
Analysis Of Variance	9	77.78	0	0	Layer Deposition	1	100	0	0
Autonomous Robotic	9	100	0	0	Linked Open Data	1	100	0	0
Communication Mechanism	9	77.78	0	0	Location Based Routing	1	100	0	0
Communication Tool	9	77.78	3	3	Low Power	1	100	0	0
Digital Fabrication	9	77.78	0	0	Machine Code	1	100	0	0
Distributed Network	9	100	0	0	Matlab	1	100	0	0
Fleet Management	9	77.78	5	5	Mechanic Arm	1	100	0	0
Industrial Controller	9	88.89	0	0	Memory Address	1	100	0	0
Local Area Network	9	88.89	1	2	Memristor	1	100	0	0
Mesh Network	9	88.89	0	0	Mesh	1	100	0	0
Modular Design	9	66.67	0	0	Mixed Model Assembly	1	100	0	0
Photogrammetry	9	88.89	1	1	Mobile Telephone	1	100	0	0
Smart Factory System	9	88.89	0	0	Modular Building Construction	1	100	0	0
Speech Recognition	9	88.89	1	2	Modular Complex System	1	100	0	0
Autonomous System	8	75	0	0	Modular Control System	1	100	0	0
Bot	8	100	0	0	Modular Conveyor System	1	100	0	0
Computer Aided Process Planning	8	75	0	0	Modular Equipment	1	100	0	0
Cybernetic	8	100	1	1	Moodle	1	100	0	0
Digital Control	8	87.5	0	1	Mosfet	1	100	0	0
Laser Melting	8	100	0	0	Multiagent System	1	100	0	0
Microelectromechanical System	8	75	8	9	Multilayered Architecture	1	100	0	0
Network Layer	8	75	0	0	Multiprocessor System On A Chip	1	100	0	0
Profibus	8	75	0	0	Network Interface	1	100	0	0
Python	8	100	0	0	Networked Rotating Machine	1	100	0	0
Selective Laser Melting	8	100	0	0	Optical Measurement System	1	100	0	0

(*continued on next page*)

Table B1 (*continued*)

Technology	N. of papers	RDGR (%)	v1.0	v1.1	Technology	N. of papers	RDGR (%)	v1.0	v1.1
Selective Laser Sintering	8	100	0	0	Optical Recognition System	1	100	0	0
Smart Assembly	8	75	0	0	Personal Area Network	1	100	0	0
Touch Screen	8	75	0	0	Phase Shift Keying	1	100	0	0
World Wide Web	8	87.5	3	3	Polyjet Technology	1	100	0	0
Cyber Physical Space	7	100	0	0	Process Control	1	100	0	0
Dsl	7	85.71	0	0	Proknow C	1	100	0	0
Github	7	100	0	0	Pylons	1	100	1	1
Indoor Positioning System	7	100	0	0	Quantum Circuits	1	100	0	0
Intelligent Manufacturing	7	71.43	0	0	Quantum Computer	1	100	0	0
Javascript	7	85.71	6	6	Random Effects Model	1	100	0	0
Linux	7	85.71	5	5	Ransomware	1	100	0	0
Multicast	7	100	0	0	Rapid Control Prototype	1	100	0	0
Networked Manufacturing System	7	85.71	0	0	Rdf Query Language	1	100	2	2
Parallel Computing	7	71.43	1	1	Real Time Computing	1	100	1	1
Peer To Peer Network	7	100	0	0	Real Time Control System	1	100	0	0
Portable Device	7	100	0	0	Real Time Sdn Controller	1	100	0	0
Printed Circuit Board	7	100	11	11	Risc V Processor	1	100	0	0
Quantum Computing	7	100	0	0	Robotic Telescope	1	100	0	0
Smartwatche	7	71.43	0	0	Robotiq	1	100	0	0
Statistical Test	7	100	1	1	Robotnik	1	100	0	0
Autoregressive Integrated Moving Average	6	100	0	0	Rom	1	100	2	2
Capacitance Sensing	6	83.33	0	0	Selective Laser Sintered	1	100	0	0
Computer Aided Engineering	6	83.33	3	3	Semantic Search	1	100	0	0
Customer Relationship Management	6	100	5	5	Sensing Tool	1	100	0	0
Intelligent Production System	6	50	0	0	Slm Process	1	100	0	0
Intelligent Transport System	6	66.67	0	0	Smart System	1	100	0	0
Intranet	6	50	0	0	Smart Automatic Picking System	1	100	0	0
Li-Fi	6	100	0	0	Smart Automatic Production System	1	100	0	0
Lidar	6	83.33	0	1	Smart Automation Tools	1	100	0	0
Man Machine Interface	6	83.33	0	0	Smart Awareness In Assembly	1	100	0	0
Modular System	6	100	0	0	Smart Awareness Of Machines	1	100	0	0
Optical Communication System	6	83.33	2	2	Smart Box Device	1	100	0	0
Oracle	6	100	9	9	Smart Box Diagnosed Device	1	100	0	0
Pervasive Computing	6	83.33	0	0	Smart Buffer Control System	1	100	0	0
Reduced Instruction Set Computer	6	100	0	0	Smart Building Management System	1	100	0	0
Sensor Network	6	83.33	0	0	Smart Cctv Camera System	1	100	0	0
Sparql	6	83.33	2	2	Smart City Management System	1	100	0	0
Biomaterial	5	100	0	0	Smart City System	1	100	0	0
Chatbot	5	100	0	0	Smart Classroom System	1	100	0	0
Cyber Computational Space	5	20	0	0	Smart Command System	1	100	0	0
Document Management System	5	100	0	1	Smart Completeness Control System	1	100	0	0
Eclipse	5	80	2	2	Smart Connected Device	1	100	1	1
Electronic Data Interchange (EDI)	5	80	0	0	Smart Connected Object	1	100	0	0
Firewall	5	100	5	5	Smart Construction Object	1	100	0	0
Holographic	5	100	0	0	Smart Control Unit	1	100	0	0
Hypertext	5	60	0	0	Smart Cutting Tool	1	100	0	0
Infrastructure As A Service	5	100	0	0	Smart Data Management Module	1	100	0	0
Integrated Development Environment	5	60	2	2	Smart Distillation Unit	1	100	0	0
Layer By Layer	5	100	1	1	Smart Edge Iot Device	1	100	0	0
Mobile Computing	5	80	0	0	Smart Entry Management System	1	100	0	0
Optical Sensor	5	60	0	0	Smart Environment Application System	1	100	0	0
Photodiode	5	80	0	0	Smart Environments Challenges System	1	100	0	0
Photogrammetric	5	80	0	0	Smart Fertigation System	1	100	0	0
Relational Database Management System	5	60	3	3	Smart Field Device	1	100	0	0
Smart Industrial System	5	80	0	0	Smart Gear System	1	100	0	0
Universal Serial Bus (USB)	5	80	0	0	Smart Gesture Control System	1	100	0	0
Wireless Mesh	5	100	0	0	Smart Irrigation System	1	100	0	0
Advanced Message Queuing Protocol	4	100	0	0	Smart Knowledge Management System	1	100	0	0
Application Specific Integrated Circuit	4	100	0	0	Smart Lighting System	1	100	0	0
Authentication System	4	100	0	0	Smart Maintenance Tool	1	100	0	0
Compiler	4	100	5	5	Smart Making Product	1	100	0	0
Controller Area Network	4	100	0	0	Smart Management Module	1	100	0	0
I2c	4	100	0	0	Smart Management System	1	100	0	0
Intelligent Maintenance System	4	75	0	0	Smart Manufacturing Of Product	1	100	0	0
Json	4	100	0	0	Smart Manufacturing Vision System	1	100	0	0
Liquid Crystal Display	4	75	1	1	Smart Mechanism	1	100	0	0
Location Based Service	4	75	0	0	Smart Mobility System	1	100	0	0
Magnetometer	4	100	0	0	Smart Monitoring System	1	100	0	0
Mobile Ad Hoc Network	4	100	0	0	Smart Monitoring Tool	1	100	0	0
Multi Tier Architecture	4	50	0	0	Smart Network Device	1	100	0	0
Multicore Processors	4	75	0	0	Smart New Automation Tool	1	100	0	0
Multiprocessor	4	100	0	0	Smart Parking Tracking Module	1	100	0	0
Multithreading	4	100	0	0	Smart Personal Protective System	1	100	0	0
Optical Character Recognition	4	100	1	2	Smart Phone Device	1	100	0	0
Robotstudio	4	100	0	0	Smart Physical Object	1	100	0	0

(*continued on next page*)

Table B1 (*continued*)

Technology	N. of papers	RDGR (%)	v1.0	v1.1	Technology	N. of papers	RDGR (%)	v1.0	v1.1
Semantic Web Rule Language	4	100	0	0	Smart Pipe System	1	100	0	0
Smart Health Device	4	100	0	0	Smart Port Logistics System	1	100	0	0
Smart Product Service System	4	75	0	0	Smart Power Plant System	1	100	0	0
Smart Virtual Product	4	100	0	0	Smart Power Tool	1	100	0	0
Smartwatch	4	100	0	0	Smart Predictive Informatics Tool	1	100	0	0
Soft Computing	4	100	0	0	Smart Process Control	1	100	0	0
Stereolithography	4	100	0	0	Smart Producing And Assembly	1	100	0	0
Thermostat	4	75	2	3	Smart Production Tool	1	100	0	0
Vehicle To Vehicle	4	100	0	1	Smart Public Transportation System	1	100	0	0
Wireless Mesh Network	4	100	0	0	Smart Rack System	1	100	0	0
Anomalies Detection	3	100	0	0	Smart Recruitment System	1	100	0	0
Authentication Method	3	100	0	0	Smart Remanufacturing	1	100	0	0
Authentication Protocol	3	66.67	0	0	Smart Robot Tool	1	100	0	0
Biotechnological	3	100	1	1	Smart Safety System	1	100	0	0
Distributed Manufacturing Network	3	33.33	0	0	Smart Security House System	1	100	0	0
Embedded Controller	3	100	0	0	Smart Sensing Device	1	100	0	0
Finite State Machine	3	33.33	0	0	Smart Sensing Unit	1	100	0	0
Fleet Management System	3	100	0	1	Smart Service Device	1	100	0	0
Fuzzy Logic System	3	100	0	0	Smart Service System	1	100	0	0
Gesture Control	3	100	0	0	Smart Shop Floor Object	1	100	0	0
Humanrobotic	3	100	0	0	Smart Software System	1	100	0	0
Hypertext Transfer Protocol	3	100	0	0	Smart Spindle Component	1	100	0	0
Hypervisor	3	100	0	0	Smart Surveillance System	1	100	0	0
Information Based Technologie	3	66.67	0	0	Smart System Component	1	100	0	0
Instant Messaging	3	100	2	2	Smart Tracking System	1	100	0	0
Intelligent Condition Monitoring System	3	100	0	0	Smart Traffic Monitoring System	1	100	0	0
Intelligent Human Machine	3	100	0	0	Smart Transportation System	1	100	0	0
Intelligent Robotic System	3	100	0	0	Smart Troubleshooting Connected Device	1	100	0	0
Intelligent Technical System	3	100	0	0	Smart Ubiquitous Device	1	100	0	0
Laser Manufacturing	3	100	0	0	Smart Ubiquitous Object	1	100	0	0
Low Power Electronic	3	66.67	0	0	Smart Vision System	1	100	0	0
Metal Deposition	3	66.67	0	0	Smart Warehouse Management System	1	100	0	0
Metamaterial	3	100	0	0	Smart Waste Management System	1	100	0	0
Optical Fiber Sensor	3	100	0	0	Smart Water Management System	1	100	0	0
Optical System	3	33.33	3	3	Smart Water System	1	100	0	0
Platform As A Service	3	66.67	0	0	Smart Welding Station System	1	100	0	0
Real Time Locating System	3	100	0	0	Smart Welding System	1	100	0	0
Remote Management	3	100	0	0	Sophos	1	100	0	0
Shape Memory Alloy	3	100	0	0	Spss Statistic	1	100	0	0
Software Development Kit	3	66.67	0	0	Star Network	1	100	0	0
Supercomputer	3	100	0	0	State Machine	1	100	0	0
Uart	3	100	0	0	Statistic Test	1	100	0	0
Visual Analytics System	3	66.67	0	0	Steam Cracking	1	100	0	0
3D Manufacturing	2	100	0	0	Subnet	1	100	0	0
3rd Generation Partnership Project	2	100	0	0	Symbolic Artificial Intelligence	1	100	0	0
Ad-Hoc On-Demand Distance Vector	2	100	0	0	System Integration	1	100	0	0
Additive Fabrication	2	100	0	0	System On Programmable Chip	1	100	0	0
Ascii	2	50	0	0	Technology Forecast	1	100	0	0
Autoid	2	50	0	0	Tele Operative	1	100	0	0
Autoregressive Moving Average	2	100	0	0	Teleoperate	1	100	0	0
Communication Network System	2	100	0	0	Telerobotic	1	100	0	0
Compactrio	2	50	0	0	Telnet	1	100	0	0
Cyber Physical Assembly	2	100	0	0	Three Dimensional Scan	1	100	0	0
Cyborg	2	100	0	0	Three Layered Architecture	1	100	0	0
Data As A Service	2	50	0	0	Tibco	1	100	0	0
Digital Light Processing	2	100	0	0	Touch Trigger Probe	1	100	0	0
Distributed Sensor Network	2	100	0	0	Virtual Manufacturing Network	1	100	0	0
Dram	2	100	0	0	Virtual Sensor	1	100	0	0
Flash Memory	2	50	0	0	Virtual Server	1	100	0	0
Fuzzy Control System	2	100	0	0	Visual Analysis System	1	100	0	0
Fuzzy Rules	2	100	0	0	Visual Basic	1	100	1	1
Google Glass	2	50	0	0	Visual Inspection System	1	100	0	0
Hijacking	2	100	0	0	Visual Servoing System	1	100	0	0
Hybrid Additive	2	100	0	0	Wire Arc Additive	1	100	0	0
Instant Messages	2	100	0	0	Wireless Biosensor Network	1	100	0	0
Intelligent Diagnostic System	2	50	0	0	Wireless Cloud Network	1	100	0	0
Intelligent Logistic System	2	100	0	0	Wireless Edge Network	1	100	0	0
Intelligent Machine Tool	2	100	0	0	Wireless Personal Area	1	100	0	0
Intelligent Material Handling System	2	100	0	0	X86	1	100	0	0
Intelligent Monitoring System	2	100	0	0	Xbee	1	100	0	0
Zwave	2	100	0	0	Xen	1	100	1	1
District Heating	2	50	0	3	Xquery	1	100	1	1

References

Abdel-Basset, M., Chang, V., Nabeeh, N.A., 2021. An intelligent framework using disruptive technologies for COVID-19 analysis. Technol. Forecast. Soc. Change 163, 120431.
Acemoglu, D., Restrepo, P., 2020. Robots and jobs: Evidence from US labor markets. J. Polit. Econ. 128 (6), 2188–2244.
Acemoglu, D., & Autor, D. (2011). Skills, tasks and technologies: Implications for employment and earnings. In *Handbook of Labor Economics* (Vol. 4, pp. 1043-1171). Elsevier.
Adamuthe, A.C., Thampi, G.T., 2019. Technology forecasting: A case study of computational technologies. Technol. Forecast. Soc. Change 143, 181–189.
Ali, M.M., Doumbouya, M.B., Louge, T., Rai, R., Karray, M.H., 2020. Ontology-based approach to extract product's design features from online customers' reviews. Comput. Ind. 116, 103175.
Ahrens, V., 2012. Inflation industrieller Revolutionen. Prod. Manage. 17 (5), 30–31.
Azimi, S., Veisi, H., Fateh-rad, M., Rahmani, R., 2020. Discovering Associations Among Technologies Using Neural Networks for Tech-Mining. IEEE Trans. Eng. Manage.
Bag, S., Pretorius, J.H.C., Gupta, S., Dwivedi, Y.K., 2021. Role of institutional pressures and resources in the adoption of big data analytics powered artificial intelligence, sustainable manufacturing practices and circular economy capabilities. Technol. Forecast. Soc. Change 163, 120420.
Baker, M.G., Peckham, T.K., Seixas, N.S., 2020. Estimating the burden of United States workers exposed to infection or disease: a key factor in containing risk of COVID-19 infection. PLoS One 15 (4), e0232452.
Baumann, M., Domnik, T., Haase, M., Wulf, C., Emmerich, P., Rösch ., C., Weil, M., 2021. Comparative patent analysis for the identification of global research trends for the case of battery storage, hydrogen and bioenergy. Technol. Forecast. Soc. Change 165, 120505.
Ben-Haim, Y., Tom-Tov,, E., 2010. . A Streaming Parallel Decision Tree Algorithm. ournal of Machine Learning Research.
Benzidia, S., Makaoui, N., Bentahar, O., 2021. The impact of big data analytics and artificial intelligence on green supply chain process integration and hospital environmental performance. Technol. Forecast. Soc. Change 165, 120557.
Blazquez, D., Domenech, J., 2018. Big Data sources and methods for social and economic analyses. Technol. Forecast. Soc. Change 130, 99–113.
Bochtis, D., Benos, L., Lampridi, M., Marinoudi, V., Pearson, S., Sørensen, C.G., 2020. Agricultural workforce crisis in light of the COVID-19 pandemic. Sustainability 12 (19), 8212.
Bonaccorsi, A., Chiarello, F., Fantoni, G., 2020. Impact for whom? Mapping the users of public research with lexicon-based text mining. Scientometrics 1–30.
Bonacini, L., Gallo, G., Scicchitano, S., 2021. Working from home and income inequality: risks of a 'new normal'with COVID-19. J. Population Econ. 34 (1), 303–360.
Boselli, R., Cesarini, M., Marrara, S., Mercorio, F., Mezzanzanica, M., Pasi, G., Viviani, M., 2018. WoLMIS: a labor market intelligence system for classifying web job vacancies. J. Intell. Inf. Syst. 51 (3), 477–502.
Catal, C., Tekinerdogan, B., 2019. Aligning education for the life sciences domain to support digitalization and industry 4.0. Procedia computer science 158, 99–106.
Chiarello, F., Belingheri, P., Bonaccorsi, A., Fantoni, G., Martini, A., 2021. Value creation in emerging technologies through text mining: the case of blockchain. Technol. Anal. Strat. Manage. 1–17.
Chiarello, F., Bonaccorsi, A., Fantoni, G., 2020. Technical sentiment analysis. Measuring advantages and drawbacks of new products using social media. Comput. Ind. 123, 103299.
Chiarello, F., Trivelli, L., Bonaccorsi, A., Fantoni, G., 2018. Extracting and mapping industry 4.0 technologies using wikipedia. Comput. Ind. 100, 244–257.
Cimino, C., Negri, E., Fumagalli, L., 2019. Review of digital twin applications in manufacturing. Comput. Ind. 113, 103130.
Colombo, E., Mercorio, F., Mezzanzanica, M., 2019. AI meets labor market: Exploring the link between automation and skills. Inf. Econ. Policy 47, 27–37.
Dai, Z., & Callan, J. (2019). Deeper text understanding for IR with contextual neural language modeling. In *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 985-988).
Daim, T.U., Rueda, G., Martin, H., Gerdsri, P., 2006. Forecasting emerging technologies: Use of bibliometrics and patent analysis. Technol. Forecast. Soc. Change 73 (8), 981–1012.
De Smedt, J., le Vrang, M., Papantoniou, A., 2015. ESCO: Towards a Semantic Web for the European Labor Market. LDOW@ WWW.
Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.
Dingel, J.I., Neiman, B., 2020. How many jobs can be done at home? J. Public Econ. 189, 104235.
Ernst, H., 1998. Patent portfolios for strategic R&D planning. J. Eng. Tech. Manage. 15 (4), 279–308.
Fareri, S., Fantoni, G., Chiarello, F., Coli, E., Binda, A., 2020. Estimating Industry 4.0 impact on job profiles and skills using text mining. Comput. Ind. 118, 103222.
Fareri, S., Melluso, N., Chiarello, F., Fantoni, G., 2021. SkillNER: Mining and mapping soft skills from any text. Expert Syst. Appl. 184, 115544.
Felten, E., Raj, M., Seamans, R., 2021. Occupational, industry, and geographic exposure to artificial intelligence: A novel dataset and its potential uses. Strat. Manage. J.
Fernández-Sanz, L., Gómez-Pérez, J., Castillo-Martínez, A., 2017. e-Skills Match: A framework for mapping and integrating the main skills, knowledge and competence standards and models for ICT occupations. Comput. Standards Interfaces 51, 30–42.
Frank, A.G., Mendes, G.H., Ayala, N.F., Ghezzi, A., 2019. Servitization and Industry 4.0 convergence in the digital transformation of product firms: A business model innovation perspective. Technol. Forecast. Soc. Change 141, 341–351.
Frey, C.B., Osborne, M.A., 2017. The future of employment: How susceptible are jobs to computerisation? Technol. Forecast. Soc. Change 114, 254–280.
Garg, P., Gupta, B., Chauhan, A.K., Sivarajah, U., Gupta, S., Modgil, S., 2021. Measuring the perceived benefits of implementing blockchain technology in the banking sector. Technol. Forecast. Soc. Change 163, 120407.
Giabelli, A., Malandri, L., Mercorio, F., Mezzanzanica, M., Seveso, A., 2021. Skills2Job: A recommender system that encodes job offer embeddings on graph databases. Appl. Soft Comput. 101, 107049.
Giabelli, A., Malandri, L., Mercorio, F., Mezzanzanica, M., 2020. GraphLMI: A data driven system for exploring labor market information through graph databases. Multimedia Tools Appl. 1–30.
Giordano, V., Chiarello, F., Melluso, N., Fantoni, G., Bonaccorsi, A., 2021. Text and Dynamic Network Analysis for Measuring Technological Convergence: A Case Study on Defense Patent Data. IEEE Trans. Eng. Manage. https://doi.org/10.1109/TEM.2021.3078231.
González, L., García-Barriocanal, E., Sicilia, M.A., 2021. Entity Linking as a Population Mechanism for Skill Ontologies: Evaluating the Use of ESCO and Wikidata. Metadata and Semantic Research 1355, 116.
Heckman, J.J., 2000. Policies to foster human capital. Research in economics 54 (1), 3–56.
Han, J., Kamber, M., Pei, J., 2011. Data mining concepts and techniques third edition. The Morgan Kaufmann Series in Data Management Systems 5 (4), 83–124.
Han, S., Pool, J., Tran, J., & Dally, W. J. (2015). Learning both weights and connections for efficient neural networks. *arXiv preprint arXiv:1506.02626*.
Handel, M.J., 2016. The O*NET Content Model: Strengths and Limitations. Journal for Labour Market Research 49, 157–176.
Howell, D. C. (2012). *Statistical methods for psychology*. Cengage Learning.
Hu, K., Wu, H., Qi, K., Yu, J., Yang, S., Yu ..., T., Liu, B., 2018. A domain keyword analysis approach extending Term Frequency-Keyword Active Index with Google Word2Vec model. Scientometrics 114 (3), 1031–1068.
Huang, J.Y., 2016. Patent portfolio analysis of the cloud computing industry. J. Eng. Tech. Manage. 39, 45–64.
Huang, J.Y., Hsu, H.T, 2017. Technology–function matrix based network analysis of cloud computing. Scientometrics 113 (1), 17–44.
Jabbour, C.J.C., de Sousa Jabbour, A.B.L., Sarkis, J., Godinho Filho, M., 2019. Unlocking the circular economy through new business models based on large-scale data: An integrative framework and research agenda. Technol. Forecast. Soc. Change 144, 546-542.
Kamprath, M., Mietzner, D., 2015. The impact of sectoral changes on individual competences: A reflective scenario-based approach in the creative industries. Technol. Forecast. Soc. Change 95, 252–275.
Khanzode, A.G., Sarma, P.R., Mangla, S.K., Yuan, H., 2021. Modeling the Industry 4.0 adoption for sustainable production in Micro, Small & Medium Enterprises. J. Cleaner Prod. 279, 123489.
le Vrang, M., Papantoniou, A., Pauwels, E., Fannes, P., Vandensteen, D., De Smedt, J., 2014. Esco: Boosting job matching in europe with semantic interoperability. Computer 47 (10), 57–64.
Lee, K., Lee, S., 2013. Patterns of technological innovation and evolution in the energy sector: A patent-based approach. Energy Policy 59, 415–432.
Liu, X., Shin, H., Burns, A.C., 2021. Examining the impact of luxury brand's social media marketing on customer engagement: Using big data analytics and natural language processing. Journal of Business Research.
Lu, Y., 2019. Artificial intelligence: a survey on evolution, models, applications and future trends. Journal of Management Analytics 6 (1), 1–29.
Lumauag, R.G., 2019. Decision Support System for Personnel Selection. International Journal of Recent Technology and Engineering 8 (1), 177, 179.
Malandri, L., Mercorio, F., Mezzanzanica, M., Nobani, N., 2021. MEET-LM: A method for embeddings evaluation for taxonomic data in the labour market. Comput. Ind. 124, 103341.
Mangla, S.K., Kazancoglu, Y., Ekinci, E., Liu, M., Özbiltekin, M., Sezer, M.D., 2021. Using system dynamics to analyze the societal impacts of blockchain technology in milk supply chainsrefer. Transportation Research Part E: Logistics and Transportation Review 149, 102289.
Manning, C., & Schutze, H. (1999). *Foundations of statistical natural language processing*. MIT press.
Massarelli, N., Wozowczyk, M., 2009. European Union Labour Force Survey-Annual results 2009. Eurostat: Data in Focus 33.
Mazzei, D., Chiarello, F., Fantoni, G., 2021. Analyzing Social Robotics Research with Natural Language Processing Techniques. Cognitive Computation 1–14.
Melluso, N., Bonaccorsi, A., Chiarello, F., Fantoni, G., 2020. Rapid detection of fast innovation under the pressure of COVID-19. PLoS One 15 (12), e0244175.
Mirski, P., Bernsteiner, R., Radi, D., 2017. Analytics in human resource management the openskimr approach. Procedia computer science 122, 727–734.
Müller, J.M., Buliga, O., Voigt, K.I., 2018. Fortune favors the prepared: How SMEs approach business model innovations in Industry 4.0. Technol. Forecast. Soc. Change 132, 2–17.
Mun, C., Yoon, S., Raghavan, N., Hwang, D., Basnet, S., Park, H., 2021. Function score-based technological trend analysis. Technovation 101, 102199.
Nadeau, D, Sekine, S, 2007. A survey of named entity recognition and classification. Lingvisticae Investigationes 30 (1), 3–26.
Nam, T., 2019. Technology usage, expected job sustainability, and perceived job insecurity. Technol. Forecast. Soc. Change 138, 155–165.
PIAAC Numeracy Expert Group. (2009). PIAAC numeracy: A conceptual framework.

Paliokaitė, A., Martinaitis, Ž., Reimeris, R., 2015. Foresight methods for smart specialisation strategy development in Lithuania. Technol. Forecast. Soc. Change 101, 185–199.
Penmetsa, P., Adanu, E.K., Wood, D., Wang, T., Jones, S.L., 2019. Perceptions and expectations of autonomous vehicles–A snapshot of vulnerable road user opinion. Technol. Forecast. Soc. Change 143, 9–13.
Piatetsky-Shapiro, G, 1991. Discovery, analysis, and presentation of strong rules. Knowledge discovery in databases 229–238.
Ranaei, S., Suominen, A., Porter, A., Carley, S., 2020. Evaluating technological emergence using text analytics: two case technologies and three approaches. Scientometrics 122 (1), 215–247.
Raut, R.D., Mangla, S.K., Narwane, V.S., Dora, M., Liu, M., 2021. Big Data Analytics as a mediator in Lean, Agile, Resilient, and Green (LARG) practices effects on sustainable supply chains. *Transportation Research Part E: Logistics and Transportation Review* 145, 102170.
R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.
RStudio Team (2020). RStudio: Integrated Development for R. RStudio, PBC, Boston, MA URL http://www.rstudio.com/.
Reischauer, G., 2018. Industry 4.0 as policy-driven discourse to institutionalize innovation systems in manufacturing. Technol. Forecast. Soc. Change 132, 26–33.
Scuotto, V., Beatrice, O., Valentina, C., Nicotra, M., Di Gioia, L., Briamonte, M.F., 2020. Uncovering the micro-foundations of knowledge sharing in open innovation partnerships: An intention-based perspective of technology transfer. Technol. Forecast. Soc. Change 152, 119906.
Shakina, E., Parshakov, P., Alsufiev, A., 2021. Rethinking the corporate digital divide: The complementarity of technologies and the demand for digital skills. Technol. Forecast. Soc. Change 162, 120405.
Shi, W., Cao, J., Zhang, Q., Li, Y., Xu, L., 2016. Edge computing: Vision and challenges. IEEE internet of things journal 3 (5), 637–646.
Silge, J., Robinson, D., 2016. tidytext: Text mining and analysis using tidy data principles in R. Journal of Open Source Software 1 (3), 37.
Song, K., Kim, K., Lee, S., 2018. Identifying promising technologies using patents: A retrospective feature analysis and a prospective needs analysis on outlier patents. Technol. Forecast. Soc. Change 128, 118–132.
Srivastava, A., Patil, P.R., 2020. The future of education, knowledge and skills: The organisation for economic co-operation and development. Journal of Critical Reviews 10 (7), 1324–1330, 2020, Pages.
Stark R., Damerau T. (2019) Digital Twin. In: The International Academy for Production Engineering, Chatti S., Tolio T. (eds) CIRP Encyclopedia of Production Engineering. Springer, Berlin, Heidelberg.
Tortorella, G.L., Saurin, T.A., Fogliatto, F.S., Rosa, V.M., Tonetto, L.M., Magrabi, F., 2021. Impacts of Healthcare 4.0 digital technologies on the resilience of hospitals. Technol. Forecast. Soc. Change 166, 120666.
Trappey, A., Trappey, C.V., Hsieh, A., 2021. An intelligent patent recommender adopting machine learning approach for natural language processing: A case study for smart machinery technology mining. Technol. Forecast. Soc. Change 164, 120511.
Tubaishat, M., Madria, S., 2003. Sensor networks: an overview. IEEE Potentials 22 (2), 20–23.
Vaidya, S., Ambad, P., Bhosle, S., 2018. Industry 4.0–a glimpse. Procedia manufacturing 20, 233–238.
Vallet, F., Puchinger, J., Millonig, A., Lamé, G., Nicolaï, I., 2020. Tangible futures: Combining scenario thinking and personas-A pilot study on urban mobility. Futures 117, 102513.
Wang, L., Luo, G.L., Sari, A., Shao, X.F., 2020. What nurtures fourth industrial revolution? An investigation of economic and social determinants of technological innovation in advanced economies. Technol. Forecast. Soc. Change 161, 120305.
Xu, S., Hao, L., Yang, G., Lu, K., An, X., 2021. A topic models based framework for detecting and forecasting emerging technologies. Technol. Forecast. Soc. Change 162, 120366.
Yuan, X., Cai, Y., 2021. Forecasting the development trend of low emission vehicle technologies: Based on patent data. Technol. Forecast. Soc. Change 166, 120651.
Zeidmane, A., & Cernajeva, S. (2011 , April). Interdisciplinary approach in engineering education. *In 2011 IEEE Global Engineering Education Conference (EDUCON)* (pp. 1096-1101). IEEE.
Zezulka, F., Marcon, P., Vesely, I., Sajdl, O., 2016. Industry 4.0–An Introduction in the phenomenon. IFAC-PapersOnLine 49 (25), 8–12.

Filippo CHIARELLO is an Assistant professor at the School of Engineering, University of Pisa. His research focuses on the use of Natural Language Processing techniques for studying technological and HR-related phenomena.

Gualtiero FANTONI is an Associate professor at the School of Engineering, University of Pisa. His research focuses on data science and its applications for management practice. He coordinates several European projects on skills mapping and evaluation. He is co-founder of several university spin-offs in the fields of technology foresight and IoT.

Terence HOGARTH is a Professor at the University of Warwick Institute for Employment Research (IER) where he is responsible to a programme of work which addresses employers' human resource investment decisions. He has over 35 five years of experience of research issues on the relationship between education and training systems with the labour market and has published widely on this subject. Prior to joining IER he was Senior Adviser at Fondazione Giacomo Brodolini.

Vito GIORDANO is a Ph.D student at the School of Engineering, University of Pisa. His research focuses on the use of Natural Language Processing techniques for studying technological and Organizational-related phenomena.

Liga BALTINA is a Senior researcher and Manager at Fondazione Giacomo Brodolini (FGB). She has more than 15 years of experience in managing research projects at the European level, much of this in the field of EU employment, skills and innovation policies. She is currently leading FGB's work programme of comparative studies related to skills, skills governance and vocational education and training. She holds a Doctorate Degree in Management Science from the University of Latvia and an Executive MBA in Security, Defence and Space from the University of Côte d'Azur in France.

Irene SPADA is a Ph.D. student in Smart Industry Program at Pisa University. Her research focuses on the use of engineering design methods and text mining techniques to study the impact of the digitalization in education and training.