Journal of Manufacturing Systems 60 (2021) 692–706

Contents lists available at ScienceDirect

Journal of Manufacturing Systems

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

Data science skills and domain knowledge requirements in the manufacturing industry: A gap analysis

Guoyan Li, Chenxi Yuan, Sagar Kamarthi, Mohsen Moghaddam, Xiaoning Jin *

Department of Mechanical and Industrial Engineering, Northeastern University, Boston, MA 02115, United States

ARTICLE INFO

Keywords:
Industry 4.0
Labor market analysis
Skills gap
Data science

ABSTRACT

Manufacturing has adopted technologies such as automation, robotics, industrial Internet of Things (IoT), and big data analytics to improve productivity, efficiency, and capabilities in the production environment. Modern manufacturing workers not only need to be adept at the traditional manufacturing technologies but also ought to be trained in the advanced data-rich computer-automated technologies. This study analyzes the data science and analytics (DSA) skills gap in today's manufacturing workforce to identify the critical technical skills and domain knowledge required for data science and intelligent manufacturing-related jobs that are highly in-demand in today's manufacturing industry. The gap analysis conducted in this paper on Emsi job posting and profile data provides insights into the trends in manufacturing jobs that leverage data science, automation, cyber, and sensor technologies. These insights will be helpful for educators and industry to train the next generation manufacturing workforce. The main contribution of this paper includes (1) presenting the overall trend in manufacturing job postings in the U.S., (2) summarizing the critical skills and domain knowledge in demand in the manufacturing sector, (3) summarizing skills and domain knowledge reported by manufacturing job seekers, (4) identifying the gaps between demand and supply of skills and domain knowledge, and (5) recognize opportunities for training and upskilling workforce to address the widening skills and knowledge gap.

1. Introduction

The fourth industrial revolution is on its way. Many businesses are currently implementing Industry 4.0, thriving on emerging trends in automation, the industrial Internet of things (IoT), big data, and cloud computing technologies [1]. With the advent of the industrial IoT, sensors in networked physical devices collect an unprecedented amount of data in real-time to enable manufacturing operations, processes, and systems to achieve significantly higher productivity, efficiency, and self-management [2,3]. The manufacturing industry requires a command of data analytics techniques, machine learning, and artificial intelligence (AI) [4–6]. Thus, a key question in this context has been: How do we upskill the manufacturing workforce with the necessary skills and domain knowledge to succeed in this new manufacturing era?

1.1. Motivation

Industry 4.0 brings changes to the labor market structure and the demand for prospective workforce skills and knowledge. The global workforce will be impacted by the adoption of AI, automation, and big data analytics. Recruiting traditional machine operators and technicians is no longer sufficient in most manufacturing and engineering businesses. According to Thomas and Rupesh [7], the current and projected skills gap in manufacturing is anticipated to leave an estimated 2.4 million manufacturing positions unfilled by 2028. A recent Deloitte-Manufacturing Institute study further suggests that industries are entering a period of long-term labor shortage, resulting in a $2.5 trillion negative impact on the U.S. economy. Dimitris's survey [8] identifies five critical issues associated with labor shortage: 1) an aging workforce, 2) outdated workforce planning, 3) less efficiency of national education, 4) poor perception of manufacturing among the young generation, and 5) the changing nature of work.

Smart manufacturing requires current workers to have a good understanding of data science-related skills and knowledge as applied to manufacturing [9]. Data science is an interdisciplinary field focusing on extracting actionable insights from data to solve problems in a broad range of application domains [10]. Foundations of data science include mathematics, statistics, inference, computer science, optimization,

* Corresponding author.
E-mail address: xi.jin@northeastern.edu (X. Jin).

https://doi.org/10.1016/j.jmsy.2021.07.007
Received 4 April 2021; Received in revised form 24 June 2021; Accepted 5 July 2021
Available online 2 August 2021

machine learning, and data management. Professionals such as data scientists, data analysts, and data engineers use techniques and theories drawn from these fundamental fields to solve practical problems [11]. By 2020, according to Miller and Hughes [12], data and analytics-related positions in the U.S. will increase by 364,000. The manufacturing sector has the third highest demand for these roles, only behind the professional services and finance sectors. A key finding of their study is that 12 % of all positions in the manufacturing industry require data science skills. The rapid growth of big data and its potential impact on manufacturing have driven a data revolution in many sectors such as supply chain, defense, biomedical, and healthcare [13]. Talents with interdisciplinary skills have gradually become targets for recruiters in the manufacturing industry. Future manufacturing jobs will require employees to master a portfolio of core manufacturing science and technology skills, manufacturing domain knowledge, computational skills associated with data science. Yet, neither the generic nor business-specific data science skills gap has been fully addressed by the manufacturing education and training programs. Therefore, there is an urgent need to upskill the incumbent manufacturing workforce and train the future manufacturing workforce on the emerging advanced data-rich computer-automated technologies of the future factories. For future workforce preparation, universities and community colleges must reimagine their curricula in concert with the emerging Industry 4.0 and data science technologies.

1.2. Objectives and outline

The overarching goal of this paper is to identify the gaps in critical skills and domain knowledge that hinder the digital transformation of the manufacturing sector. This comprehensive gap analysis can be used to inform and guide the efforts of academia, government, and industry to develop necessary workforce training initiatives for the transforming manufacturing sector. The study is based on data collected from the Emsi labor market analytics database [14]. The study utilizes job posting data (i.e., demand for skills/domain knowledge) and professional profiles (i.e., supply of skills/domain knowledge) in the U.S. manufacturing industry to understand the type, severity, and significance of the gaps associated with skills and domain knowledge in the manufacturing sector. The study has four major steps:

1) Extract the comprehensive job posting data in the U.S. manufacturing labor market, including job titles, required educational level, skills and domain knowledge.
2) Identify the details of the skills and domain knowledge required by employers (i.e., demand) and the skills and domain knowledge available in online professional profiles such as LinkedIn (i.e., supply).
3) Identify potential gaps associated with each skill or domain knowledge topic in the recent manufacturing job market, and the type, severity, and significance of each identified gap.
4) Create a new understanding of the opportunities and priorities for training and upskilling the manufacturing workforce in data science to bridge the gaps identified.

The remainder of the paper is organized as follows. Section 2 discusses the state-of-the-art literature on the topic of skills and domain knowledge requirements in data science enabling manufacturing. Section 3 presents the study methodology and data description. Section 4 demonstrates the analysis and key findings of the current labor market, including overall trends, job requirement variation, and existing gaps of skills and domain knowledge. Section 5 presents insights from the current labor market and skills gap analysis. Section 6 summarizes the key findings and provides directions for future research.

2. Related work

The skills gap analysis for the manufacturing industry in the era of Industry 4.0 has received significant attention from academia in recent years. The National Association of Manufacturers (NAM) conducted a survey of around 1000 manufacturing company leaders in 2011 and 2014. The survey results indicate that an overwhelming majority of manufacturing company leaders—83 % in 2011 and 79 % in 2014—believe there is a moderate to severe shortage of skilled production workers in the U.S. [15]. Another study [16], focusing on middle-skill jobs that typically need less than a bachelor's degree, provides a picture of potential skill gaps in different regions across the U.S. Burning Glass Technologies [17] reports that manufacturing employers uniformly cite a considerable shortage of manufacturing workforce with the skills and qualifications they need, indicating a misalignment between the skills possessed by workers and the skills needed by employers.

The literature on the topic of the skills gap in the manufacturing workforce can be divided into two major categories: manufacturing science and technology skills gap and data science-related skills gap. The former category focuses on identifying trends in manufacturing skills requirements such as material processing, production operations, process control, and quality assurance. The latter category focuses on trends in data analytics, data management, big data, cloud computing, machine learning in manufacturing.

After decades of emigration, manufacturing has been making a comeback in the U.S. Since the Great Recession, production jobs have increased, and so have the skills and knowledge requirements. Domestic manufacturing is increasingly looking like a high-tech industry that needs a smaller and specialized workforce.

An overview of recent studies on the manufacturing skills gap is provided in Table 1. These studies, which were published between 2015 and 2020, cover many regions, including the U.S., Europe, Asia (especially China and India), and Australia. Berger and Frey [18] report that 25 % of workers in Europe have zero or low digital skills. They also report that disadvantaged regions in the world are more likely to be negatively impacted by the rapid growth in automation and artificial intelligence technologies. They recommend that the diffusion of basic digital skills across EU-28 is a key priority. Similarly, Moldovan [19] concludes that workers need to develop skills in a set of new digital technologies to fit into the Industry 4.0 environments. Chenoy et al. [20] suggest that the impact of next-generation digital technologies is evident in India's automotive, textile, and apparel sectors. Similar studies conducted in the U.S [21,22] emphasize skills that require industry certification but not necessarily a bachelor's degree.

Most of the aforementioned skills gap studies are based on surveys and interviews with manufacturing workers or executives. Using Portuguese manufacturing data from 2007 to 2014, Nogueira et al. [28] built a linear regression model to estimate the importance of different skills and identify the wage gap between the labor force with tertiary education and workers with primary and secondary education. Vista [21] uses a centrality-metrics method based on graph theory to find what skills are needed for the 21st century manufacturing workforce. Xu et al. [26] develop a Skills-Net framework following a heuristic method used in social network analysis to measure the popularity of job skills in the Chinese IT community.

For these survey-based studies, content analysis is a commonly used data-driven method used for job market skills requirement analysis, which uses quantitative methods that help identify relevant information and enumerate the number of occurrences of a specific word or phrase. Several studies [18,19,21,22] adopt web-scraper technologies to collect job posting data from public workforce websites such as LinkedIn and Indeed, which provide comprehensive job market data without bias. Some other studies [14,16,23] apply the latent Dirichlet allocation method for content analysis on job market data. These data-driven analysis methods can help identify topics in any document by

Table 1
Summary of previous studies on manufacturing skills gaps/requirements literature.

Study	Key findings	Dataset	Methodology	Research Target / Region
Regional analysis on addressing the skills gap [22].	• The middle skills gap has gained particular attention. • Strong partnerships among private and public leaders may be required to grow the talent community.	The U.S. Bureau of Economic analysis.	Unemployment analysis.	The Northwest Indiana, U.S.
Skills for 21st century workforce [21].	The skills needed now and in the future (current and future demand).	O*Net version 23.2 Database.	Various centrality metrics-based graph theory.	Recent engineering graduates in U.S.A.
Employability skills lacking in the STEM industry [23].	A list of 16 essential skills for employability in the five manufacturing industries.	Survey responses of 250 Human Resource Managers.	Survey response exploration.	STEM educational institutions in the U. S.
Employee competences in manufacturing companies [24].	Competences relevant to bridging the gap between academics and practice as evaluated by managers.	An online survey of German public companies.	Survey response exploration.	Manufacturing companies in Germany.
Filling the skills gap in U.S. manufacturing [25].	• A unique paradigm for integrating design and manufacturing skills. • Integration of two industrial engineering courses through practical course projects.	Spring 2014 survey of industrial engineering students at Penn State University Park.	Survey response exploration.	Industrial engineering students.
Measuring the popularity of job skills in recruitment market [26]	A novel Skill Popularity based Topic Model (SPTM) for integration of different job criteria to rank job skills.	Job posting dataset from the online recruiting market.	"Skill-Net" following the heuristic method used in social network analysis.	IT community in China.
Skill demands and mismatch in U.S. manufacturing [27].	• Three-quarters of U.S. manufacturing plants show no sign of hiring difficulties. • Estimated upper bound on potential skills gap.	Survey of U.S. manufacturing establishments.	Survey response exploration.	The manufacturing industry in the U.S.
Skill premium in Portuguese manufacturing industry [28].	Explanation of the wage gap between the labor force with tertiary education (skilled) and secondary/ primary education (unskilled).	Portuguese manufacturing data between 2007 and 2014.	Skills importance estimation following a linear regression model.	Manufacturing industries in Portuguese.
Strategies for closing the skills gap [18].	• 25 % workers in the EU-28 have none or low digital skills. • Disadvantaged regions are more likely to be negatively affected by automation.	Multiple EU databases. European Skills and Jobs (ESJ) survey.	Survey response exploration.	European manufacturing industry.
Developing the right skills to address the growing skill gap in various manufacturing sectors in India [20].	• The impact of new-age technologies is evident in India's automotive, textile, and apparel sectors. • Emerging job roles in the manufacturing sector, such as machine learning-based vehicle cybersecurity experts.	White papers about India manufacturing.	A white paper and survey exploration.	India manufacturing industry.
Closing the advanced manufacturing talent gap [29].	A 12-month course curriculum to transform students seeking skills for a career transition to the advanced manufacturing sector.	Prospective students from Massachusetts.	Survey response exploration.	Community college students.
Current skills gap in manufacturing [8].	• Root causes of skills gap. • Future skillsets adopted in Europe.	European public dataset.	Data exploration and data visualization.	European countries.
Developing the next generation of engineers for intelligent and sustainable manufacturing [30].	Manufacturing education needs to create a holistic-engineer profile from a system perspective.	"Experiencing I-Design" project from Jan. 2010 – May 2012.	Survey response exploration and curriculum development.	Manufacturing education in Italy.
Skills gap on the topic of Industry 4.0 and the requirements for work-based learning [19].	• Only one or two skills are not enough when it comes to Industry 4.0, but a set of skills is necessary. • Proposed iNduce 4.0, a practice-focused training course.	Survey from 6 countries in Nov. 2017-Jan. 2018.	Training module design.	Six European countries.
Jobs and skills in Industry 4.0 [31].	The latest needs of manufacturing in terms of Industry 4.0 skills.	Interviews with 70 participants.	Survey response exploration	European manufacturing industry.
Challenges and roadmaps in bridging the skills gap of works in Industry 4.0 [32].	Four main research domains which are expected to contribute to bridging the skills gap of workers in Industry 4.0.	White papers about Industry 4.0.	Skills metrics analysis.	Workers in Industry 4.0.

observing all the words and generating a topic distribution.

Even after being informed by several skills gap analysis studies, manufacturing companies struggle to find employees with the right mix of technical and business skills [34]. There is no comprehensive data-driven analysis on job market data that addresses both skills and domain knowledge requirements (demand) and professional profiles (supply). To create a more precise and insightful understanding of preparing the manufacturing workforce for the ongoing industrial revolution, we present a comprehensive job market data-based analysis of Industry 4.0 related skills and domain knowledge in the U.S. manufacturing industry. This current analysis is motivated by four major limitations in the recent manufacturing skills gap analysis studies in the literature: (1) surveys and interviews are limited to structured, predefined questions and cannot fully reflect a comprehensive status of the labor market; (2) current studies have predominantly focused on the recruiter-side data (i.e., demand data), disregarding the skillsets offered by the incumbent and future workforce; (3) there is a lack of a comprehensive skills gap research covering various aspects of industry 4.0 including cyber manufacturing, cloud computing, and sensor engineering; and (4) there is an absence of data science-related skills gap analysis on the current manufacturing industry job market.

3. Methodology

This section presents the methodology for understanding the gaps between the skills and domain knowledge requirements listed by manufacturers on their job postings and the relevant skills and domain knowledge reported in professional profiles by job seekers. The data for this analysis is obtained from Emsi database.

3.1. Definitions

Skills are multi-dimensional and typically represented in three main categories [35]: occupation- or industry-specific skills, academic skills (e.g., STEM), and non-cognitive or interpersonal skills. The occupation- or industry-specific skills and academic skills, which are teachable and measurable abilities, are generally defined as *hard skills,* while non-cognitive or interpersonal skills are defined as *soft skills*. This study exclusively focuses on the hard skills associated with data science in the manufacturing industry. We propose to divide occupation- or industry-specific competencies into two categories: skills and domain knowledge, where "skills" refers only to hard skills (occupation- or industry-specific skills). In contrast to general knowledge, *domain knowledge* refers to the knowledge of a specialized discipline or field, such as software engineering and manufacturing [36]. Requirements for hard skills and domain knowledge are listed on job postings. For example, employers in the data science job market often require applicants to have experience in programming languages such as R, Python, or SQL, as well as knowledge in machine learning. In the rest of the paper, “hard skills” and “skills” are used synonymously. Further, we define a *skills gap* as the mismatch between the needs of employers for skilled talent and the skills offered by the workforce available in the job market. We propose a quantitative method to measure the skills gap between the employer side and the workforce side. In this context, *demand* is defined as the occurrence frequency (count) of a particular skill in online job postings, and *supply* is defined as the occurrence frequency (count) of skill in workforce profiles. The skills gap is thus measured as the difference between the demand and supply.

3.2. Preliminary for data source

Emsi labor market analytics database provides two comprehensive data sets. The first one is Job Posting Analytics (JPA), which covers roughly 165.6 million online job postings from over 130 different online sources (e.g., LinkedIn, Indeed) in the U.S. It provides details about skills requirements in the current labor market. The second one is Profile Data (PD), which provides public self-reported information about individuals’ job history, education history, and skills. To handle the massive volume of job postings and profiles, Emsi applies built-in data mining methods to identify labor market patterns and frequency of occurrences of specific skill-describing keywords/phrases. Emsi database supports customized search by applying different filters. Results may be filtered by keywords in the full job posting text. In this paper, the job posting data and profile data are filtered, ranked and sorted by skills, job title, region, timeframe, and source industry. Another unique feature of the Emsi job posting database is that it is up-to-date by retrieving real-time labor market data. Note that the number of postings is not equal to the actual number of hires or the actual number of job positions. Postings could outnumber hires when a company is trying hard to find talent. In contrast, posting could be significantly lower than the number of hires in certain job types (e.g., blue-collar jobs) because these positions are not typically advertised online. In addition, the same job posting may be posted on multiple websites, Emsi JPA function removed duplicates across sources. The Emsi Profile Data (PD) database, which has about 143.4 million U.S. profiles, covers about 86 % of the workforce in the U.S. Considering the nature of the data described above, we assume that the Emsi data provides a realistic representation of the distributions of skills and domain knowledge topics in job postings and online profiles.

3.3. Analysis framework

We collected the 2017–2020 data of skills and domain knowledge requirements from the JPA module and the professional profiles data from the PD module of the Emsi database. The Emsi’s hierarchical data extraction filters enable us to design a sequential data acquisition scheme, as presented in Fig. 1. The overall analysis framework consists of four main steps as described next.

3.3.1. Step 1: define technology clusters

The clustering of technological advancement in the manufacturing labor market is driven by the nine pillars of Industry 4.0:1) Cyber-physical systems, 2) Internet of Things (IoT), 3) Big data, 4) Robotics, 5) Advanced processing technology, 6) Simulation, 7) Augmented reality, 8) Cloud computing and, 9) Cyber security [37–41]. As it evolves, Industry 4.0 is tending towards automation and data exchange in manufacturing processes to leverage artificial intelligence to improve reliability, flexibility and productivity [5,42]. However, smart manufacturing would not be possible without sensors [43].

We group these nine pillars into four technology clusters: 1) Data Science, 2) Automation, 3) Cyber, and 4) Sensors. Fig. 2 shows the members of the main pillars in the four clusters. These clusters have been identified based on the expertise and judgment of the authors and did not necessarily reflect a universal categorization. For each cluster, we identified specific keywords to retrieve skills and qualifications from the Emsi database.

- *Data Science Cluster*: This technology cluster includes skills and qualifications containing "data science." The five data science-related keywords for this cluster are “data mining,” “data management,” “big data,” “programming languages,” and “mathematical skills.” It should be noted that data science skills are not just limited to jobs with those keywords in their titles. Other jobs such as laser engineer and biomedical engineer may also require data science-related skills. Maxwell [12] has shown that engineers represent a tapped pool of talent for data science skills.
- *Automation Cluster*: This technology cluster includes skills and qualifications containing keywords "automation" and "robotics." The skills relevant to automation and robotics pave the way for traditional manufacturing workers to improve their skills and career transformation.
- *Cyber Cluster:* This technology cluster includes skills and qualifications containing the keywords "cyber" and “*cloud*.” These skills predominantly revolve around the Internet of things, cloud computing, cybersecurity.
- *Sensor Cluster.* This technology cluster includes skills and qualifications containing keywords "sensor" and "signal." These skills are mostly related to sensor analytics and signal processing.

3.3.2. Step 2: extract skills and qualifications

For each technology cluster, we apply the keyword filter to the Emsi JPA module to generate related skills and qualifications. Table 2 provides examples of the skills and qualifications extracted for each technology cluster. Here, we limit our search space to the manufacturing industry by selecting codes 31 through 33 in the North American Industry Classification System (NAICS). This highest-level filter setting ensures that all data we collected comes from manufacturing companies and manufacturing workers.

3.3.3. Step 3: extract related job titles

In this step, we collected titles of the job postings that include the cluster-specific skills and qualifications extracted in Step 2. Job titles

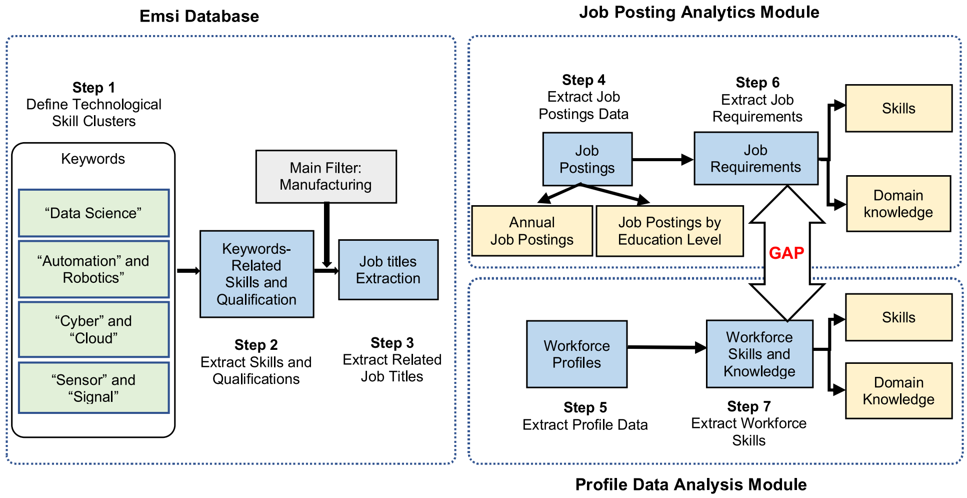

Fig. 1. Overall data collection and processing scheme.

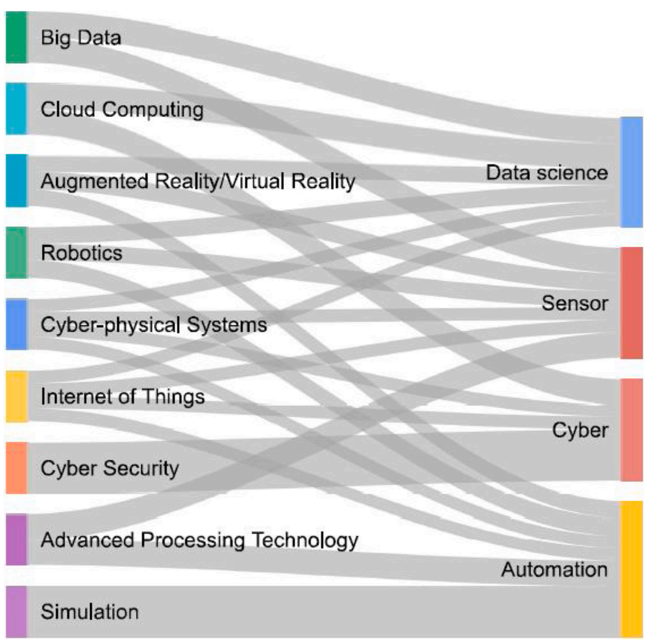

Fig. 2. Four technology clusters classification.

that appear fewer than 1000 times in the database during 2017–2020 were excluded from the analysis. The job titles extracted in each technology cluster are listed in Table 3.

3.3.4. Steps 4 and 5: extract job postings and profiles

In this step, we use the job titles in the Emsi JPA and PD for conducting a gap analysis between the demand and supply, respectively. We collected both the demand and supply data (e.g., number of job postings and number of profiles that appear in the database) from 2017 to 2020 at two different levels: (1) the U.S. national level, and (2) the state level to facilitate our skills gap analysis at both national and regional levels. Further, the study also takes into consideration the geographical distribution of job postings as well as the required educational level for different jobs.

3.3.5. Steps 6 and 7: extract job requirements and workforce skillsets

For each job posting and profile, we extracted skill requirements, including the total number of job postings, the total number of relevant profiles/resumes, and the top 50 skills based on the frequency of their occurrence in each job posting category. We performed a gap analysis by comparing the frequency of a given skill in the demand and supply data.

Table 2
Skills and qualifications for different technology clusters.

Technology Cluster	Keyword	Number of Skills and Qualification Retrieved	Example of skills and qualifications
Data Sciences	Data Mining	10	Data Mining, Text Mining, Forecasting, Machine Learning
	Data Management	11	SQL, Relational Databases, Oracle Databases
	Big Data	11	Spark, MapReduce, NoSQL
	Programming languages	11	Python, C#, C++, VB, PERL, C, Java
	Mathematical skills	12	Statistics, Hypothesis Testing, Operation Research, Probability
Automation	Automation	93	Manufacturing Automation, Logistic Automation, Process Automation
	Robotics	17	Robot Framework, Robot Welding, Robotic Programming
Cyber	Cloud	54	Cloud Computing, Cloud Application, Cloud Collaboration
	Cyber	16	Cyber Resilience, Cyber Engineering, Cyber Defense
Sensor	Sensor	16	Sensor Fusion, Image Sensor, Wireless Sensor Network
	Signal	49	Signal Conditioning, Signal Compression, Signal handling

4. Analysis and discussion

This section elaborates on the trends in skills and domain knowledge requirements, educational requirements, emerging/declining skills, and the existing gaps between demand and supply of skills and domain knowledge in the manufacturing industry.

Table 3
List of job titles in each technology cluster.

Data Science	Automation	Cyber	Sensor
• Data Analysts	• Manufacturing Engineers	• Software Engineers	• Software Engineers
• Data Scientists	• Mechanical Engineers	• Systems Engineers	• Systems Engineers
• Data Engineers	• Electrical Engineers	• Electrical Engineers	• Electrical Engineers
• Database Administrators	• Software Development Engineers in Test	• Cloud Security Architects	• Algorithm Engineers
• Big Data Engineers	• Automation Controls Engineers	• Cybersecurity Engineers	• Hardware Engineers
• Data Architects	• Test Automation Engineers	• Cybersecurity Analysts	• Embedded Software Engineers
• Data Visualization Engineers	• Systems Engineers	• Information Systems Security Officers	• Digital Design Engineers
• Data Science Engineers	• Test Engineers	• Cybersecurity Architects	• Electronics Technicians
• Machine Learning Engineers	• Controls Engineers	• Network Engineers	• Signal Processing Engineers
• Business Analysts	• Project Engineers	• Solutions Architects	• Signal Integrity Engineers
	• Maintenance Technicians	• Cybersecurity Systems Engineers	• Firmware Engineers
	• Software Engineers	• Security Counselors	• Radar Systems Engineers
	• Project Managers	• Cyber Engineers	• Computer Scientists
	• Machine Operators	• Cloud Engineers	• Research Scientists
	• Software Test Engineers	• Cloud Security Engineers	
	• Software Quality Engineers	• Systems Administrators	
	• Systems Administrators	• Cloud DevOps Engineers	
	• Quality Engineers	• Cybersecurity Specialists	
	• Automation Technicians	• Cyber Systems Engineers	
	• Quality Assurance Automation Engineers	• Cloud Architects	
	• DevOps Engineers	• Cloud Software Engineers	
	• Manufacturing Technicians		
	• Controls Technicians		
	• Automation Engineers		
	• Building Automation Specialists		
	• Application Engineers		
	• Process Engineers		
	• Software Developers		
	• Building Automation Technicians		
	• Quality Assurance Engineers		

4.1. *Trends in the U.S. manufacturing jobs postings*

Fig. 3 shows the total number of unique job postings of the four technology clusters between 2017 and 2020. Results show an impressive 50 % growth in job postings from 2017–2020, which is indicative of the rising demand for workforce for Industry 4.0 transformation. In addition, the Automation cluster accounts for more than 70 % of the total number of manufacturing job postings, which mainly cover traditional

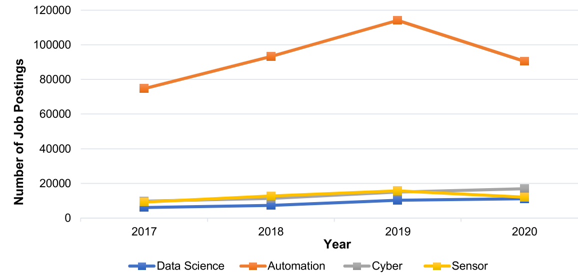

Fig. 3. Overall trends in job postings in the four technology clusters from 2017 to 2020.

jobs such as machine operators, manufacturing engineers, and automation engineers. During 2020, however, the total number of Automation job postings declined by 20 %, while the total number of jobs associated with the Data Science and Cyber clusters experienced continuous growth, 13 % and 8%, respectively. Although the total number of Data Science job postings account for only 6% of the total number of manufacturing job postings, the number of Data Science job postings has posted 82 % growth over the past four years, which is the highest among the four technology clusters.

Fig. 4 summarizes the number of job postings associated with the top seven most frequently listed job titles in the Data Science cluster. Results show a steep growth in data science-related jobs such as Data Scientist, Data Engineer, and Data Analyst. This cluster's most popular job title is Data Engineer, whose demand has increased by 300 % since 2017. It is interesting that the total number of job postings for Machine Learning Engineer has grown from 872 in 2019–2106 in 2020 (142 % increase), despite the mobility and economic challenges posed by the COVID-19 pandemic. On the contrary, the total number of Business Analyst job postings for—a non-data science-related analytics job—has declined by 16 % over the years from 2017 to 2020.

4.2. *Patterns in job requirements*

To analyze the overall pattern of the current Industry 4.0 workforce requirements, we studied the top 50 job requirements by dividing them into two categories.

1) Skills (occupation- or industry-specific hard skills, which are teachable and measurable abilities)
2) Domain knowledge (as described in Section 3.3)

Skills and domain knowledge are respectively ranked in the descending order of the percentage of job postings in each technology cluster. We also studied the distribution of educational level requirements in the job postings between 2017 and 2020.

4.2.1. *In-demand skills requirements*

Results presented in Table 4 lists the frequency of occurrence of skills in recent Industry 4.0 positions. We observe that the most common skills across technology clusters are programming languages, indicating that Industry 4.0 positions would need multiple programming languages skills. However, different technology clusters have different priorities for programming languages. For example, the demand for C++ programmers is much higher in the Cyber cluster than in the Data Science cluster. It is worth noting that Python is the top in-demand programming skill in the Automation technology cluster. In general, Python, C++, and Java are the most demanded programming skills in manufacturing job postings.

In addition to programming skills, technology clusters have their own cluster-specific skills requirements. Big data management tools are highly required skills for data science-related jobs, where Apache Hadoop, Apache Spark, Apache Kafka, and Apache Hive are among the

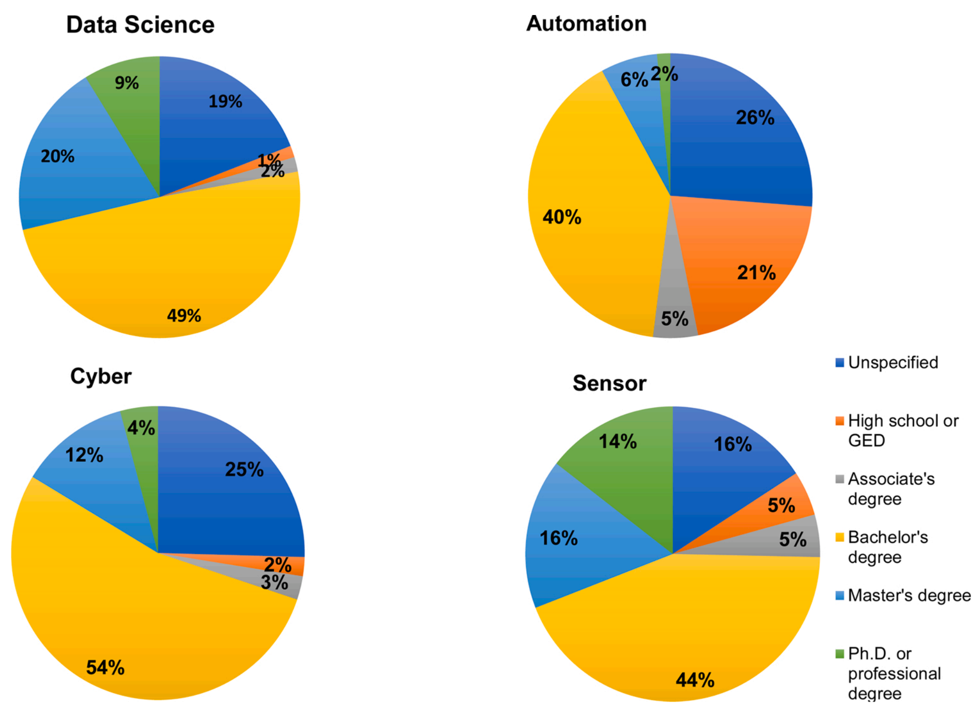

Fig. 4. (a) Data Science job postings in the U.S. manufacturing industry. (b) Required educational level distribution for job positions in four technology clusters.

Table 4
Skills requirement statistics.

Data Science		Automation		Cyber		Sensor	
Skill	Frequency	Skill	Frequency	Skill	Frequency	Skill	Frequency
SQL	41.0 %	Python	7 %	Linux	27 %	C++	35 %
Python	38.0 %	C++	7 %	C++	22 %	C	32 %
Java	20.0 %	Linux	7 %	Java	21 %	Debugging	24 %
R	19.0 %	Java	6 %	Python	19 %	Firmware	23 %
Hadoop	18.0 %	Debugging	6 %	OS	18 %	Python	21 %
Spark	17.0 %	C	5 %	C	17 %	MATLAB	17 %
AWS	14.0 %	AutoCAD	5 %	AWS	14 %	Linux	15 %
Tableau	15.0 %	SolidWorks	5 %	Debugging	13 %	Oscilloscope	15 %
Scripting	14.0 %	C#	4 %	C#	13 %	Real-TimeOS	11 %
C++	13.0 %	SQL	4 %	Scripting	12 %	OS	10 %
Scala	12.0 %	Scripting	4 %	JavaScript	12 %	Scripting	8%
NoSQL	11.0 %	OS[a]	4 %	Unix	11 %	Git	6%
Dashboard	8.0 %			SQL	11 %	Java	6%
Kafka	8.5 %			Azure	9 %		
Azure	7.5 %			Git	9 %		
PostgreSQL	9.0 %			Scrum	8 %		
Apache Hive	7.0 %			OOP[a]	8 %		
				Jenkins	8 %		
				Docker	7 %		

[a] OOP: Object-Oriented Programming. OS: Operating System.

most in-demand skills. For Automation-related jobs, traditional computer-aided design tools such as AutoCAD and SolidWorks are still widely required by most manufacturing job positions. The Cyber-related jobs place a higher emphasis on operating systems and cloud service platforms, while the Sensor-related jobs have cluster-specific requirements for MATLAB programming.

4.2.2. Trends in skill requirements

Tables 5–8 present the frequency of appearance of the top 50 skills year-wise for the four technology clusters. In general, the Automation technology cluster's hard skills requirements are relatively more stable than those for the other three clusters. The Data Science and Cyber technology clusters have witnessed many emerging skills, especially in 2020.

Table 5 shows the annual demand for Data Science skills. From 2017–2020, the changes in the skill requirements of Data Science-related jobs have been mainly focused on programming languages and big data management tools. MATLAB was eliminated from the top 50 skills in 2018, while Scala (programming language) appeared in the list since 2018, and its frequency of occurrence in job postings has increased by 243 % since then. Further, the demand for traditional database management tools such as Microsoft SQL Servers and Microsoft Access has declined, while big data management tools such as Apache Kafka and Apache Cassandra are among the highly demanded skills in 2020. As for the domain knowledge requirements, Deep Learning first appeared in the top 50 skills in 2019, and its frequency of occurrence has doubled since then. The growing demand for Deep Learning has given rise to PyTorch, which is an emerging skill in 2020.

Table 5
Annual demand for Data Science-related skills.

Skills	2017	2018	2019	2020
SQL	34 %	40 %	40 %	49 %
Python	25 %	35 %	39 %	51 %
R	19 %	23 %	22 %	16 %
Apache Hadoop	16 %	17 %	16 %	23 %
Java	14 %	16 %	17 %	30 %
Tableau	11 %	14 %	16 %	19 %
Apache Spark	11 %	14 %	16 %	24 %
Scripting	9%	12 %	10 %	24 %
Microsoft Access	8 %	7 %	6 %	–
SAS	8 %	7 %	7 %	–
Microsoft SQL Servers	8 %	7 %	–	–
Linux	7 %	8 %	6 %	–
Apache Hive	7 %	8 %	8 %	–
C++	6 %	9 %	8 %	25 %
C	6 %	8 %	6 %	–
MATLAB	6 %	–	–	–
Amazon Web Services	–	8 %	11 %	27 %
Scala	–	7 %	8 %	24 %
NoSQL	–	6 %	7 %	17 %
Power BI	–	–	7%	–
Object-Oriented Programming	–	–	–	19 %
Apache Kafka	–	–	–	16 %
Microsoft Azure	–	–	–	15 %
PostgreSQL	–	–	–	14 %
Apache Cassandra	–	–	–	14 %
PyTorch	–	–	–	10 %

Table 6
Annual demand for Automation-related skills.

Skills	2017	2018	2019	2020
C++	7 %	7 %	8 %	8 %
Linux	7 %	7%	7 %	7 %
Java	6 %	6 %	6 %	7 %
Debugging	6 %	6 %	6%	6 %
C	5 %	5 %	5%	5 %
Python	5 %	6 %	8 %	9 %
AutoCAD	5 %	5 %	5 %	5 %
SolidWorks (CAD)	5 %	5 %	5 %	4 %
Scripting	5 %	4 %	4 %	5 %
SQL	4 %	4 %	5 %	5 %
C#	4 %	4 %	5 %	5 %

According to Table 6, the demand for Data Science-related programming skills associated with the Automation cluster has gradually grown from 2017 to 2020. Specifically, the demand in 2020 for programming languages Python and SQL has increased by 80 % and 20 %, respectively. By contrast, the frequency of SolidWorks skill for computer-aided design/manufacturing/engineering has dropped by 20 % between 2017 and 2020.

The changes in skills requirements for the Cyber technology cluster primarily focus on programming automation servers and cloud platforms. As shown in Table 7, the demand for Amazon Web Services has increased by 175 % between 2017 and 2020, which has the highest growth rate among all Cyber-related skills. Moreover, the demand for Microsoft Azure has also increased by 112 % between 2017 and 2020. On the other hand, software skills such as Scrum and Git are no longer among the top 50 in-demand skillsets in 2020. Instead, new skills for programming automation servers such as Docker, Puppet, and the configuration management tool Chef have emerged on the list of tops skills in 2020.

The Sensor technology cluster has also experienced a significant change in the requirements for programming languages. As shown in Table 8, Python is the most uprising skill requirement in this technology cluster. Although MATLAB has not been a competitive programming skill for data-related jobs, the demand for MATLAB has increased in Sensor-related job postings in 2020.

Table 7
Annual demand for Cyber-related skills.

Skills	2017	2018	2019	2020
Linux	28 %	29 %	27 %	25 %
Java	23 %	23 %	21 %	20 %
C++	22 %	24 %	24 %	21 %
Operating Systems	20 %	19 %	18 %	15 %
C	18 %	19 %	18 %	15 %
Python	17 %	19 %	21 %	19 %
Debugging	14 %	13 %	13 %	11 %
Scripting	13 %	12 %	13 %	12 %
Unix	13 %	12 %	11 %	10 %
C#	13 %	13 %	14 %	12 %
JavaScript	12 %	12 %	12 %	11 %
Amazon Web Services	8 %	10 %	13 %	22 %
SQL	11 %	12 %	11 %	10 %
Firewall	11 %	11 %	10 %	–
Object-Oriented Programming (OOP)	9 %	8 %	8 %	–
Git	7 %	9 %	9 %	–
Scrum	7 %	8 %	8 %	–
Perl	7 %	–	–	–
Microsoft Azure	–	–	8 %	17 %
Jenkins	–	–	6 %	14 %
Docker	–	–	–	15 %
Puppet	–	–	–	11 %
Chef (Configuration Management Tool)	–	–	–	11 %
GitHub	–	–	–	10 %
Gitlab	–	–	–	10 %

Table 8
Annual demand for Sensor-related skills.

Skills	2017	2018	2019	2020
C	28 %	33 %	29 %	32 %
C++	26 %	32 %	32 %	35 %
Debugging	23 %	27 %	23 %	24 %
Firmware	22 %	23 %	20 %	23 %
Python	14 %	15 %	18 %	21 %
MATLAB	13 %	18 %	15 %	17 %
Linux	13 %	12 %	14 %	15 %
Oscilloscope	13 %	19 %	13 %	15 %
Operating Systems	10 %	9 %	10 %	10 %
Scripting	9 %	7 %	7 %	8 %
Java	6 %	–	7 %	6 %
Perl	5 %	–	–	–
Git	–	8 %	6 %	6 %
VHSIC Hardware Description Language	–	6 %	–	–
Assembly Language	–	6 %	–	–

4.2.3. *In-demand domain knowledge requirements*

In line with the evolution of the technical skills required by recent Industry 4.0 job postings, the demand for domain knowledge has also significantly changed over recent years. Domain knowledge refers to the knowledge of specialists or experts in a particular field. In the present work, domain knowledge refers to the knowledge of manufacturing, including methodologies, technologies, and know-how.

A summary of the four-year domain knowledge requirements in job postings is presented in Table 9. Employers expect more high-level requirements for domain knowledge than for the skills requirements. Although some domain knowledge requirements are domain-specific, many others are shared across the four technology clusters. For example, Agile methodology, a highly demanded project management process once mainly used for software development, is more and more important in manufacturing companies. Agile methodology is a modern practice that promotes continuous iteration of development and testing during the whole lifecycle of the project. The digital transformation of manufacturing provides an opportunity to rethink the production processes with the implementation of Industry 4.0 technologies. The combination of Agile methodology and data analysis will make the existing manufacturing process smart.

Table 9
In-demand domain knowledge requirement summary for years 2017 through 2020.

Data Science		Automation		Cyber		Sensor	
Domain Knowledge	Freq.	Domain Knowledge	Freq.	Domain Knowledge	Freq.	Domain Knowledge	Freq.
Machine Learning	27 %	Automation	13 %	Agile Methodology	22 %	Embedded Software	27 %
Big Data	24 %	Corrective & Preventive Action	11 %	Cyber Security	15 %	Algorithms	22 %
Statistics	17 %	Lean Manufacturing	9%	Automation	15 %	Embedded Systems	16 %
Algorithms	17 %	Machine Operation	9 %	Vulnerability	11 %	Electronics	16 %
Data Engineering	14 %	Auditing	8 %	Firewall	10 %	Simulations	13 %
Agile Methodology	14 %	Continuous Improvement Process	8 %	TS/SCI Clearance	9 %	Signal Processing	12 %
Extract Transform Load	14 %	Tooling	8 %	DevOps	9 %	Field-Programmable Gate Array	10 %
Data Modeling	14 %	Agile Methodology	7 %	Software Development Life Cycle	9 %	Hardware Architecture	10 %
Data Warehousing	12 %	Process Engineering	7%	Algorithms	9 %	Agile Methodology	10 %
Data Visualization	12 %	Programmable Logic Controllers	6 %	Scalability	8 %	Digital Signal Processing	10 %
Database Administration	12 %	Quality Management Systems	6 %	Network Engineering	8%	Test Equipment	10 %
Relational Databases	11 %	Packaging & Labeling	6 %	Cloud Computing	8%	Serial Peripheral Interface	8 %
Data Mining	10 %	Machining	6 %	Virtualization	8%	Automation	7 %
Business Intelligence	10 %	Six Sigma Methodology	6 %	Configuration Management	7%	Printed Circuit Board	6 %
Scalability	10 %	Good Manufacturing Practices	5 %	Unit Testing	7 %	Radio Frequency	6 %
Mathematical Optimization	9 %	Root Cause Analysis	5 %	.NET Framework	7 %	Machine Learning	6 %
Data Architecture	9 %	Statistical Process Control	5 %	Web Services	7 %	Logic Analyzer	6 %
Automation	9 %	Preventive Maintenance	5 %	TCP/IP	7 %	Systems Integration	6 %
Artificial Intelligence	8 %	Control Systems	5 %	Cyber Engineering	6 %	Software Design	6 %
Data Management	8 %	Test Planning	5 %	Solution Architecture	6 %	Device Drivers	6 %
Operations Research	8 %	Quality Management	5 %	Auditing	6 %	Signal Generators	6 %
Deep Learning	7%	Computer-Aided Design	5 %			Prototyping	6 %
Data Quality	7 %	HVAC	4 %			Test Planning	5 %
		Hydraulics	4 %			Secret Clearance	5 %
						Version Control	5 %

From 2017 to 2020, the domain knowledge requirements have experienced significant shifts in different technology clusters. Two important findings are as follows: 1) Machine Learning, the most in-demand domain knowledge in data-related jobs, has emerged in the top 50 skill requirements list since 2019. 2) Several cloud-based domain knowledge requirements such as Cloud Infrastructures and Cloud Computing Security have been the main requirements for Cyber-related jobs. Table 10 lists the top 5 uprising domain knowledge requirements in each technology cluster since 2017.

4.2.4. Educational level requirements

This section analyzes the patterns of educational level requirements in manufacturing job postings between 2017 and 2020. Most job postings specify the minimum educational level required. Using Emsi JPA dataset, we extracted the number of job postings associated with different educational levels for each technology cluster, as shown in Fig. 4. Among the job titles collected, nearly half of the job postings require a bachelor's degree or higher. The job postings with master's or Ph.D. degree requirements are primarily associated with the Data Science and Sensor technology clusters. The Automation technology cluster has the highest proportion of job postings with high school or General Education Diploma (GED) educational level requirements (e.g., Machine Operator). These observations indicate that the manufacturing industry needs to retrain and retool the existing skilled workforce across all education levels of the employees to address the needs of the Industry 4.0 workforce for automating and digitizing labor-intensive manufacturing processes.

4.3. Workforce profile data analysis

This section analyzes the frequency of occurrence of different skills in job applicant profiles (i.e., supply) collected from the Emsi Profile Data (PD) analytics tool. The PD data was collected using the data collection procedure described earlier for the Emsi JPA. The ten most-supplied skills and domain knowledge topics associated with the four technology clusters are presented in Tables 11 and 12, respectively.

In the Data Science technology cluster, the three most available (supply) skills are programming languages (SQL, Python, Java), which are also highly in-demand skills (demand) as discussed in Section 4.2. However, the supply of specific skills such as Microsoft SQL Servers and Microsoft Access significantly exceeds the employers' demand in the Data Science technology cluster. Further, MATLAB, SolidWorks, and AutoCAD are still the most available skills in the Automation technology cluster, while the employers' demand in that cluster is shifting towards programming languages such as Python. Moreover, in the Cybertechnology cluster, AWS and Microsoft Azure do not appear on the PD's top 10 skills, despite the rising demand by employers for those skills. These findings shed light on the gap between the demand and supply of skills

Table 10
Top 5 most increasing domain knowledge.

Rank	Data Science		Automation		Cyber		Sensor	
	Domain Knowledge	Growth Rate	Domain Knowledge	Growth Rate	Domain Knowledge	Growth Rate	Domain Knowledge	Growth Rate
1	Data Engineering	163 %	Machine Operation	43 %	DevOps	183 %	Signal Processing	50 %
2	Relational Databases	113 %	Good Manufacturing Practices	40 %	Cloud Computing	150 %	Agile Methodology	43 %
3	Data Visualization	100 %	Agile Methodology	29 %	Automation	83 %	Digital Signal Processing	43 %
4	Machine Learning	84 %	HVAC	25 %	Cyber security	81 %	Field-Programmable Gate Array	25 %
5	Statistics	77 %	Hydraulics	25 %	Agile Methodology	61 %	Algorithms	22 %

Table 11
Top 10 most supplied skills.

Rank	Data Science		Automation		Cyber		Sensor	
	Skill	Freq.	Skill	Freq.	Skill	Freq.	Skill	Freq.
1	SQL	38 %	MATLAB	10 %	Operating Systems	18 %	C	15 %
2	Python	23 %	SolidWorks (CAD)	10 %	Linux	18 %	C++	15 %
3	Java	18 %	AutoCAD	9 %	SQL	16 %	MATLAB	15 %
4	Microsoft SQL Servers	17 %	Java	9 %	TCP/IP	13 %	Debugging	10 %
5	R	16 %	C++	9 %	Firewall	13 %	Linux	9 %
6	C++	15 %	C	8 %	Unix	13 %	Java	8 %
7	C	13 %	SQL	7 %	Java	10 %	Python	8 %
8	Hyper Text Markup Language (HTML)	13 %	Linux	5 %	Microsoft Access	10 %	Firmware	6 %
9	MATLAB	12 %	Hyper Text Markup Language (HTML)	5 %	Microsoft SQL Servers	9 %	LabVIEW	5 %
10	Microsoft Access	12 %	Python	5 %	Dynamic Host Configuration Protocol (DHCP)	9 %	SQL	4 %

Table 12
Top 10 most available (supply) domain knowledge topics.

Rank	Data Science		Automation		Cyber		Sensor	
	Domain Knowledge	Freq.	Domain Knowledge	Freq.	Domain Knowledge	Freq.	Domain Knowledge	Freq.
1	Database Administration	29 %	Machine Operation	22 %	Network Engineering	31 %	Electronics	18 %
2	Machine Learning	18 %	Lean Manufacturing	11 %	Active Directory	26 %	Hardware Architecture	14 %
3	Business Intelligence	13 %	Continuous Improvement Process	10 %	Windows Servers	25 %	Embedded Software	13 %
4	Data Warehousing	12 %	Process Engineering	9%	Technical Support	21 %	Embedded Systems	13 %
5	Data Mining	10 %	Automation	8 %	Network Administration	20 %	Test Equipment	8 %
6	Statistics	9 %	Computer-Aided Design	7 %	System Administration	19 %	Simulations	8 %
7	Big Data	9 %	Six Sigma Methodology	7 %	Network Switches	19 %	Algorithms	7 %
8	Extract Transform Load (ETL)	9 %	Electronics	6 %	Network Security	16 %	Soldering	7 %
9	Algorithms	9 %	Root Cause Analysis	5 %	Local Area Networks	14 %	Printed Circuit Board	7 %
10	Data Modeling	9 %	Machining	5 %	Disaster Recovery	14 %	Automation	7 %

in the technology clusters.

The analysis of domain knowledge data related to the Cyber technology cluster reveals an absolute mismatch between the top 10 domain knowledge topics offered by job applicants and the top 10 topics wanted in the job postings. Further, there is a surplus of specific domain knowledge topics in other technology clusters. For example, Business Intelligence in the Data Science technology cluster and Process Improvement in the Automation technology cluster are among the ten most-supplied domain knowledge topics despite the low demand for them in the 2020 job postings. This significant supply-demand gap in domain knowledge topics is yet another indication of the discrepancy between current educational/workforce training programs and the manufacturing industry's actual needs.

4.4. Identified gaps in skills and domain knowledge

In this section, we identify the most critical gaps in skills and domain knowledge topics related to *Data Science* that are specifically applied in the manufacturing discipline. *Data Science* here refers to a collection of interdisciplinary such as mathematics, statistics, inference, computer science, optimization, machine learning, and data management applied to manufacturing. In this context, a *gap* is defined as the difference between the relative occurrence frequencies of a given skill or domain knowledge topics across job postings and profiles/resumes. Fig. 5 through 12 provide the details of major gaps in skills and domain knowledge topics associated with the four technology clusters. The gap analysis can inform educational and workforce training programs, government initiatives, and policy/decision-makers in the manufacturing industry.

4.4.1. Programming languages

For the Data Science technology cluster, Python is the most significant skills gap, despite it being one of the most available skills in the job-seeker profiles (see Fig. 5). Similarly, there is a significant skills gap associated with Python in the Automation technology cluster (see Fig. 7). For Sensor technology cluster, C++ and C are the most demanded hard skills (see Fig. 11). Verma et al. [33] identified Java as the most in-demand programming language for extensive data analysis in software engineering. However, this happens to be different for manufacturing sector. The demand for Java is relatively lower in the

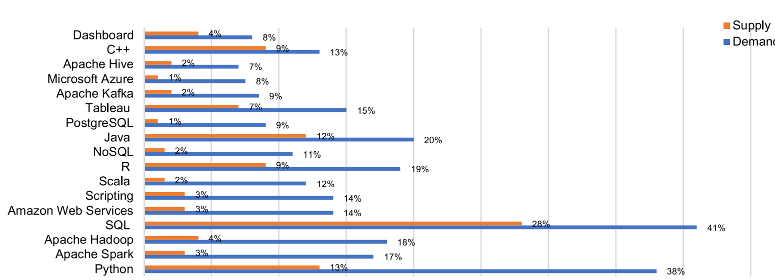

Fig. 5. Data Science-related skills demand, supply, and gaps.

manufacturing industry, and the supply of Java skill far exceeds the demand in all technology clusters except for Data Science cluster (see Figs. 5, 7, 9, and 11).

4.4.2. *Big data processing and database management tools*

Big data processing and data management tools are the two core aspects of data-related job requirements for data acquisition, processing, and storage. With the uprising demand for data engineer positions, the skills gap associated with those tools is evident. Apache Hadoop, a popular tool for big data processing, posts the second-largest skills gap in the Data Science technology cluster, followed by Apache Spark, a real-time big data processing tool (see Fig. 5). Further, NoSQL, a relatively new database management tool, is among the skills that face significant shortage (see Fig. 5).

4.4.3. *Software automation server and cloud platforms*

The primary skills gap in the Cloud technology cluster is related to software automation servers and cloud platforms. AWS, the most widely used cloud platform, accounts for the largest skills gap in the Cloud technology cluster (see Fig. 9). Similarly, there is a significant shortage of AWS skills in the Data Science cluster (see Fig. 5). The rising demand for cloud platforms has also widened the skills gap for Microsoft Azure (see Figs. 5 and 9). Further, there is an emerging demand for skills related to software automation servers and configuration management tools in the Cyber cluster. Additional examples of skills with supply shortages in 2019 and 2020 include Jerkins and Docker (see Fig. 9).

4.4.4. *Data-related domain knowledge*

The digital, connected, and complex Industry 4.0 environments call for a new set of domain knowledge topics. Current data science-related labor markets demand talents with multidisciplinary knowledge in mathematics, data science, and computer science. Major domain knowledge gaps associated with data-related jobs include big data, machine learning, statistics, algorithms, and data engineering (see Fig. 6). On the contrary, the demand for traditional data-related domain knowledge topics such as database administration and business intelligence is gradually declining despite their high frequency of appearance on job applicant profiles.

4.4.5. *Industry 4.0 domain knowledge*

The most significant domain knowledge gap in the Automation technology clusters is corrective and preventive action (CAPA), followed by automation, auditing, tooling, and preventive maintenance (see Fig. 8), indicating relative stability of demand for these relatively mature domain knowledge topics. Cybersecurity accounts for the most prominent domain knowledge gap in the Cyber technology clusters, followed by vulnerability and agile methodology (see Fig. 10), which shows the growing concerns for securing cyber-manufacturing systems against cyberattacks. In the Sensor technology cluster, algorithms, embedded software, and machine learning are the top 3 domain knowledge topics with large gaps (see Fig. 12), which points to the increasingly widespread use of sensors and sensing technologies in factories and the needs for intelligent, real-time signal processing and sensor analytics methods. In addition to cluster-specific domain knowledge gaps, several topics are shared across the clusters. Examples include algorithms, machine learning, and automation.

5. Discussion

The comprehensive manufacturing labor market analysis presented in this paper provides both the big picture and fine-grained details of the emerging and future trends in the manufacturing workforce skills and domain knowledge requirements and gaps associated with Data Science, Automation, Cyber, and Sensor technology clusters. This knowledge can be leveraged by decision-makers in educational institutions, policy-makers in government agencies, and the executives in the manufacturing industry to better prepare the workforce for current and future talent needs through training and educational pipelines.

5.1. *Current and future trends in skills and domain knowledge*

The total demand for data science-related jobs in manufacturing in the U.S., as estimated by unique job postings, has increased over the years since 2017. It is worth mentioning that the total demand for data science-related jobs has not been affected by the COVID-19 pandemic. As shown in Fig. 2, compared with the total of job postings in 2019, the total number of Data Science job postings in 2020 has increased by 20 %. In 2020, Data Engineer was the most in-demand job category compared with nine other data science-related job titles. We project that the demand for data engineers and machine learning engineers will dominate the labor market. Interestingly, we observed some opposite trends in many data science-related job positions in the manufacturing industry. For example, the total number of posts requiring business intelligence expertise has declined.

Our analysis of skills requirements reveals that SQL, Python, and Java are the top three in-demand programming skills in the Data Science technology cluster, where Python accounts for the largest supply shortage. Further, results show that the frequency of occurrence of C++ programming languages in job postings has increased by 317 % over the past four years. Compared with the programming language requirements in the other three technology clusters, the Data Science technology cluster requires workers with more diverse programming skills. Job seekers who are proficient in SQL, Python, and C++ will be favored in the labor market in the near future. Besides the requirements for programming languages, big data processing tools play a key role in Data Science technology cluster. The high demand for data engineers has in turn, driven a growing demand for Apache Hadoop and Apache Spark skills, which are the top two most in-demand tools in this

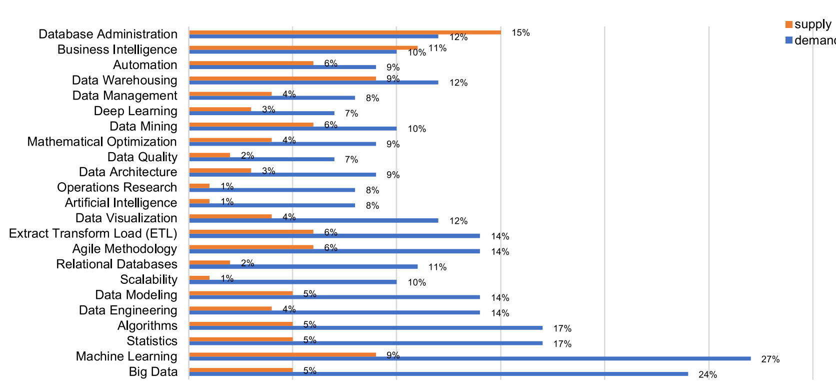

Fig. 6. Data Science-related domain knowledge demand, supply, and gaps.

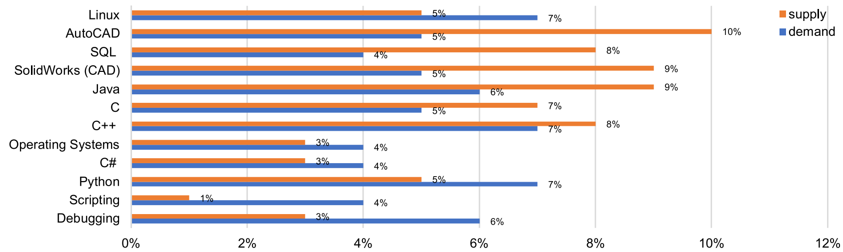

Fig. 7. Automation-related skills demand, supply, and gaps.

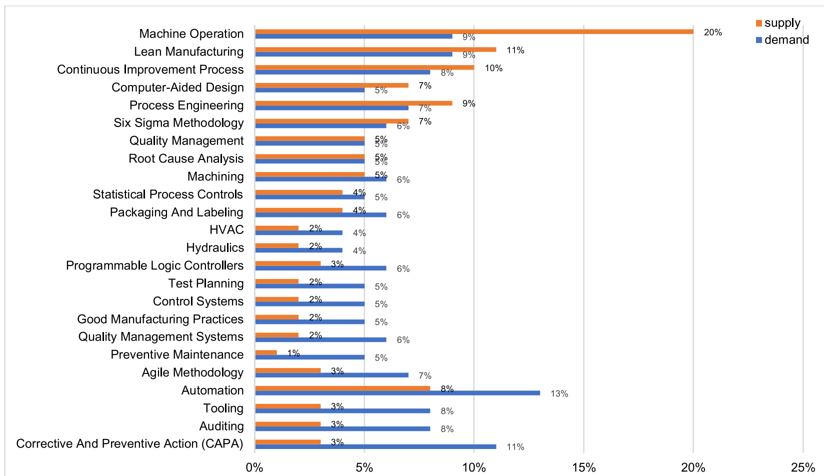

Fig. 8. Automation-related domain knowledge demand, supply, and gaps.

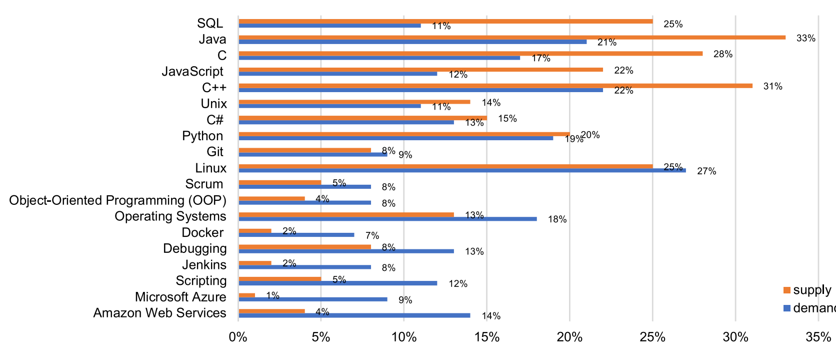

Fig. 9. Cyber-related skills demand, supply, and gaps.

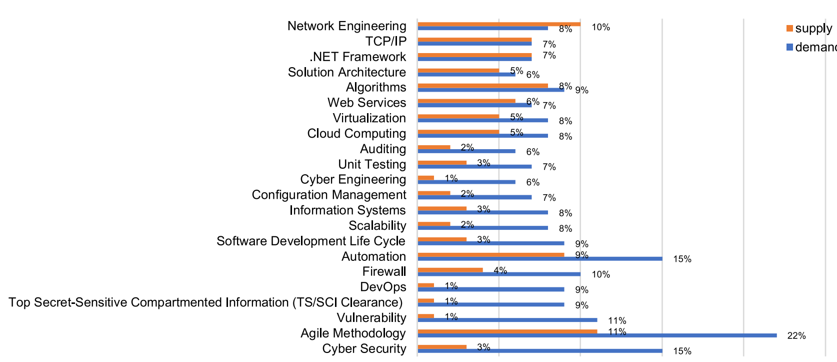

Fig. 10. Cyber related domain knowledge demand, supply, and gaps.

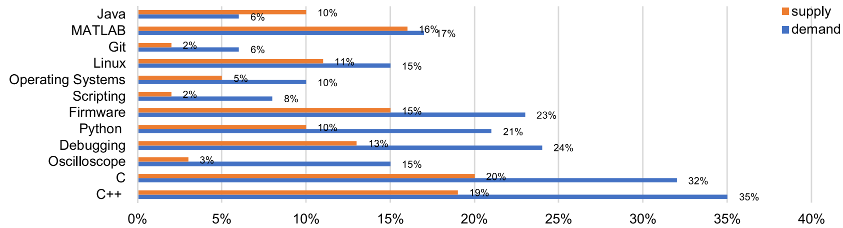

Fig. 11. Sensor-related skills demand, supply, and gaps.

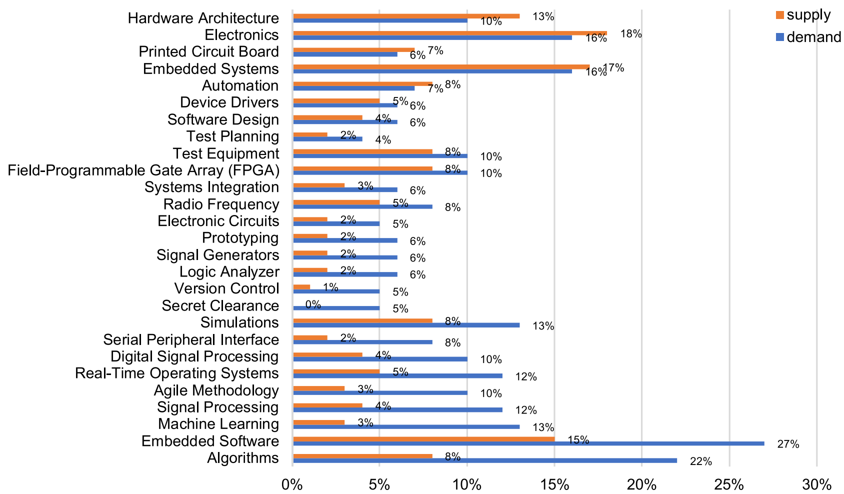

Fig. 12. Sensor-related domain knowledge demand, supply, and gaps.

technology cluster. The significant skills gap associated with Apache Hadoop and Apache Spark makes them the most disruptive software skills for those who desire to be a Data Engineer or a Big Data Engineer.

As job requirements continue to change, the manufacturing workforce needs to master new hard skills to meet the needs of future labor markets. Cloud service platforms such as AWS and Microsoft Azure are highly demanded in the Data Science technology cluster in 2020. In addition, Apache Kafka and Apache Cassandra are two emerging big data processing tools with a high frequency of occurrence in job postings in 2020. However, the occurrence frequency of these two skills is decimal in the current workforce profiles. Apache Kafka and Apache Cassandra will be the next disruptive big data processing tools for data-related job seekers in the future.

5.2. Key insights

5.2.1. Programming languages for data science

Numerous factors impact the Industry 4.0 environment, which drives the shifts in skillset needs in the manufacturing industry. Current Automation-related job requirements have shifted from the demand from computer-aided design tools to programming-oriented skills such as Python. Given that the most acquired skills in the Automation cluster are MATLAB and computer-aided design tools, the current supply of computer-aided design tools such as CAD far exceeds the demand of employers. The increasing volume of data in modern factories has driven the demand growth for programming languages such as Python and SQL—the top two programming skills in high demand and short supply in the Data Science technology cluster.

5.2.2. Cloud service platforms

Emerging cloud service platforms and programming tools do not meet employers' demands. Workers should appropriately shift their learning focus to programming languages and emerging skills such as Microsoft Azure and Dockers. The rising demand for cloud-based and security-related domain knowledge poses new challenges to traditional manufacturing practitioners. The manufacturer's need for talent with digital and computational skills will increase. Traditional manufacturing companies will benefit from adequately investing in training their workforce in advanced digitalization and computing technologies.

5.2.3. Advanced data science and analytics

Data Science is one of the most in-demand and fast-growing professional fields in the manufacturing industry. Data Science-related jobs have an extremely dynamic and competitive working environment with an ever-changing and progressing demand for advanced data science and analytics, including computation science, machine learning, algorithms, and optimization. The present gap analysis reveals the top in-demand domain knowledge topics in data science. This study demonstrates that the current workforce in the manufacturing industry requires a solid fundamental knowledge base in data science to be competitive. The skills gap reveals that Industry 4.0 environments demand interdisciplinary skillsets and knowledge spanning manufacturing science/engineering/technologies, data science, and even computer science.

5.2.4. Bridging the gap – on the talent training and recruitment

While many skills identified have been around for a while, it was only after the 2010s that data science has gained popularity in the manufacturing industry. Therefore, the upskilling of the current workforce with strong data science skills and attracting talents with strong competencies in data science, machine learning, programming languages are critical for the manufacturing sector transition to Industry 4.0. Although the data used for this gap analysis doesn't directly recommend the strategies of talent recruitment, we believe that the identified top skills in shortage provide valuable information for companies to (re-)train their employees and offer them opportunities in education and training. This would benefit the manufacturing industry in addressing the widening skills gap in the process of digital transformation. In addition, identifying these crucial gaps would also shed light on the transforming traditional educational curriculum in manufacturing to a more data and analytics centered. Manufacturing curriculum in Mechanical Engineering, Industrial Engineering, and Manufacturing Engineering programs need to be updated and renovated to include data science-related skills and domain knowledge.

The skills gap and domain knowledge deficiency indicate the necessity for a gap-driven learning approach based on multidisciplinary collaboration between academia and industry. Future data science learning is anticipated to consist of the following five competency domains: machine learning, big data, statistics, algorithm, and data engineering, where large gaps currently exist between the employers' talent demand and the talent supply in the job market. Although the priorities of domain knowledge topics differ from one position to another, employers in data science generally demand a collection of knowledge of these five domains. Industry 4.0 workforce needs to be upskilled or reskilled to cover a broader spectrum of skills and domain knowledge topics. This work presents a wide range of insights into the trends of programming language requirements in manufacturing. For example, SQL and Python are the most in-demand programming languages for database-related workers, while Scala and Java are the most demanded programming languages for data engineers. Workers should customize their learning priorities according to the importance of these programming languages in their job positions.

6. Conclusions

This paper has investigated how data-driven technologies, skills, and domain knowledge have manifested in manufacturing job positions. It identified the skills and knowledge gap in the current manufacturing workforce. Data for this study is sourced from the Emsi JPA and PD modules. This paper developed a novel analysis framework to compare the skill and knowledge requirements from job postings to those reported the workforce's self-reported profiles. Analysis of the skill requirements confirmed that skills associated with many Industry 4.0 job positions require multidisciplinary skills and domain knowledge topics. The results indicated that many manufacturing employees lag behind data science and computer science-related skills for Industry 4.0. Furthermore, our analysis also highlighted that the manufacturing workforce needs programming skills to narrow the gap between current job requirements and workforce skills.

The present study identified and ranked gaps in skills and domain knowledge topics in four major technology clusters. The implications of the research findings are two-fold. First, they can guide manufacturing graduates and employees to upgrade their skills according to the emerging and future skill requirements. Second, they provide insights that could be useful for designing data science-related curricula for undergraduate and graduate manufacturing programs.

Our investigation has potential limitations. The data in Emsi is a representative sample of the labor market, and thus the data may not be comprehensive. Moreover, in this study, we have emphasized technical jobs, with an intended consequence of underrepresenting managerial-level jobs. Therefore, our results may not inform individuals pursuing managerial and executive positions. Future research should consider analysis of different job levels. We will expand our analysis by including more emerging job titles in various Industry 4.0 fields such as digital twin and augmented reality, and compare our results with those of other countries and the global labor market in our future work.

Declaration of Competing Interest

The authors report no declarations of interest.

Acknowledgments

This material is based upon work supported by the National Science Foundation under Grant No. 1935646. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References

[1] Tao F, Qi Q, Liu A, Kusiak A. Data-driven smart manufacturing. Int J Ind Manuf Syst Eng 2018;48:157–69. https://doi.org/10.1016/j.jmsy.2018.01.006.
[2] Sahal R, Breslin JG, Ali MI. Big data and stream processing platforms for Industry 4.0 requirements mapping for a predictive maintenance use case. Int J Ind Manuf Syst Eng 2020;54:138–51. https://doi.org/10.1016/j.jmsy.2019.11.004.
[3] Gao Y, Li X, Wang XV, Wang L, Gao L. A review on recent advances in vision-based defect recognition towards industrial intelligence. Int J Ind Manuf Syst Eng 2021. https://doi.org/10.1016/j.jmsy.2021.05.008.
[4] Kergroach S. Industry 4.0: new challenges and opportunities for the labour market. Foresight STI Gov 2017;11:6–8. https://doi.org/10.17323/2500-2597.2017.4.6.8.
[5] Wang J, Ma Y, Zhang L, Gao RX, Wu D. Deep learning for smart manufacturing: methods and applications. Int J Ind Manuf Syst Eng 2018;48:144–56. https://doi.org/10.1016/j.jmsy.2018.01.003.
[6] Tuptuk N, Hailes S. Security of smart manufacturing systems. Int J Ind Manuf Syst Eng 2018;47:93–106. https://doi.org/10.1016/j.jmsy.2018.04.007.
[7] Deloitte A. A deloitte and the manufacturing institute series on the skills gap and future of work in manufacturing. 2018.
[8] Skevi A, Szigeti H, Perini S, Oliveira M, Taisch M, Kiritsis D. Current skills gap in manufacturing: towards a new skills framework for factories of the future. IFIP Adv Inf Commun Technol 2014;438:175–83. https://doi.org/10.1007/978-3-662-44739-0_22.
[9] Moghaddam M, Cadavid MN, Kenley CR, Deshmukh AV. Reference architectures for smart manufacturing: a critical review. Int J Ind Manuf Syst Eng 2018;49: 215–25. https://doi.org/10.1016/j.jmsy.2018.10.006.
[10] Dhar BV. Dhar_data_science_prediction. Commun ACM 2013;56:64–73.
[11] Hayashi C. What is data science? Fundamental concepts and a heuristic example. In: Hayashi C, Yajima K, Bock H-H, Ohsumi N, Tanaka Y, Baba Y, editors. Data sci. classif. relat. methods. Tokyo: Springer Japan; 1998. p. 40–51.
[12] Maxwell. The crunch. Police J Theory Pract Princ 1968;41:115–8. https://doi.org/10.1177/0032258x6804100305.
[13] Bortolini M, Galizia FG, Mora C. Reconfigurable manufacturing systems: literature review and research trend. Int J Ind Manuf Syst Eng 2018;49:93–106. https://doi.org/10.1016/j.jmsy.2018.09.005.
[14] EMSI database, n.d., https://api.emsidata.com/.
[15] Gap S. Current issues in HR. 2013. p. 1–6.
[16] International EMS. The skills gap: a national issue that requires a regional focus. 2014.
[17] Restuccia D, Taska B. Different Skills, Different Gaps: Measuring and Closing the Skills Gap. Dev Ski a Chang World Work 2018:207–26. https://doi.org/10.5771/9783957103154-207.
[18] Berger T, Benedikt Frey C. Digitalization, jobs, and convergence in Europe: strategies for closing the skills gap. Prepared for the European Commission DG Internal Market, Industry, [1]. In: Berger T, Benedikt Frey C, editors. Digitalization, jobs, and convergence in Europe: strategies for closi. Oxford Martin Sch; 2016.
[19] Moldovan L. State-of-the-art analysis on the knowledge and skills gaps on the topic of industry 4.0 and the requirements for work-based learning. Procedia Manuf. 2019;32:294–301. https://doi.org/10.1016/j.promfg.2019.02.217. Elsevier B.V.
[20] Chenoy D, Ghosh SM, Shukla SK. Skill development for accelerating the manufacturing sector: the role of 'new-age' skills for 'Make in India'. Int J Train Res 2019;17:112–30. https://doi.org/10.1080/14480220.2019.1639294.
[21] Vista A. Data-driven identification of skills for the future: 21st-century skills for the 21st-century workforce. SAGE Open; 2020. p. 10. https://doi.org/10.1177/2158244020915904.
[22] Addressing the Skills Gap: A Regional Analysis, n.d.
[23] McGunagle D, Zizka L. Employability skills for 21st-century STEM students: the employers' perspective. High Educ Ski Work Learn 2020;10:591–606. https://doi.org/10.1108/HESWBL-10-2019-0148.
[24] Meyer G, Brünig B, Nyhuis P. Employee competences in manufacturing companies – an expert survey. J Manag Dev 2015;34:1004–18. https://doi.org/10.1108/JMD-06-2014-0056.
[25] Internships P, Experiences C, Engineering II, Knowledge M. Filling the skills gap in U.S. Manufacturing. 2016.
[26] Xu T, Zhu H, Zhu C, Li P, Xiong H. Measuring the popularity of job skills in recruitment market: a multi-criteria approach. 32nd AAAI Conf. Artif. Intell. AAAI 2018 2018:2572–9.
[27] Weaver A, Osterman P. Skill demands and mismatch in U.S. Manufacturing. Ind Labor Relations Rev 2017;70:275–307. https://doi.org/10.1177/0019793916660067.
[28] Nogueira MC, Afonso Ó, Soukiazis E. Skill premium in Portuguese manufacturing industries. Appl Econ Lett 2018;25:1015–8. https://doi.org/10.1080/13504851.2017.1391993.
[29] Javdekar C, Watson E, Kapilow V, Bograd M, Boyer P, Zeid I, et al. Closing the advanced manufacturing talent gap. Procedia Manuf 2016;5:1197–207. https://doi.org/10.1016/j.promfg.2016.08.094.
[30] Library NEDS. Rapid #: -16717580 CROSS REF ID : LENDER : developing the next generation of engineers for intelligent and sustainable manufacturing: a case study*. 2020.

[31] Lödding H, Thoben K, Von Cieminski G. Advances in production management systems. 2017. https://doi.org/10.1007/978-3-319-66923-6.
[32] Ras E, Wild F, Stahl C, Baudet A. Bridging the skills gap of workers in industry 4.0 by human performance augmentation tools – challenges and roadmap. ACM Int. Conf. Proceeding Ser., vol. Part F1285; 2017. p. 428–32. https://doi.org/10.1145/3056540.3076192.
[33] Verma A, Yurov KM, Lane PL, Yurova YV. An investigation of skill requirements for business and data analytics positions: a content analysis of job advertisements. J Educ Bus 2019;94:243–50. https://doi.org/10.1080/08832323.2018.1520685.
[34] Mikalef P, Krogstie J. Investigating the data science skill gap: an empirical analysis. IEEE Glob Eng Educ Conf EDUCON 2019;(April-2019):1275–84. https://doi.org/10.1109/EDUCON.2019.8725066.
[35] Kimmell J, Martin S. Sorting out the skills gap. 2015.
[36] Chris K. Domain knowledge: what is it and examples. 2020.
[37] Erboz G. How to define industry 40: the main pillars of industry 4.0. Manag Trends Dev Enterp Glob Era 2017:761–7.
[38] Yang F, Gu S. Industry 4.0, a revolution that requires technology and national strategies. Complex Intell Syst 2021;7:1311–25. https://doi.org/10.1007/s40747-020-00267-9.
[39] Vaidya S, Ambad P, Bhosle S. Industry 4.0 - a glimpse. Procedia Manuf 2018;20:233–8. https://doi.org/10.1016/j.promfg.2018.02.034.
[40] Rüßmann M, et al. Future of productivity and growth in manufacturing. Bost consult. 2015. https://doi.org/10.1007/s12599-014-0334-4.
[41] Council NS of T. Strategy for american leadership in advanced manufacturing. Natl Sci Technol Counc 2018:1–40.
[42] Gazzaneo L, Padovano A, Umbrello S. Designing smart operator 4.0 for human values: a value sensitive design approach. Procedia Manuf 2020;42:219–26. https://doi.org/10.1016/j.promfg.2020.02.073.
[43] Schütze A, Helwig N, Schneider T. Sensors 4.0 - Smart sensors and measurement technology enable Industry 4.0. J Sensors Sens Syst 2018;7:359–71. https://doi.org/10.5194/jsss-7-359-2018.