Journal of Higher Education Policy and Management

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/cjhe20

What are the digital skills sought by scientific employers in potential candidates?

Guangyao Zhang, Lili Wang, Furong Shang & Xianwen Wang

To cite this article: Guangyao Zhang, Lili Wang, Furong Shang & Xianwen Wang (13 Jul 2024): What are the digital skills sought by scientific employers in potential candidates?, Journal of Higher Education Policy and Management, DOI: 10.1080/1360080X.2024.2374392

To link to this article: https://doi.org/10.1080/1360080X.2024.2374392

View supplementary material

Published online: 13 Jul 2024.

Submit your article to this journal

View related articles

View Crossmark data

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=cjhe20

JOURNAL OF HIGHER EDUCATION POLICY AND MANAGEMENT
https://doi.org/10.1080/1360080X.2024.2374392

Routledge
Taylor & Francis Group

Check for updates

What are the digital skills sought by scientific employers in potential candidates?

Guangyao Zhang[a,b], Lili Wang[b], Furong Shang[c] and Xianwen Wang[a]

[a]WISE Lab, Institute of Science of Science and S&T Management, Dalian University of Technology, Dalian, China; [b]UNU-MERIT, Maastricht University, Maastricht, The Netherlands; [c]College of geography and environment, Shandong Normal University, Jinan, China

ABSTRACT
The growth in digitalisation has led to an increasing demand for digital skills in various job sectors. In particular, employers in scientific job areas have shown interest in candidates possessing digital competencies. This study aims to analyse the digital skill requirements for candidates in scientific job opportunities. The content analysis is based on a dataset comprising 126,360 scientific job advertisements from Science Careers spanning the period from 2019 to 2023. The findings of this study reveal an upward trend in the demand for digital skills. Furthermore, jobs offering higher salaries are more likely to require digital skills compared to jobs offering lower salaries. Notably, skills such as data analysis, statistics, Python, and R demonstrate increasing demand, while other skills such as data collection show a declining trend. These results significantly contribute to our understanding of the impact of digitalisation on the scientific landscape.

ARTICLE HISTORY
Received 14 December 2023
Accepted 26 June 2024

KEYWORDS
Digitalisation; scientific job; employability; research education; higher education management

Introduction

As universities worldwide continue to produce a growing number of Ph.D. graduates, the pool of practitioners engaged in scientific fields steadily expands. Science and engineering employment in the United States, for example, has outpaced the overall workforce growth since 2020, constituting approximately 5.0% (around 7 million) of all U.S. jobs (NSB, 2020). Within this realm, scientific occupations have exhibited a pressing demand for digital skills (Autor, 2015), making the impacts of digitalisation on work a focal point of scientific interest (Ribeiro et al., 2023).

Digitalisation refers to the process of integrating digital technologies into various aspects of science, business, and industry. It involves the use of digital tools and systems to transform traditional methods, enhance efficiency, and create new opportunities for innovation (Ribeiro et al., 2023). Digitalisation has permeated the job market (Brynjolfsson et al., 2023). Specifically, its impacts on work are multifaceted: on one hand, digitalisation stimulates the need for highly skilled

CONTACT Lili Wang wang@merit.unu.edu UNU-MERIT, Maastricht University, Boschstraat 24, Maastricht 6211 AX, The Netherlands; Xianwen Wang xianwenwang@dlut.edu.cn WISE Lab, Institute of Science of Science and S&T Management, Dalian University of Technology, No.2 Linggong Road, Dalian 116024, China
Supplemental data for this article can be accessed online at https://doi.org/10.1080/1360080X.2024.2374392
© 2024 Association for Tertiary Education Management

employees, leading to the creation of new positions and improved work efficiency (Atalay et al., 2018). On the other hand, the efficiency gains from digitalisation may reduce employment and wages (Acemoglu & Restrepo, 2020) and exacerbate social tensions (Im et al., 2019). Nonetheless, the irreversible trend of digitalisation embedding current jobs and intensifying skill changes cannot be denied (Berger & Frey, 2016).

In particular, digitalisation is increasingly influencing the field of science. For example, reports from Burning Glass Technologies indicate a growing demand for data scientists (Burning Glass Technologies, 2017). Specifically, there were already 2,350,000 data science and analytics (DSA) job listings in 2015 in the United States, and by 2020 the number of DSA job listings is projected to grow by nearly 364,000 listings. In addition, according to U.S. Bureau of Labor Statistics (2023), employment of data scientists is projected to grow 35.0% from 2022 to 2032, which is much faster than the average for all occupations (3.0%). Furthermore, advancements in digitalisation in science have spurred the need for employees well-versed in digital skills (Meyer, 2019). Notably, the medical field offers ample opportunities to integrate digital skills, with numerous medical companies and schools utilising machine learning, deep learning, and artificial intelligence to accelerate drug development (Eisenstein, 2020) as well as enhance medical training and education (Pucchio et al., 2021). Machine learning is also projected to improve diagnostic accuracy and potentially replace much of the work of radiologists and anatomical pathologists (Obermeyer & Emanuel, 2016).

Many conventional experimental science disciplines, including physics, chemistry, and materials science, have undergone significant transformations due to digitalisation. Subfields such as computational physics and computational chemistry, requiring proficient programming, statistical mechanics simulations, and visualisation skills, have emerged as essential areas of research (Hanwell et al., 2012; Hirel, 2015). Moreover, the expansion of datasets has underscored the importance of skills related to handling and analysing large datasets and operating cyberinfrastructures in geographic science (Schwanghart & Scherler, 2014; Wright & Wang, 2011). Consequently, comprehending the demand for digital skills in science becomes of paramount importance.

For educational institutions and governments, aligning students' skills with employers' needs holds significance for educational efficacy and economic development (Abell & Becker, 2021). On a broader scale, scientific jobs contribute to scientific progress and societal advancements. Moreover, amidst the apparent oversupply of researchers for available positions (Bourne, 2013), there has been growing interest in aligning Ph.D. skills with scientific employers' requirements. For junior scientists, understanding demands of scientific jobs is instrumental to their scientific careers.

Analysing job skills offers a vital perspective for understanding the structure of labour demand, as the evolution of skill requirements in positions reflects the dynamics of skill-biased technological change (Deming & Kahn, 2018). Labour demand plays a crucial role in the enhancement of the educational system, as it serves to identify the authentic requirements of employers and facilitate the optimisation of alignment between job prerequisites and educational provisions. Furthermore, within the dynamic background of the digital transformation era, where labour demand undergoes continuous evolution, conducting an analysis of the requisites for digital skills within the realm of scientific occupations provides valuable insights that can serve as a reference for other sectors of

the industry. Despite the increasing importance of digital skills in scientific jobs, few studies have examined the demands for such skills in scientific positions.

In general, unveiling the demands for digital skills within the scientific environment serves as an important way to comprehend the genuine digital trends in the current science and the extent to which digitalisation is integrated into the advancement of science.

This study delves into the content of scientific job postings to elucidate trends and identify the digital skills sought after in this domain. Employing content analysis across multiple disciplines and drawing from a substantial dataset of 126,360 job advertisements, the study aims to ascertain the digital skills required by scientific employers and offer valuable insights to scientific developers and policymakers. The research questions addressed in this study are as follows:

(1) How do digital skills required for scientific jobs vary across multidisciplinary contexts, organisations, and countries?
(2) Are jobs offering higher salaries more likely to require digital skills compared to jobs offering lower salaries?
(3) What are the most critical digital skills demanded in scientific jobs?
(4) What are the emerging digital skills experiencing high demand?

Literature review

Some preliminary studies have explored the PhD training from the perspective of development of specialised industries through surveys and interviews (Adams et al., 2010; Sproles & Clemons, 2019). For example, Durette et al. (2016) conducted a national survey among 2,794 PhDs to identified PhDs' 'core' competencies. Several initiatives, such as Vitae (2023), have carried out a detailed analysis of various domains in which scientists' skills are developed. These studies, conducted via surveys and interviews, provide valuable insights for Ph.D. training. As the availability of datasets increases and methodologies become more sophisticated, the analysis of job advertisements, which centres on employers' requirements, offers an additional and targeted approach for Ph.D. training.

With the advent of the internet, big datasets of job advertisements on online platforms have emerged, providing real-time information on the state of the labour market. Skills requirements in job advertisements considered to be a good indicator to capture employers' demand (Deming & Kahn, 2018). Researchers can utilise these data to identify skill requirements, offer insights for human resource planning, and inform targeted efforts towards addressing skill needs (Atalay et al., 2018; Deming & Kahn, 2018; Gorsky & MacLeod, 2016). Content analysis is becoming a common method in understanding the qualifications and skills from positions. Pitt and Mewburn (2016) conducted text analysis by manually coding based on samples of job descriptions presented on Australian universities' websites. They found that employers need multi-talented workers.

Previous studies analysing job advertisements have often favoured manual coding over automated analysis (Harper, 2012). Manual coding offers the advantage of ensuring that words are analysed within their context as well as their frequency through human interpretation. However, the manual coding of extensive datasets is time-consuming and may result in

biased or inconsistent coding. Furthermore, studies relying on manual coding often face limitations in conducting further statistical analysis due to insufficient sample sizes. In contrast, automated analysis offers the advantages of speed and efficiency in consolidating large amounts of data compared to manual coding methods. Several studies employing natural language processing (NLP) and machine learning (ML) offer valuable insights into the analysis of job advertisement content (Boselli et al., 2017). Grüger and Schneider (2019), for instance, proposed a system for automatically identifying skills in German-language job postings. Similarly, Lukauskas et al. (2023) utilised NLP and data clustering techniques to investigate skill requirements in the Lithuanian labour market, identifying specific skill dynamics and employee profiles. In the context of Ph.D. graduate employment, Chen et al. (2020) employed a machine learning-natural language processing (ML-NLP) approach to extract skill requirements from research-intensive job advertisements.

Furthermore, as technology progresses, employers may have novel demands for candidates with digital skills. Thus, online job advertisements are rich sources for investigating technological changes and trends. Meyer (2019) analysed 198 data scientist positions from Indeed.com and identified varying demands for digital skills across different organisations. Sheriff and Sevukan (2023) applied Latent Dirichlet Allocation (LDA) to uncover trends in research data management based on the analysis of 810 job advertisements.

Despite the evident high demand for digital skills among scientific employers, few studies have explicitly focused on their needs. As Pitt and Mewburn (2016) underscored, content analysis is exceptionally labour-intensive. Moreover, studies relying on small and limited samples may not be sufficient to generalise valid and comprehensive results. Consequently, it is imperative to gain a comprehensive understanding of the digital skill demands in scientific jobs.

Methods and data

Methods

The research inquiries are addressed through an examination of the content within scientific job postings, wherein the requisite skills sought by scientific employers are identified. Following the study by Michel et al. (2011) which employed extensive data analysis utilising the Google Books, numerous subsequent studies have leveraged word-usage dynamics as a means to elucidate the evolution of societal and economic phenomena. The prevalent methodology involves the retrieval and extraction of information pertaining to word frequencies within a defined temporal framework, subsequently unveiling the dynamic patterns of word usage within textual data. In accordance with this approach, we posit that there exists a pertinent relationship between the actual scientific demand for digital skills and the prevalence of digital skills-related terminology within scientific job listings. Thus, by analysing the word-usage dynamics present in the details of scientific job postings, we can reveal the demand for digital skills within the realm of scientific environment.

In contrast to manual content analysis, which is suited for small sample size datasets (Meyer, 2019; Pitt & Mewburn, 2016), and the term frequency – inverse document frequency (TF-IDF) (and LDA) method, which may not effectively extract skill-related words (Zhanget al., 2022), the current approach stands out as the most appropriate for

our research purposes. Additionally, to facilitate the distinction of skills, we implement the inclusion of the suffix 'programming' after each computer language in our text, with the exception of Python. This modification allows for clearer identification and categorisation of specific programming skills in the dataset. In the specific process of skill identification, we created a list of 5,981 keywords categorised under 'Information Technology' sourced from Emsi (https://lightcast.io/open-skills/categories). Emsi, a widely utilised platform in skills analysis (Li et al., 2021), offers an extensive skill dictionary curated through rigorous cleaning and fact-checking procedures applied to each skill. This dictionary draws from a vast repository comprising hundreds of millions of online job postings and profiles from diverse sources. The resulting keywords list encompasses a significantly larger pool of terms compared to previous studies employing similar methodologies (Michel et al., 2011), thus ensuring a thorough and comprehensive search process. Subsequently, using this skills keywords list, we conducted a skill search for each job detail and obtained the frequency of skills ($No.\ of\ skill_{i,}\ of\ job_k$). Finally, we conducted statistical analysis, regression analysis, and visualisation techniques to address the aforementioned research questions.

Specifically, as to statistical analysis, we calculated the average word occurrences. The *Mean No. of skills*, denoted as the average occurrence frequency of digital skills in the dataset, is calculated following Zhang et al. (2022). Additionally, to account for potential repetition of skills within each job text, we calculate the *Mean No. types of skills*. Given that longer job descriptions may inherently contain more skills, to address the influence of job detail length on the number of skills, we further compute the *Mean ratio* by dividing the length of job details. The equations for these calculations are as follows:

$$Mean\ No.\ of\ skills = \frac{\sum_{k=1}^{N}\sum_{i=1}^{n} No.\ of\ skill_i\ of\ job_k}{The\ total\ number\ of\ jobs} \tag{1}$$

$$Mean\ No.\ of\ types\ of\ skills = \frac{\sum_{k=1}^{N}\sum_{i=1}^{n} No.\ of\ types\ of\ skill_i\ of\ job_k}{The\ total\ number\ of\ jobs} \tag{2}$$

$$Mean\ ratio = \frac{\sum_{k=1}^{N}\left(\sum_{i=1}^{n} No.\ of\ skill_i / Detail\ length\ of\ job_k\right)}{The\ total\ number\ of\ jobs} \tag{3}$$

Then, to show the differences in the number of digital skills, we conduct a regression analysis of digital skills with respect to multidisciplinary contexts, organisation types and country affliction in the job postings collected from *Science Careers*. The dependent variable is the number of digital skills required for each job ($No.\ of\ skills\ of\ job_k$). *Science Careers* does not show job's specific sub-disciplines. Considering the data limitation, we draw on the experience of Squazzoni et al. (2021) and Jia et al. (2023) and divided these jobs into three main disciplines, namely Life Sciences, Physical Sciences, and Health Sciences. Then, we introduce a dummy variable (*Multidiscipline*) to capture the multidisciplinary nature of the jobs. Specifically, *Multidiscipline* takes the value of 0 if a job belongs to only one discipline and 1 otherwise.

To account for the variations in organisation types, we define a dummy variable (*Organisation Type*) that includes four categories: Industry, Hospital, Academia, and Other. We further consolidate some specific categories into more general ones. For

instance, 'All Industry', 'Product Supplier', 'Biotech', 'Other Corporate', and 'Pharma' are recategorised as 'Industry', while 'Healthcare & Hospital' is recategorised as 'Hospital'. Additionally, 'Govt', 'Law Firm', 'Media/Communications', 'Non-Profit', and 'Consultant' are combined under 'Other'.

We transform location information into country information. In order to consider the location aspect, we establish a binary dummy variable (*NonUS*) to distinguish between jobs located in non-US countries (taking the value 0) and those in the US (taking the value 1). We also control for the time variable (*Year*) based on the job closing date and the length of each job description (*Length of job details*). *Salary* is a ratio variable. If a job displays a salary range, we computed the average salary within that range as the job's salary.

Table 1 provides a comprehensive overview of the variable definitions along with the corresponding descriptive statistics. Our data analysis reveals that out of the total 126,360 scientific jobs 22,663 jobs (approximately 17.9%) exhibit a multidisciplinary character. The distribution of jobs across different organisational types is as follows: academia (19,419 jobs, approximately 15.3%), industry (79,336 jobs, approximately 62.8%), hospitals (24,028 jobs, approximately 19.0%), and others (3,567 jobs, approximately 2.8%). This aligns with prior studies (Gofman & Jin, 2024; Jurowetzki et al., 2021) highlighting the significance of the industry as a pivotal source of scientific employment opportunities. Most of jobs (more than 91.3%) are situated in the United States. We also provided the list of top 11 Non-US countries ranked by number of jobs in the online supplementary material. Additionally, we have included Figure S1 in the online supplementary material to illustrate distribution of the number of skills. Upon examination, we observe that digital skills follow a power law distribution across all jobs, with 62.5% of jobs mentioning at least one digital skill. We included Table S1 in the online supplementary material to illustrate that Physical Sciences demonstrate the highest demand for digital skills, followed by Health Sciences, while Life Sciences exhibit the lowest demand.

With the relevant variables defined, we proceed to the regression model, seeking to investigate the relationship between the number of digital skills and various attributes of the scientific jobs. The formulation of the regression model is as follows:

Table 1. Descriptions of variables and descriptive statistics.

Variables	Type	Description	Obs.	Mean	SD	Min	Max
No. of skills of job$_k$	Ratio	The number of digital skills for each job.	126,360	2.34	4.05	0	85
Multidiscipline	Dummy	0 if the job belongs to only one discipline; 1 otherwise	126,360	0.18	0.38	0	1
Organisation type	Dummy	0 if academia; 1 if industry; 2 if hospital; 3 if other	126,350	2.09	0.67	1	4
NonUS	Dummy	0 if US; 1 otherwise	126,360	0.08	0.28	0	1
Length of job details	Ratio	The number of words for each job detail.	126,360	778.27	344.14	0	4459
Year	Dummy	0 if 2019; 1 if 2020; 2 if 2021; 3 if 2022; 4 if 2023	126,360	2.19	1.02	0	4
Salary	Ratio	Salary in US dollars.	3,013	85,138.61	61,636.50	10,000	1,112,222

$$No.\ of\ skills\ of\ job_k = \alpha + Multidiscipline\beta_1 + Organizationtype\beta_2 + \ln(Length\ of\ job\ details)\beta_3 + NonUS\beta_4 + Year\beta_5 + \varepsilon_i \quad (4)$$

While only a limited number of job listings specify salaries, it is essential to note that salary information plays a pivotal role as it may reflect the significance of the respective job positions. Consequently, we seek to conduct a further examination of the association between salaries and the number of digital skills required for a given job. To ensure the comparability of salary data, we have chosen to restrict our analysis to jobs within the United States, resulting in a dataset comprising 3,013 jobs. The formulation of the regression model is as follows:

$$No.\ of\ skills\ of\ job_k = \theta + \ln(Salary)\gamma_1 + Multidiscipline\gamma_2 + Organizationtype\gamma_3 + \ln(Length\ of\ job\ details)\gamma_4 + Year\gamma_5 + \varepsilon_i \quad (5)$$

ε_i is an error term.

No. of skills of job_k involves non-negative integers. We sequentially employed poisson models, negative binomial models, and zero-inflated negative binomial models. The conclusive Vuong test results ($p < 0.001$) advocate for the utilisation of zero-inflated negative binomial models. So, we adopt zero-inflated negative binomial models. We utilise the natural logarithm of the length of job details, denoted as *ln(Length of job details)*, as a means to reduce heteroscedasticity. Then, we employ the variable *ln(Length of job details)* elucidate the high incidence of zero values in the dependent variable.

Finally, we aim to provide insights into the changing landscape of digital skills in scientific jobs through visualisation. Using the VOSviewer clustering algorithm (van Eck & Waltman, 2010), we constructed a co-occurrence distribution of skills. The level of detail of the clustering of a network is contingent on the resolution parameter. As suggested by van Eck and Waltman (2017), users are encouraged to experiment with various values of the resolution parameter and select the value that yields the most pertinent outcomes tailored to their specific requirements. This approach can help uncover underlying patterns in the data.

Data

In this study, we collected data from Science Careers (https://jobs.sciencecareers.org), which is associated with *Science* - one of the prominent multidisciplinary peer-reviewed journals. Science Careers serves as a comprehensive platform offering numerous job opportunities worldwide, catering to scientists seeking to advance their careers. We collected a dataset of 105,157 job advertisements from *Science Careers* on 28 July 2022 and a supplementary set of 21,205 items on 20 June 2023. Finally, we retained 126,360 unique job advertisements for our analysis after data cleaning. As Figure 1 shows, the web page provides job title, employer, location, salary, closing date, discipline, position type and job description. The information for discipline, organisation type and location are available.

There are four main job-discipline categories, including health sciences, life sciences, physical sciences, and others. However, other job information is non-specific and incomplete. Less than 2.6% of jobs clarified salary. About 65.0% of jobs have position type information. 64.5% of jobs have job title, i.e., research scientist and director. Finally,

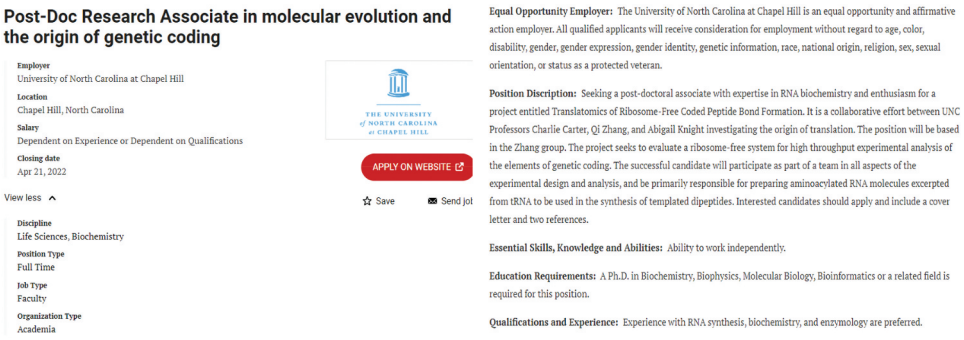

Figure 1. Screenshot of a job advertisement. (https://jobs.sciencecareers.org/job/592664)

we get structured discipline, organisation type, location information and salary from our dataset.

Results

Time trends

We aim to show the time trends in digital skills through the analysis of average word occurrences. As illustrated in Figure 2, a notable ascending trend is evident in the annual occurrence of digital skills and types of digital skills between the years 2019 and 2022, with figures ranging from 1.4 to 2.6. The diminished value in the year 2023 May be attributed to the incompleteness of our data collection during that year. Concurrently, the ratio line displays a consistent upward trajectory during the period spanning from 2019 to 2023, with values progressively increasing from 0.3% to 0.4%. From Figure 2, it is discernible that there exists a generally upward trend in the demand for digital skills in the realm of scientific job.

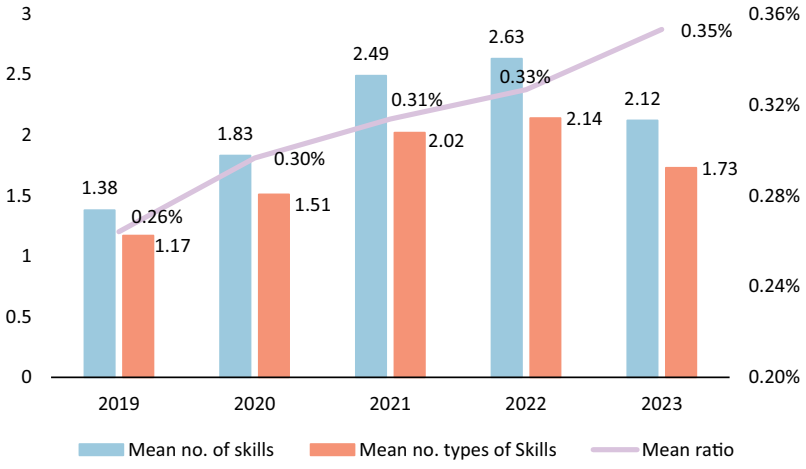

Figure 2. No. of skills and ratio of information technology (2019–2023).

Differences in the number of digital skills

Table 2 presents the results of the regression model represented by Equation (4). The coefficients of variables presented in Table 2 reveal the difference of digital skills required for scientific jobs across multidisciplinary contexts, organisations, and countries. The estimates obtained through zero-inflated negative binomial regression are expressed as incidence rate ratios (IRR), with robust standard errors indicated in parentheses. The IRR allows us to compare the incident rate between different groups. An 'IRR Greater than 1' indicates that the incident rate is higher in an exposed group compared to an unexposed group. For instance, the IRR of 1.058 for *Multidiscipline* implies that the occurrence rate of digital skills among multidisciplinary jobs is 1.058 times as high as the rate among single-discipline jobs. The model includes variables for *Multidiscipline* (m1), *Organisation Type* (m2), and *NonUS* (m3). All IRRs are found to be positive and statistically significant. Taking the model m4, which includes all the variables, as an example, the IRR for *Multidiscipline* is 1.058, implying that multidisciplinary jobs require approximately 5.8% more digital skills than single-discipline jobs.

Regarding the *Organisation Type* variable, with Academia as the baseline, the IRRs for industry, hospital, and other are 1.422, 1.017, and 1.306, respectively. These values indicate that jobs in the industry sector have the highest demand for digital skills,

Table 2. Regression results.

	m1	m2	m3	m4	m5
Multidiscipline	1.081***			1.058***	1.055***
	(0.011)			(0.011)	(0.012)
Industry		1.381***		1.422***	1.427***
		(0.020)		(0.021)	(0.023)
Hospital		0.961**		1.017	1.000
		(0.015)		(0.017)	(0.017)
Other		1.326***		1.306***	1.308***
		(0.038)		(0.037)	(0.039)
NonUS			1.223***	1.213***	1.220***
			(0.019)	(0.019)	(0.021)
2020	1.151***	1.095***	1.154***	1.097***	1.094***
	(0.023)	(0.022)	(0.023)	(0.022)	(0.022)
2021	1.298***	1.083***	1.303***	1.085***	1.073***
	(0.023)	(0.020)	(0.023)	(0.020)	(0.020)
2022	1.345***	1.148***	1.354***	1.146***	1.134***
	(0.024)	(0.021)	(0.024)	(0.021)	(0.021)
2023	1.331***	1.175***	1.287***	1.146***	
	(0.031)	(0.027)	(0.030)	(0.027)	
ln(Length of job details)	1.655***	1.498***	1.698***	1.531***	1.531***
	(0.019)	(0.017)	(0.020)	(0.018)	(0.019)
inflate					
ln(Length of job details)	0.060***	0.056***	0.062***	0.058***	0.054***
	(0.003)	(0.003)	(0.004)	(0.003)	(0.003)
lnalpha	1.466***	1.458***	1.464***	1.457***	1.452***
	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)
Observations	126357	126347	126357	126347	118190
Vuong test	0.00	0.00	0.00	0.00	0.00
Log likelihood	−247567.33	−247059.46	−247506.19	−246970.79	−231800.81
AIC	495154.66	494142.93	495032.38	493969.59	463627.62
BIC	495252.13	494259.89	495129.85	494106.04	463753.46

*$p < 0.1$, **$p < 0.05$, ***$p < 0.01$.
Exponentiated coefficients; Standard errors in parentheses; Vuong test of Zero-inflated negative binomial vs. standard negative binomial: $p = 0.00$.

with approximately 42.2% more digital skills required compared to academic jobs. Conversely, jobs in academia (and hospital) have the lowest demand for digital skills among the considered organisational types. Jobs outside the United States exhibit a demand for digital skills that is approximately 21.3% higher than their counterparts within the US.

Furthermore, the IRRs for the *Year* variable are found to be statistically significant. As the year increases between 2019 and 2022, the IRRs also demonstrate an increasing trend, indicating a growing trend in digital skill requirements for scientific jobs during this period. The data structure for the year 2023 is different from the previous years, and we also conduct a separate regression (m5) by excluding year 2023, which yields results consistent in significance and magnitude with model 4.

These results reveal that digital skill demands in scientific jobs vary based on multidisciplinary contexts, hiring organisation types and countries. Jobs involving multiple disciplines require more digital skills compared to those confined to a single discipline. Industries prioritise digital skills to support product development and enhance their competitive edge. On the other hand, jobs within health systems focus on care performance improvement, and academic positions centre on scientific publishing, leading to a relatively lesser emphasis on digital skills for these roles. Additionally, jobs located outside the United States exhibit a greater demand for digital skills compared to jobs within the United States.

Then, we examined the relationship between *Salary* and *No. of skills of* job_k, as well as utilise the natural logarithm of *Salary*, denoted as *ln(Salary)*, to reduce heteroscedasticity. In order to mitigate the influence of extreme salary values on our analysis, we identified and removed the top 2.0% of salary values as outliers (Zhang et al., 2023). Subsequently, our dataset consisted of 2,951 observations, which were employed in the regression models. Model 6 (including solely *ln(Salary)*) and model 7 (comprising all variables) were executed using negative binomial models, while model 8 (containing solely *ln(Salary)*) and model 9 (including all variables) were examined using zero-inflated negative binomial models. The results of the Vuong test support the superior fit of the zero-inflated negative binomial models. The coefficients of *ln(Salary)* can elucidate the association between salary levels and the demand for digital skills. Table 3 shows that all coefficients of *ln(Salary)* are positive and statistically significant. These findings suggest that jobs with higher salaries are more likely to require digital skills compared to those with lower salaries.

Co-occurrence analysis

In this visualisation (Figure 3), nodes are sized based on their occurrence frequencies, offering a reflection of their relative importance to some extent. The most prominent programming skills identified are R and Python. Four distinct clusters emerge in Figure 3:

Cluster 1, labelled 'Computer system', encompasses traditional digital skills, such as computer science, information technology, and software development.

Cluster 2, named 'Data analysis', includes skills related to data collection, forecasting, word processing, and calculations.

Table 3. Regression results.

	NB m6	NB m7	ZINB m8	ZINB m9
ln(Salary)	0.530***	0.272**	0.425***	0.303***
	(0.112)	(0.134)	(0.098)	(0.112)
Multidiscipline		0.267***		0.291***
		(0.095)		(0.087)
Industry		0.787***		0.734***
		(0.159)		(0.175)
Hospital		−0.206		−0.182
		(0.142)		(0.114)
Other		−0.211*		−0.197*
		(0.124)		(0.101)
ln(Length of job details)		1.067***		0.814***
		(0.063)		(0.091)
Constant	−5.692***	−9.171***	−4.181***	−7.890***
	(1.247)	(1.432)	(1.102)	(1.248)
lnalpha	1.272***	0.995***	0.755***	0.850***
	(0.046)	(0.050)	(0.076)	(0.079)
inflate				
ln(Length of job details)			−2.002***	−1.908***
			(0.229)	(0.439)
Constant			10.489***	8.950***
			(1.188)	(2.178)
Observations	2951	2951	2951	2951
Vuong test	-	-	0.00	0.02
Log likelihood	−4213.14	−4017.86	−4081.10	−4009.64
AIC	8432.29	8059.72	8172.20	8047.28
BIC	8450.26	8131.60	8202.15	8131.13
Year	Yes	Yes	Yes	Yes

*$p < 0.1$, **$p < 0.05$, ***$p < 0.01$.
Standard errors in parentheses; Vuong test of Zero-inflated negative binomial vs. standard negative binomial: $p = 0.00$. Year means year fixed effect.

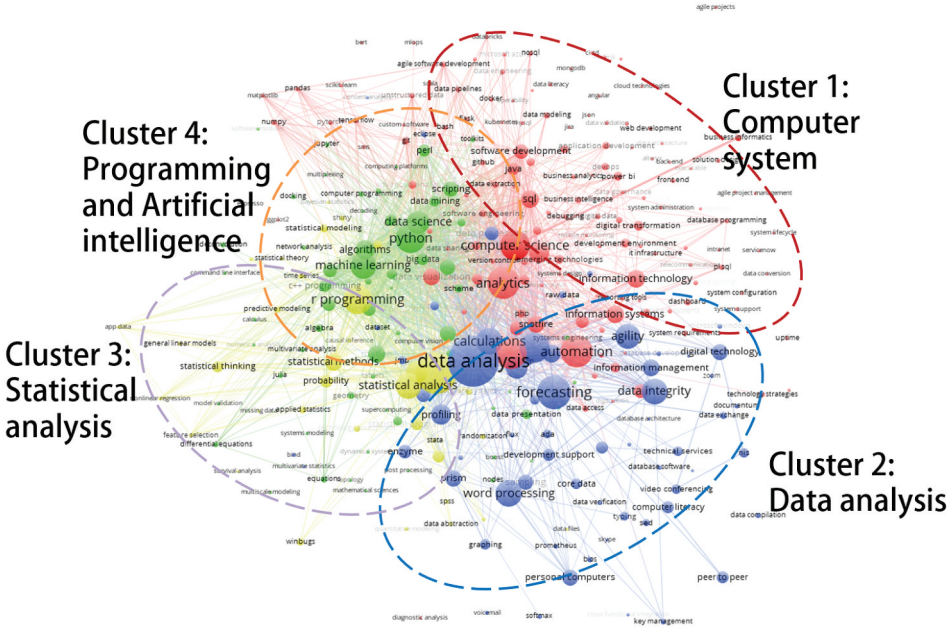

Figure 3. Clusters of digital skills in scientific job details.

Cluster 3, termed 'Statistical analysis', comprises skills like SAS, R programming, and statistical programming.

Cluster 4, designated 'Programming and artificial intelligence', encompasses skills such as machine learning, artificial intelligence, data mining, and algorithms. In comparison to 'Statistical analysis', which focuses on deriving conclusions from data analysis, 'Programming and artificial intelligence' combines statistical and programming skills and is more technically oriented, demanding higher academic qualifications.

To illustrate the evolution of these clusters over time and the emergence of new skills, Figure 4 presents a timeline showcasing the progression of skills in each cluster. Skills are listed in descending order of their average occurrence over time. Older skills are depicted in colder colours, while newer skills are displayed in warmer colours.

In the 'Computer system' cluster, skills have evolved from data conversion and computer systems to data pipelines and agile project management.

Within the 'Data analysis' cluster, skills have advanced from Skype and technical assistance software to database management and core data.

The 'Statistical analysis' cluster demonstrates the development of skills from data abstraction and statistical reporting to computing, nonlinear regression, and future selection.

The 'Programming and artificial intelligence' cluster exhibits the evolution of skills from software modules and data retrieval to machine learning, artificial intelligence, and deep learning.

Digital skills within scientific jobs are evolving as scientific employers seek to enhance the depth of their analytics work. These results indicate that potential employees should possess at least general statistical and data analysis skills. Furthermore, the advancement of technology imposes higher requirements on current employees. They need to be proficient in machine learning, knowledge management, and artificial intelligence. A 'data scientist' who demonstrates adeptness in programming languages and possesses expert-level data analysis skills may hold a distinct advantage in the job market compared to a conventional graduate equipped with only rudimentary data analysis skills.

Figure 5 presents the top 30 words based on skill intensity (Eq. 6). For instance, 'data analysis' appears 12.96 times in every 100 job details. Additionally, we calculated the intensity of skills per year and identified trends and slopes of intensity. Most of the top skills exhibit upward trends, with the exception of data collection.

$$Intensity\ of\ Skill_i = 100 \cdot \frac{\sum_{k=1}^{N} No.\ of\ skill_i\ of\ job_k}{The\ total\ number\ of\ jobs} \tag{6}$$

Discussion

This study contributes valuable insights into the demand for digital skills and the specific digital skills sought by scientific employers in potential candidates. Our methodology and results are of particular interest to researchers investigating the impact of digitalisation in

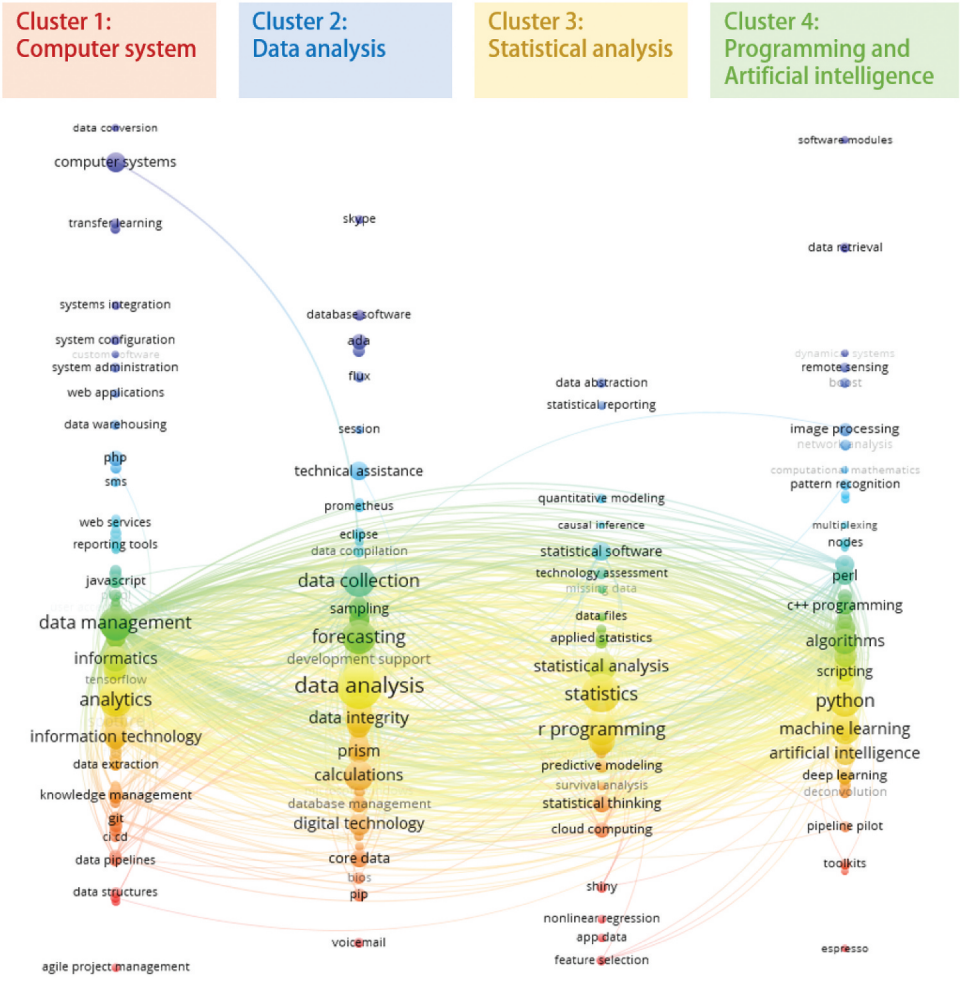

Figure 4. Timeline of digital skills in scientific job details.

the scientific domain. We conducted a comprehensive analysis using a substantial dataset of scientific advertisements, surpassing the scope of previous studies (Meyer, 2019; Pitt & Mewburn, 2016; Shahbazi & Hedayati, 2016), and consequently, we unveil the upward trends and distribution patterns of digital skills within scientific jobs.

The findings revealed a consistent upward trend in the annual demand for digital skills. The comparative analysis, based on regression models, demonstrated that multidisciplinary scientific jobs displayed a greater need for digital skills. Furthermore, when considering organisation types, jobs in the industry sector exhibited the highest demand for digital skills, whereas jobs in academia displayed the lowest demand. Additionally, employers outside the United States exhibited a greater demand for digital skills compared to their counterparts within the U. S. This indicates that nations across the globe accord paramount importance to digital skills and actively pursue candidates who possess advanced digital competencies.

Order	Skill	Intensity	Trends	Slope	Order	Skill	Intensity	Trends	Slope
1	data analysis	12.96		1.543	16	machine learning	3.22		0.380
2	statistics	7.90		1.164	17	technical support	2.51		0.036
3	forecasting	5.94		0.250	18	data science	2.40		0.503
4	automation	5.33		0.639	19	informatics	2.07		0.100
5	analytics	5.07		0.926	20	profiling	2.07		0.379
6	data management	5.03		0.247	21	algorithms	1.87		0.058
7	data collection	4.66		-0.169	22	sas	1.76		0.269
8	python	4.62		0.663	23	data quality	1.71		0.283
9	r programming	4.23		0.620	24	information systems	1.68		0.114
10	computer science	4.08		0.178	25	statistical methods	1.59		0.075
11	word processing	3.69		0.228	26	information technology	1.58		0.270
12	calculations	3.61		0.667	27	information management	1.49		0.265
13	statistical analysis	3.52		0.289	28	data visualization	1.44		0.332
14	data integrity	3.44		0.406	29	prism	1.36		0.246
15	agility	3.24		0.333	30	personal computers	1.35		0.224

Upward
Downward

Figure 5. List of top 30 skills.

In addition, these results show that jobs offering higher salaries are more likely to require digital skills compared to jobs offering lower salaries. Against the backdrop of a burgeoning number of Ph.D. graduates surpassing the availability of academic positions, careers in industry have emerged as a significant alternative for Ph.D. holders. Notably, individuals with Ph.D. qualifications employed in industry settings may command higher salaries compared to their counterparts in academia (Muurlink et al., 2023). This phenomenon is attributed partly to the growing demand for digital skills in the labour market, leading to challenges for employers in filling positions with requisite digital expertise (Burning Glass Technologies, 2017). Salaries in roles requiring digital competencies experience upward pressure, rendering them relatively lucrative compared to other occupations. Hence, it is imperative to comprehend the actual needs of employers and cultivate digital skills among Ph.D. candidates.

Furthermore, the visualisation cluster map, generated using VOSviewer software, identified subject-specific themes, namely 'Computer system', 'Data analysis', 'Statistical analysis', and 'Programming and artificial intelligence'. Current job demands notably emphasise data analysis, with Python and R emerging as the most vital skills. Concurrently, certain emerging skills such as machine learning, artificial intelligence, and deep learning are witnessing increasing demand from employers, whereas some traditional skills like computer systems and software modules are experiencing a decline in recent years.

Conclusion

This study investigates the development and trends of digital skills within scientific job advertisements spanning the period from 2019 to 2023, utilising word-

usage analysis. A comprehensive analysis of 126,360 scientific job advertisements sourced from Science Careers served to illuminate the progression of digitalisation in the scientific domain. The presence and impact of digital skills in scientific positions were discerned through the frequency of these skills within the advertisement texts.

The implications of our findings extend to a diverse range of stakeholders. Potential candidates, policymakers, and educational institutions can all derive benefits from the insights provided by this study, which shed light on the critical role of digitalisation in the scientific job market. For education institutions, the escalating demand for digital skills necessitates enhanced professional training and education. It is essential that universities and scientific institutes take proactive measures and formulate strategies to recruit employees equipped with digital skills or to provide training for existing staff to enhance their competitiveness. Notably, the growing significance of digital skills aligns with the increasing emphasis on interdisciplinary approaches in the scientific arena. For policy makers and educational systems, to meet market demand effectively, the cultivation of data literacy should be integrated into the formal education system. Educational system ought to expand learning and training avenues for students, with a particular focus on bolstering their frontier digital skills (e.g., machine learning), particularly for those aspiring to enter multidisciplinary and industrial occupations.

Limitation

We acknowledge that our research is subject to several limitations that warrant consideration. Firstly, the scope of this study is confined to analysing advertisements exclusively posted on Science Careers, predominantly situated in the United States, thereby potentially excluding positions from other sources. Consequently, the skill requirements in the broader scientific job market may not be fully represented in our analysis, leading to a bias concerning the skill subjects. Secondly, our investigation of digital skills primarily focuses on the common skill requirements specified in the analysed job postings. However, it may not encompass other specialised scientific skills crucial for conducting research comprehensively across various disciplines. Lastly, it is important to recognise that the essential digital skills sought after in scientific jobs are likely to evolve over time, driven by the changing needs of the job market. Therefore, conducting further longitudinal studies is essential to keep pace with these evolving trends and ensure accurate and up-to-date insights.

Acknowledgements

This work was supported by the grant from the National Natural Science Foundation of China [grant numbers 71974029] and the China Scholarship Council [grant numbers 202106060129]. The authors thank two anonymous reviewers and Bart Verspagen for their insightful comments and suggestions.

An earlier version of this study was posted at the Proceedings of the 26th International Conference on Science and Technology Indicators held at Granada, Spain.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

The work was supported by the China Scholarship Council [202106060129]; National Natural Science Foundation of China [71974029].

References

Abell, D., & Becker, K. (2021). Enhancing university employer attractiveness for the next generation of academics. *Journal of Higher Education Policy & Management*, *43*(5), 457–473. https://doi.org/10.1080/1360080X.2020.1847241

Acemoglu, D., & Restrepo, P. (2020). Robots and jobs: Evidence from US labor markets. *Journal of Political Economy*, *128*(6), 2188–2244. https://doi.org/10.1086/705716

Adams, K., Ripper, M., Zander, A., & Mullins, G. (2010). *The valuing of creativity in the workplace roles of engineering research graduates*. 3rd International Symposium for Engineering Education, University College Cork, Ireland.

Atalay, E., Phongthiengtham, P., Sotelo, S., & Tannenbaum, D. (2018). New technologies and the labor market. *Journal of Monetary Economics*, *97*, 48–67. https://doi.org/10.1016/j.jmoneco.2018.05.008

Autor, D.H. (2015). Why are there still so many jobs? The history and future of workplace automation. *Journal of Economic Perspectives*, *29*(3), 3–30. https://doi.org/10.1257/jep.29.3.3

Berger, T., & Frey, C. B. (2016). Did the computer revolution shift the fortunes of U.S. cities? Technology shocks and the geography of new jobs. *Regional Science and Urban Economics*, *57*, 38–45. https://doi.org/10.1016/j.regsciurbeco.2015.11.003

Boselli, R., Cesarini, M., Mercorio, F., & Mezzanzanica, M. (2017). Using machine learning for labour market intelligence. In Y. Altun, K. Das, T. Mielikäinen, D. Malerba, J. Stefanowski, J. Read, M. Žitnik, M. Ceci, & S. Džeroski (Eds.), *Machine learning and knowledge discovery in databases* (pp. 330–342). Springer International Publishing. https://doi.org/10.1007/978-3-319-71273-4_27

Bourne, H.R. (2013). A fair deal for PhD students and postdocs. *eLife*, *2*, e01139. https://doi.org/10.7554/eLife.01139

Brynjolfsson, E., Li, D., & Raymond, L. R. (2023). *Generative AI at Work* (*WorkinG PAPer 31161*). National Bureau of Economic Research. https://doi.org/10.3386/w31161

Burning Glass Technologies. (2017). *The quant crunch: How the demand for data science skills is disrupting the job market*. https://www.burning-glass.com/wp-content/uploads/The_Quant_Crunch.pdf

Chen, L., Mewburn, I., & Suonimen, H. (2020). A machine-learning based model to identify PhD-level skills in job ads. In M. Kim, D. Beck, & M. Mistica (Eds.), *Proceedings of the 18th annual workshop of the Australasian language technology association* (pp. 72–80). Australasian Language Technology Association. https://aclanthology.org/2020.alta-1.8

Deming, D., & Kahn, L. B. (2018). Skill requirements across firms and labor markets: Evidence from job postings for professionals. *Journal of Labor Economics*, *36*(S1), S337–S369. https://doi.org/10.1086/694106

Durette, B., Fournier, M., & Lafon, M. (2016). The core competencies of PhDs. *Studies in Higher Education*, *41*(8), 1355–1370. https://doi.org/10.1080/03075079.2014.968540

Eisenstein, M. (2020). Active machine learning helps drug hunters tackle biology. *Nature Biotechnology*, *38*(5), 512–514. Article 5. https://doi.org/10.1038/s41587-020-0521-4

Gofman, M., & Jin, Z. (2024). Artificial intelligence, education, and entrepreneurship. *The Journal of Finance*, *79*(1), 631–667. https://doi.org/10.1111/jofi.13302

Gorsky, D., & MacLeod, A. (2016). Shifting norms and expectations for medical school leaders: A textual analysis of career advertisements 2000–2004 cf. 2010–2014. *Journal of Higher Education Policy & Management*, *38*(1), 5–18. https://doi.org/10.1080/1360080X.2015.1126893

Grüger, J., & Schneider, G. (2019). Automated analysis of job requirements for computer scientists in online job advertisements. *Proceedings of the 15th International Conference on Web Information Systems and Technologies*, 226–233. https://doi.org/10.5220/0008068202260233

Hanwell, M.D., Curtis, D. E., Lonie, D. C., Vandermeersch, T., Zurek, E., & Hutchison, G. R. (2012). Avogadro: An advanced semantic chemical editor, visualization, and analysis platform. *Journal of Cheminformatics*, *4*(1), 17. https://doi.org/10.1186/1758-2946-4-17

Harper, R. (2012). The collection and analysis of job advertisements: A review of research methodology. *Library and Information Research*, *36*(112), 29–54. https://doi.org/10.29173/lirg499

Hirel, P. (2015). Atomsk: A tool for manipulating and converting atomic data files. *Computer Physics Communications*, *197*, 212–219. https://doi.org/10.1016/j.cpc.2015.07.012

Im, Z.J., Mayer, N., Palier, B., & Rovny, J. (2019). The "losers of automation": A reservoir of votes for the radical right? *Research & Politics*, *6*(1), 2053168018822395. https://doi.org/10.1177/2053168018822395

Jia, P., Xie, W., Zhang, G., & Wang, X. (2023). Do reviewers get their deserved acknowledgments from the authors of manuscripts? *Scientometrics*, *128*(10), 5687–5703. https://doi.org/10.1007/s11192-023-04790-7

Jurowetzki, R., Hain, D., Mateos-Garcia, J., & Stathoulopoulos, K. (2021). The privatization of AI research(-ers): Causes and potential consequences – from university-industry interaction to public research brain-drain? (arXiv: 2102.01648). *arXiv*. https://doi.org/10.48550/arXiv.2102.01648

Li, G., Yuan, C., Kamarthi, S., Moghaddam, M., & Jin, X. (2021). Data science skills and domain knowledge requirements in the manufacturing industry: A gap analysis. *Journal of Manufacturing Systems*, *60*, 692–706. https://doi.org/10.1016/j.jmsy.2021.07.007

Lukauskas, M., Šarkauskaitė, V., Pilinkienė, V., Stundžienė, A., Grybauskas, A., & Bruneckienė, J. (2023). Enhancing skills demand understanding through job ad segmentation using NLP and clustering techniques. *Applied Sciences*, *13*(10), 6119. Article 10. https://doi.org/10.3390/app13106119

Meyer, M.A. (2019). Healthcare data scientist qualifications, skills, and job focus: A content analysis of job postings. *Journal of the American Medical Informatics Association*, *26*(5), 383–391. https://doi.org/10.1093/jamia/ocy181

Michel, J.-B., Shen, Y. K., Aiden, A. P., Veres, A., Gray, M. K., THE GOOGLE BOOKS TEAM, Pickett, J. P., Hoiberg, D., Clancy, D., Norvig, P., Orwant, J., Pinker, S., Nowak, M. A., & Aiden, E. L. (2011). Quantitative analysis of culture using millions of digitized books. *Science*, *331*(6014), 176–182. https://doi.org/10.1126/science.1199644

Muurlink, O., Chen, L., Boorman, R., Pearson, D., & Cohen, G. (2023). Stakeholder perceptions of what industry wants from doctoral students: A systematic literature review. *Higher Education Research & Development*, *43*(4), 1–14. https://doi.org/10.1080/07294360.2023.2269871

NSB. (2020). *Science and engineering indicators: The state of US science and engineering*. https://ncses.nsf.gov/pubs/nsb20201

Obermeyer, Z., & Emanuel, E. J. (2016). Predicting the future—big data, machine learning, and clinical medicine. *New England Journal of Medicine*, *375*(13), 1216–1219. https://doi.org/10.1056/NEJMp1606181

Pitt, R., & Mewburn, I. (2016). Academic superheroes? A critical analysis of academic job descriptions. *Journal of Higher Education Policy & Management*, *38*(1), 88–101. https://doi.org/10.1080/1360080X.2015.1126896

Pucchio, A., Eisenhauer, E. A., & Moraes, F. Y. (2021). Medical students need artificial intelligence and machine learning training. *Nature Biotechnology*, *39*(3), 388–389. Article 3. https://doi.org/10.1038/s41587-021-00846-2

Ribeiro, B., Meckin, R., Balmer, A., & Shapira, P. (2023). The digitalisation paradox of everyday scientific labour: How mundane knowledge work is amplified and diversified in the biosciences. *Research Policy*, *52*(1), 104607. https://doi.org/10.1016/j.respol.2022.104607

Schwanghart, W., & Scherler, D. (2014). Short communication: TopoToolbox 2 – MATLAB-based software for topographic analysis and modeling in Earth surface sciences. *Earth Surface Dynamics*, *2*(1), 1–7. https://doi.org/10.5194/esurf-2-1-2014

Shahbazi, R., & Hedayati, A. (2016). Identifying digital librarian competencies according to the analysis of newly emerging IT-based LIS Jobs in 2013. *The Journal of Academic Librarianship*, *42*(5), 542–550. https://doi.org/10.1016/j.acalib.2016.06.014

Sheriff, N., & Sevukan, R. (2023). Discovering research data management trends from job advertisements using a text-mining approach. *Journal of Information Science*, 01655515231193845. https://doi.org/10.1177/01655515231193845

Sproles, C., & Clemons, A. (2019). The migration of government documents duties to public services: An analysis of recent trends in government documents librarian job advertisements. *The Reference Librarian*, *60*(2), 83–92. https://doi.org/10.1080/02763877.2019.1570419

Squazzoni, F., Bravo, G., Farjam, M., Marusic, A., Mehmani, B., Willis, M., Birukou, A., Dondio, P., & Grimaldo, F. (2021). Peer review and gender bias: A study on 145 scholarly journals. *Science Advances*, *7*(2), eabd0299. https://doi.org/10.1126/sciadv.abd0299

U.S. Bureau of Labor Statistics. (2023, September 6). *Data scientists: Occupational outlook handbook*. https://www.bls.gov/ooh/math/data-scientists.htm#tab-6

van Eck, N.J., & Waltman, L. (2010). Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics*, *84*(2), 523–538. https://doi.org/10.1007/s11192-009-0146-3

van Eck, N.J., & Waltman, L. (2017). Citation-based clustering of publications using CitNetExplorer and VOSviewer. *Scientometrics*, *111*(2), 1053–1070. https://doi.org/10.1007/s11192-017-2300-7

Vitae. (2023, June 14). *Researcher professional development [Homepage]*. Vitae. http://www.vitae.ac.uk

Wright, D.J., & Wang, S. (2011). The emergence of spatial cyberinfrastructure. *Proceedings of the National Academy of Sciences*, *108*(14), 5488–5491. https://doi.org/10.1073/pnas.1103051108

Zhang, G., Geng, Y., Wang, L., & Wang, X. (2022). What do academic employers want from candidates? A multi-disciplinary research. *From Global Indicators to Local Applications*. 26th International Conference on Science, Technology and Innovation Indicators (STI 2022), Granada, Spain. https://doi.org/10.5281/zenodo.6966543

Zhang, G., Shang, F., Wang, L., Xie, W., Jia, P., Jiang, C., & Wang, X. (2023). Is peer review duration shorter for attractive manuscripts? *Journal of Information Science*, 016555152311743. https://doi.org/10.1177/01655515231174382

Zhang, G., Wang, L., Xie, W., Shang, F., Xia, X., Jiang, C., & Wang, X. (2022). "This article is interesting, however": Exploring the language use in the peer review comment of articles published in the BMJ. *Aslib Journal of Information Management*, *74*(3), 399–416. https://doi.org/10.1108/AJIM-06-2021-0172