Received: 1 December 2022 | Accepted: 3 August 2023

DOI: 10.1111/ntwe.12278

RESEARCH ARTICLE

Mind the gender gap: Inequalities in the emergent professions of artificial intelligence (AI) and data science

Erin Young | **Judy Wajcman** | **Laila Sprejer**

Public Policy Programme, The Alan Turing Institute, London, UK

Correspondence
Judy Wajcman
Email: j.wajcman@lse.ac.uk

Abstract

The emergence of new prestigious professions in data science and artificial intelligence (AI) provide a rare opportunity to explore the gendered dynamics of technical careers as they are being formed. In this paper, we contribute to the literature on gender inequality in digital work by curating and analysing a unique cross-country data set. We use innovative data science methodology to investigate the nature of work and skills in these under-researched fields. Our research finds persistent disparities in jobs, qualifications, seniority, industry, attrition and even self-confidence in these fields. We identify structural inequality in data and AI, with career trajectories of professionals differentiated by gender, reflecting the broader history of computing. Our work is original in illuminating gendering processes within elite high-tech jobs as they are being configured. Paying attention to these nascent fields is crucial if we are to ensure that women take their rightful place at forefront of technological innovation.

KEYWORDS

artificial intelligence, careers, data science, gender, inequalities, professionalisation

INTRODUCTION

The question of how automation will affect the nature of jobs, skills, occupations, professions and labour markets is once again in the spotlight. The rise of generative artificial intelligence (AI)—widely associated with ChatGPT—is the new focal point of excitement and anxiety.[1] The novelty follows a familiar pattern: we are at an inflection point in relation to both job loss and job creation, as we head towards an unrecognisable world of work. While the extent of the looming disruption remains difficult to predict and will unfold unevenly, the burgeoning fields of data science and AI attest to the fact that major changes have already taken place. In this context, it is particularly important to investigate which groups are accessing these newly created jobs that are set to be the well-paid, prestigious and intellectually stimulating jobs of the future.

According to Berman and Bourne (2015), the new data science and AI professions potentially offer a rare opportunity to disrupt the traditionally male-dominated fields of computing and engineering, narrow the gender gap in science, technology, engineering and mathematics (STEM) and make diversity a priority early on. Yet the absence of women in the fast-growing AI and data science fields is already evident. Women make up 32% of workers in AI and data roles worldwide (World Economic Forum, 2021), and only 18% of users across the largest online global data science platforms (Young et al., 2021). Loukides' (2021) data/AI salary survey found that women's data science salaries are already significantly lower than men's, equating to 84% of the average salary for men regardless of education or job title. Despite these top-level estimations, however, there is a distinct scarcity of quality, disaggregated data on women in data and AI. This is essential to interrogate and tackle inequities in the AI and data science workforce (Zhang et al., 2021), and it is particularly difficult to make gender gaps in industry visible without granular data. Such gender data gaps are not only the result of this data not being made public, or even collected in the first place, but also due to the nascent—and crucially fast-moving and rapidly growing—nature of these fields and professions (Dorschel, 2021). In this paper, we analyse a new cross-country data set using innovative data science methodology, including creative data curation and novel classification methods, to explore the gendered dynamics of data and AI careers. A key contribution of this paper is to show the depth of exploration into gendered careers in AI which is possible when detailed data is made available.

Feminist scholars have long pointed out the ways in which labour markets are shaped by gender relations and the key role that the historical association of technical skills with masculinity has played. The under-representation of women in STEM education and professions is well documented (Blackburn, 2017). To date, however, research on the gender segregation of technical occupations typically examines the broad information and communication technology (ICT) sector (e.g., Segovia-Pérez et al., 2019). Our work goes beyond such studies, specifically honing in on the newly emerging professions of data science and AI, about which little is known (Dorschel, 2021). These elite occupations not only confer social and economic capital, but are also in the vanguard of determining how the latest technologies will be designed. As such it is especially apposite to examine the extent to which the link between gender and technical expertise is being disrupted or reproduced in these nascent fields. As Howcroft and Rubery (2019) note, if the women who do succeed in entering tech are stratified into 'less prestigious' subfields and specialties, rather than obtaining those jobs at the forefront of technical innovation, gender gaps in the future world of work, including the gender pay gap, will be widened.

This article presents a unique data science and AI career data set. Data curation and analysis are novel in several ways. First, we negotiated access to seed data collected by a recruitment firm specialising in the data, AI and advanced analytics fields to ensure focus on those specifically and actively working in such areas. This is particularly important considering their 'hard to define' nature as the fields configure, and thus using a seed list of individuals who 'self-define' as working in the area is one way to approach this. Second, our use of data from LinkedIn profiles connected to this database is another way to access the 'self-declared' job and education history of data science and AI professionals. LinkedIn is the world's largest online professional network, and a key online platform for job search, so to the best of our knowledge this cannot be found in such detail elsewhere. LinkedIn data is also uniquely rich since it enables us to analyse career trajectories in a longitudinal way. Relatedly, since the site is kept up-to-date by users, this is key for tracking the emerging professions in AI and data science which are in formation and as such constantly evolving.

We make a two-fold contribution to debates on gender inequality and the future of work. While much of the literature on gender and technology concentrates on the systematic exclusion of women from science, technology and engineering occupations (Howcroft & Taylor, 2023, p. 361), we focus on the terms of women's inclusion, specifically in the under-researched data and AI professions. Our overall aim is to explore the gendered dynamics of careers in these fields. In addition, we endeavour to illuminate how curating and analysing detailed and granular data through novel approaches can facilitate this. Our work is original in examining these new high-tech, prestigious professions as they are being formed, and as the gendering processes within them are being configured. The research reveals extensive gendered disparities in jobs, qualifications, seniority, industry, attrition, skills and even self-confidence. We begin by discussing the emergence of data science and AI as professional fields, and then review existing statistics and datasets on diversity in AI as a baseline. Sections 3 and 4 lay out our methodology and present and discuss our findings. We conclude by outlining the study's limitations and implications.

BACKGROUND

Framing emerging AI and data science professions

Over a decade ago, Harvard Business Review named data scientist as 'the sexiest job of the 21st century' (Davenport & Patil, 2012). Yet data science was only formally established as an occupational title in 2008, after years of academic debate between statisticians and computer scientists regarding the impact of computerisation on the types and methods of data analysis. In actuality, data science is still in its formative period of professionalisation (Abbott, 1988; Brandt, 2016; Dorschel, 2021), and, as Roca (2019, p. 3) points out, 'Artificial Intelligence is not a job title'. Indeed, rather than following the traditional professional-expert path, members of nascent technical occupations seem to be constructing their professional identity through an omnivorous approach to skills acquisition (Avnoon, 2021). As several authors have observed, data scientists are being constructed as hybrids, combining generalist and specialist roles as 'technicians and communicators; data exploiters and data ethicists' (Dorschel, 2021, p. 1). Aiming to provide foundations for a knowledge framework for professional standards in data science, Fayyad and Hamutcu (2020) provide an overview of the emergence and current state of the field in industry. They too describe the multidisciplinary nature of data science, with data scientists blending wide-ranging knowledge with a diverse set of skills (Fayyad & Hamutcu, 2022).

What is in no doubt is that the fields of data science and AI are fast-growing, bringing with them new forms of work. Undergoing processes of construction, yet already institutionalised in various ways globally, there are a wide array of ways to define, discuss and describe data science and AI and the associated skills, job roles, tools and methods (Davenport, 2020; Garten, 2018; O'Neil & Schutt, 2013; Passi & Sengers, 2020; Slota et al., 2020). Fayyad and Hamutcu (2022) explore the lack of classification of data science job roles and definition of occupational labels (see Section 3.1.2). We recognise that data science and AI are newly emerging, and as such ill-defined, professional fields. To delineate the scope of our study at the outset, the working definitions we use here are as follows:

> *Data science:* Using data to achieve specified goals by designing or applying computational methods for inference or prediction. (Fayyad & Hamutcu, 2020)

> *Artificial Intelligence:* When a machine or system performs tasks that would ordinarily require human (or other biological) brainpower to accomplish. (The Alan Turing Institute, 2021)

To understand the gender dynamics of AI professions, it is worth recalling that feminist scholars have been examining the ways in which technical knowledge and expertise are associated with masculinity for decades (Cockburn, 1985; Wajcman, 1991). Relatedly, the sociological literature on professions (Abbott, 1988) has been critiqued for failing to recognise the role of professionalisation in ensuring men's monopoly over the provision of certain skills and competencies in various occupations (Charles & Grusky, 2004; Witz, 1992). Building on this theoretical work, a number of writers have highlighted the role of gender relations in the very definition and configuration of computing as a profession. As feminist historian Hicks (2017) recalls, computer programming was originally the purview of women. At the advent of electronic computing following the Second World War, software programming in industrialised countries was largely considered 'women's work', and the first 'computers' were young women. However, as computer programming became professionalised, structural discrimination shifted this, edging women out of the newly prestigious computing jobs (Ensmenger, 2012; Misa, 2010; Thompson, 2019). As Hicks (2017, p. 313) explains, 'histories of hidden or devalued computing labor connect powerfully with current trends in information technology and prompt questions about the categories of privilege that silently structure our computing systems today'.

What is important to emphasise here is that technical skill is often deployed as a proxy to keep certain groups in positions of power (Abbate, 2012). Technical cultures and gender stereotypes encourage discrimination and occupational segregation within computing and engineering (Acker, 1990; Segovia-Pérez et al., 2019; Wynn & Correll, 2018). As we have noted above, the skill profiles that are associated with data science and AI are still not well understood and are continually changing though occupational practices. This study, therefore, examines the extent to which gender–technology relations are being disrupted or reproduced in new fields. We want to rewrite the narrative, heightening awareness of the gendered history of computing to avoid its replication in the nascent fields of AI and data science. This is crucial if we are to ensure that women take their rightful place at forefront of technological innovation.

As feminist Science and Technology Studies (STS) scholars have emphasised, the gender gap in AI is not only an equal opportunity issue, but also a matter of how the world we live in is designed and for whom (Wagman & Parks, 2021). Technologies, whether hardware or software, are socially shaped by gender power relations and cultural beliefs that influence the design, technical content and use of such artefacts. Automated decision-making systems employing algorithms thus pose the risk of reflecting and amplifying existing patterns of gender inequities (Kuhlman et al., 2020; Leavy, 2018). Indeed, as we argue elsewhere, the stark lack of diversity in the AI workforce results in a feedback loop whereby gender bias is encoded into machine learning systems (Wajcman & Young, 2023). At a time when data science and AI are marketed as the solution to all social problems, we must include a wide range of perspectives and experiences in the professional AI workforce to build better and more inclusive technologies.

Women in AI and data science: What does the existing data tell us?

While research on the gendering of technical occupations typically examines computing and engineering (e.g., Segovia-Pérez et al., 2019), much less is known about the subfields of data science and AI. This is largely because of the lack of clarity in defining and classifying them, in turn stunting the collection of data about the demographics of such new professions. As West et al. (2019, pp. 10–13) note, 'the current data on the state of gender diversity in the AI field is dire…The diversity and inclusion data AI companies release to the public is a partial view'. National labour force statistics also lack detailed information about job titles and pay within these fields. For example, a recent paper by Handel (2022) analysing the United States of America (US) Bureau of Labor Statistics notes the rapid growth of statistician jobs and the emergence of data scientist as a recognised occupation, but does not contain information on gender differences.[2] This is a significant barrier to research and understanding.

Given the scarcity of raw data available, researchers have drawn on other sources including online data science platforms, surveys and academic and conference data (e.g., Freire et al., 2021, see Case Study from World Economic Forum, 2020). These approaches provide mounting evidence of serious gaps in the gender diversity of the AI research and development workforce. For example, an independent survey of 399 data scientists by the recruiting firm Burtch Works found that 15% were women, although this figure shrank to 10% for those in the most senior roles (Burtch, 2018). Additionally, Gheorghe's (2022) State of Data Science and Machine Learning Report suggests that the industry is highly imbalanced, with 23% women and 77% men responding to the survey. Figure 1 also shows gendered differences in response rates between countries.

In 2018, WIRED and Element AI reviewed the AI research pages of leading technology companies and found that only 10%–15% of machine learning researchers were women (Simonite, 2018). Notably, Google's AI pages listed 641 people working on machine intelligence, but only around 60 were women. Related research found that on average only 12% of authors who had contributed work to the leading three machine learning conferences (neural information processing systems, international conference on machine learning and international conference on learning representations) in 2017 were women (Mantha & Hudson, 2018; Simonite, 2018) (see Figure 2).

Honing in on women working in AI research in academia, Stathoulopoulos and Mateos-Garcia (2019) found that only 13.8% of AI research paper authors were women: at Google, it was 11.3%, at Microsoft, 11.95%, with a slightly higher proportion of women at IBM (15.66%). Additionally, men constitute a majority of AI department faculty, making up 80% of AI professors on average (Perrault et al., 2019). Moreover, diversifying AI faculty along gender

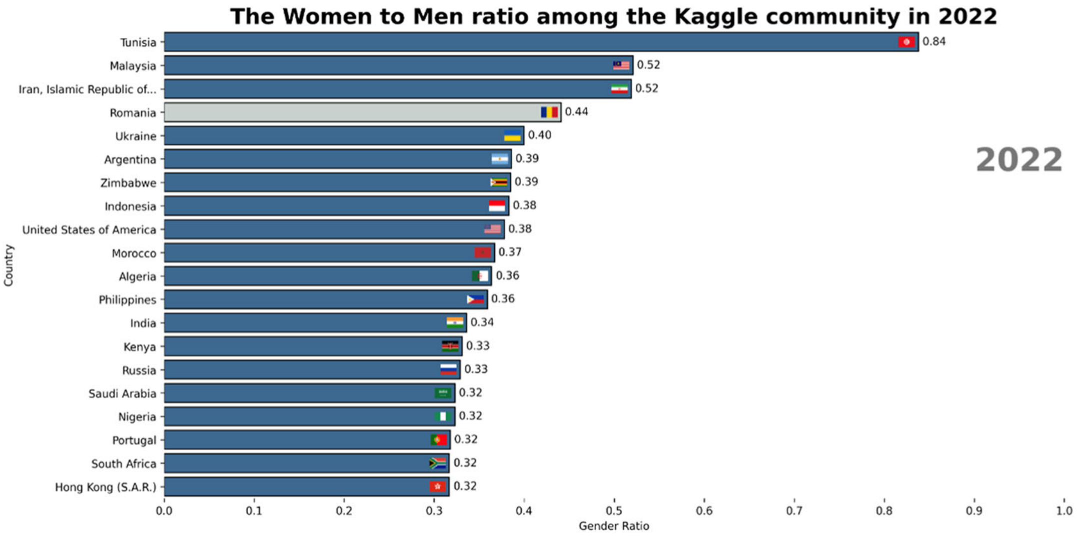

FIGURE 1 The gender imbalance in the 2022 Kaggle community across artificial intelligence. research across 20 countries. https://www.kaggle.com/code/loredanagheorghe/zoom-on-gender-in-the-kaggle-world. [Color figure can be viewed at wileyonlinelibrary.com]

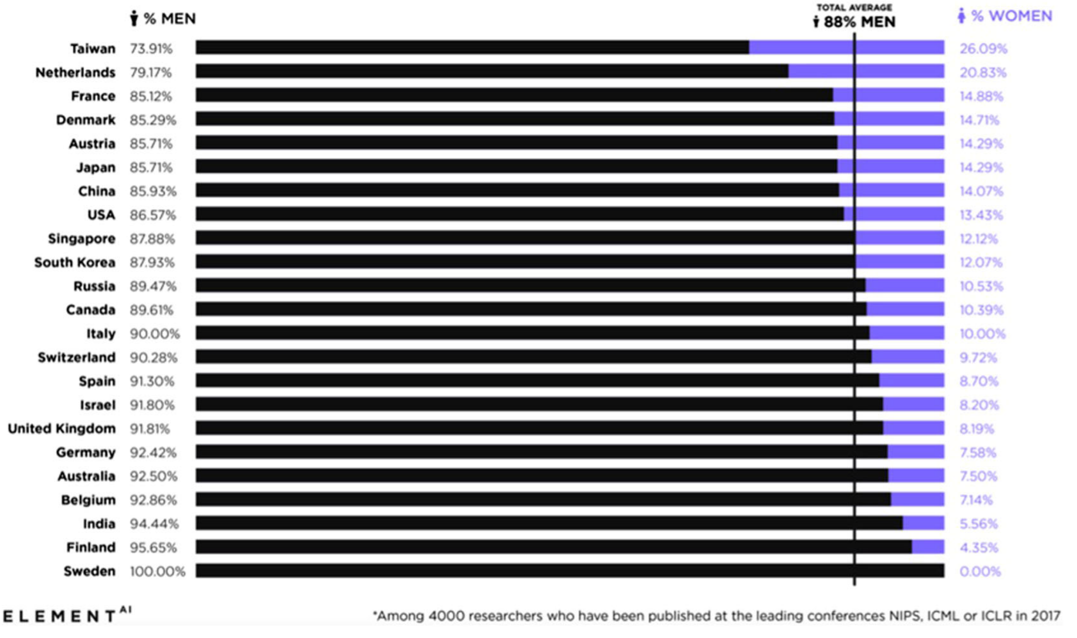

FIGURE 2 The gender imbalance in artificial intelligence (AI) research across 23 countries. Chart by Element AI: https://medium.com/element-ai-research-lab/estimating-the-gender-ratio-of-ai-researchers-around-the-world-81d2b8dbe9c3. [Color figure can be viewed at wileyonlinelibrary.com]

lines has not shown significant progress—with women comprising less than 20% of the new faculty hires in 2018.

The statistics and data we have reviewed confirm that the 'newest wings of technology', that is, data science and AI, have poor representation of women (West et al., 2019). D'Ignazio and Bhargava (2020) point out that 'white men remain overrepresented in data-related fields, even as other STEM fields have managed to narrow their gender gap' (p. 207). In other words, the more prestigious and vanguard the field, the fewer the number of women working in it. As the AI and data science fields are rapidly growing as predominant subfields within the tech sector, so is the pervasive gender gap within them. To fully grasp the nature of this problem, we urgently need better data and analysis.

METHODS

We now describe our methodology, which involves using a novel, large and diverse data science and AI career data set. Due to a lack of accessible data on data and AI professionals, and to curate a data set suitable for responsibly investigating gender gaps in these industries, we partnered with Quotacom, an executive search and consulting firm specialising in data science, advanced analytics and AI. From there, we developed a methodology to first identify the relevant data profiles, second obtain information on their career trajectories from LinkedIn, third process their education, work experience and skills into manageable categories and fourth test for group differences by gender.

Data collection

Initial seed database

It is crucial to understand the sources and methods from which our initial data seed list was created to ensure an ethical approach and robustness in our findings. We interviewed Quotacom about their data sources and data collection methods to understand potential biases in our sample. The Quotacom data set consists of more than 10,000 'Candidates' (potential recruits) and 90,000 'Contacts' (company contacts), that voluntarily subscribed, either searching for a job or for potential hires. Quotacom scouts across industries, focusing particularly on the data pipeline in Europe, the Middle East and Africa, US and Asia-Pacific. Data were collected over the last 5 years, and a general data protection regulation-compliant privacy notice was provided to candidates and contacts before signing up to the database. Each person's job title and LinkedIn profile are provided.

Quotacom creates talent lists for candidates through the use of X-ray, Lusha, Owler, Skrapp, LinkedIn/Xing or similar, personal networks, referrals, recommendations, websites, industry forums, blogs, competitions, speaker lists, conference attendee lists and industry press. They then approach candidates via Internet-sourced contact details. Alternatively, candidates can approach Quotacom via responses to advertisements, although they are not typically added to the database unless they have relevant skills within digital transformation, data science or AI.

Contact profiles—that is, individuals based at Quotacom's partner companies—are sourced in a different way. Companies are initially added as target prospects. Quotacom then perform various outreach marketing campaigns to stakeholders within those companies, usually via email, phone and LinkedIn. Quotacom typically uses LinkedIn, business directories, CrunchBase and Google to develop the initial prospect companies lists. These ‘prospects’ span small to large companies, and there are no specific criteria apart from the fact that they operate or have specialist business units in digital or data. Once prospect lists are compiled in Excel, they are loaded onto the central database via CSV.

Identifying and classifying data and AI profiles

Despite Quotacom's focus on data and AI companies, we found that many ‘contacts’ did not sit squarely in the data science field (e.g., HR administrators, sales executives and account managers). Consequently, we selected relevant profiles using the International Standard Classification of Occupations (ISCO-08) categorisations from the international labour organization (ILO), which categorises jobs into different groups according to their tasks and duties. We evaluated and compared multiple occupational classifications alongside ISCO-08; these were Standard Occupation and Classification (SOC) system, Occupational information Network (O*NET) and European Skills, Competences, Qualifications and Occupations (ESCO). Although none were entirely appropriate for the new jobs emerging from the data science and AI fields (see Section 5), we chose ISCO-08 since it contains data science roles—whereas others do not—and our data set is international. Next, we matched profiles' job titles with these groups, keeping only those belonging to the ILO codes ‘25 (Information and Communication Technology Professionals)’ and ‘133 (Information and Communication Technology Service Managers)’. The matching was done by standard text preprocessing—that is, lowercase, stop words removal and stemming—together with word vectors and similarity scores to find the closest standard title for each profile and its sub-major category.[3] To ensure that the ILO job list covered most data science job titles, we did a first round of matching in which we sampled 1000 profiles, looking for data-related job titles that were misclassified. We then added these to the standard ILO job list and performed a new matching. This way, we managed to match over 80,000 unique job titles to their 7000 titles and 43 subgroups, which left us with 22,373 data profiles. We tested the precision in our detection by randomly sampling 1000 of the selected profiles and looking to see if the job titles were correctly matched. Out of those, only 92 were wrongly classified (90.8% precision). Similarly, we estimated a 76% recall (i.e., how many data profiles were left out of our sample) by manually validating a random sample of 1000 profiles from our complete list.

LinkedIn

LinkedIn is the world's largest professional network with nearly 740 million members in more than 200 countries and territories worldwide, hosting self-reported information on individual's professional and educational backgrounds and skills. As both Case et al. (2012) and Li et al. (2019) recognise, the LinkedIn database is a valuable information repository for understanding the dynamics of careers. Indeed, since our study was carried out, Fayyad and Hamutcu (2022) have also used a LinkedIn sample (2726 US- and UK-based profiles) to categorise roles in the data science field. However, as pointed out below, our sample is both larger and more diverse.

Following ethical approval, we scraped LinkedIn to collect the complete educational, professional and skill set information of the individuals on our reduced list of profiles.[4]

No personal information, such as phone numbers and email addresses, was collected, and data was fully anonymised in storage. It is important to note here that the vast majority of LinkedIn information is self-reported and optional. As such, some information may well be missing, exaggerated or subject to different qualification standards (e.g., when stating proficiency in a particular skill). We mitigate these by looking at gender differences in the aggregated data and focusing on the relative gaps.

Data cleaning and characterisation

As stated, one of our major concerns when dealing with LinkedIn data is its level of completeness, especially when each field of information is 'optional'. To ensure comparability between users, we only considered profiles with some professional experience, and with at least 50 contacts. We also removed 25 outlier profiles who reported more than 45 total years of experience.

Our final sample consists of 19,535 profiles, out of which 2203 (11.3%) are women, mostly from the United States, France, Germany or the United Kingdom (countries in the top 10 in AI rankings). Our exploratory analysis showed that, as anticipated by Quotacom, our sample is skewed towards seniority, with an average of almost 20 years of work experience. Further, over 55% of our sample hold a graduate or postgraduate degree (see Table 1).

Variables and data processing

This section explains the processing steps taken for each variable in our analysis (see Table 2), with some descriptive statistics.

Gender

To infer each profile's gender, first names were passed to an API that returns a gender with a probability score (Genderize API). The API, designed to predict the gender probability of a person given their name, is based on more than 100 million data points collected across 242 countries. Even though we are aware of the limitations of this method (see Section 5), given the lack of information on self-perceived gender (Karimi et al., 2016; Terrell et al., 2017), we adopt a probability-based approach based on names rather than alternatives such as image recognition or scraping personal webpages. To maximise precision, we only kept profiles with more than 80% probability and removed the ones we could not classify (less than 1%). This left us with 11% women in our data.

TABLE 1 Characterisation of the sample.

	Female	Male	Graduate degree	Senior jobs
% of total	11.3%	88.7%	55.6%	59.2%
N	2203	17,332	8793	10,431

TABLE 2 Complete list of variables and their sources.

Variable	Source	Description
LinkedIn profile	Quotacom	LinkedIn URL
Gender	Genderize API	Inferred gender (binary)
Job history	LinkedIn	Includes: self-declared job title, company, industry and years.
Seniority	Own authors	Inferred from job title based on keywords
Role (e.g., consultant, engineer, analyst)	Own authors	Inferred from job title based on keywords
Industry	LinkedIn	Industry associated with each job company
Start and end date	LinkedIn	Start and end date of employment
Education history	LinkedIn	Includes: self-declared degree, discipline, institution and years
Max degree	Own authors	Maximum degree achieved. Classified into undergrad, postgrad and none
Skills (LinkedIn)	LinkedIn	List of self-declared skills and their LinkedIn categories
Data skills	Own authors	Subset of data skills and their category based on HDSR
Location	LinkedIn	Inferred location based on their last job

Abbreviation: HDSR, harvard data science review.

Location

We used the last available job location for each profile to determine their country of residence, assisted by *pycountry* in cases where the city name or country code was mentioned instead of the country name. We found that 50% of our sample corresponded to the United States (22.7%), France (10.7%), Germany (10.1%) and the United Kingdom (9%), with no significant differences in gender gaps between them in the years of experience, roles duration or number of skills. The other 50% is divided between 14 countries that make up 1%–5% of the sample each, and 50+ more countries at under 1% each.

Work experience, seniority and job fields

For each profile, we scraped available job history including job title, start and end dates, company name, industry and location. Using each role duration, we estimated the total years of experience within the same company, industry and across whole careers. Further, we used last available work experience to infer our sample's seniority by classifying job titles into five different categories (see Table 3). As anticipated, we found that our sample is very senior, with over 50% having CXO roles, as well as a trajectory of 20 years of experience across seven different roles (see Table 4).

Finally, we looked at job fields by classifying all job titles into Consultancy, Engineering, Development, Analytics, Architecture, Science and Research. We should note that this

TABLE 3 Keywords used for seniority classification.

Seniority	Keywords
Junior	Junior, assistant, intern, trainee, associate
Mid	Lead, manager, supervisor, project director
Senior	Senior, executive, director, head, principal
CXO	Chief X officer, CXO
Board	VP, president, chairman, board, founder, partner, owner

TABLE 4 Work experience statistics after removing outliers with more than 3.5 standard deviations over the mean years of experience.

	Total years of experience	Number of different roles	Number of different companies	Number of industries
Mean (SD)	19.88 (7.22)	7.32 (2.53)	5.29 (2.43)	3.64 (2.71)
Minimum	1.00	1	1	1
Maximum	45.33	17	14	13

categorisation was only made for Junior, Mid- and some Senior roles, given that generally this does not make sense for CXO and Board roles.

Industry

When available, we used the industry from each company's LinkedIn page associated with our profiles' jobs. We then grouped 147 unique LinkedIn industry codes into 13 major categories and looked at the gender distribution of roles in each. Figure 3 shows the number of profiles that have held at least one job in each industry. Unsurprisingly, Technology/Information Technology (IT) is the most common, with over 50% of the individuals having worked in a tech company.

Skills

LinkedIn allows users to add up to 50 skills, and automatically classifies them into one of five categories: Industry Knowledge, Tools and Technologies, Interpersonal Skills, Languages and Other Skills. We found that 7% of our sample had no skills on their LinkedIn, with little difference by gender (6.9% for men and 7.4% for women). For the rest, the prevalence of the types of skills is very uneven, with industry skills encompassing over 60% of the sample (see Figure 4A). To specifically detect data science and AI skills, we used the framework proposed by Fayyad and Hamutcu (2020) to reclassify all skills, creating eight new data categories. In our overall sample, data skills represent over 15% of the total (see Figure 4B).

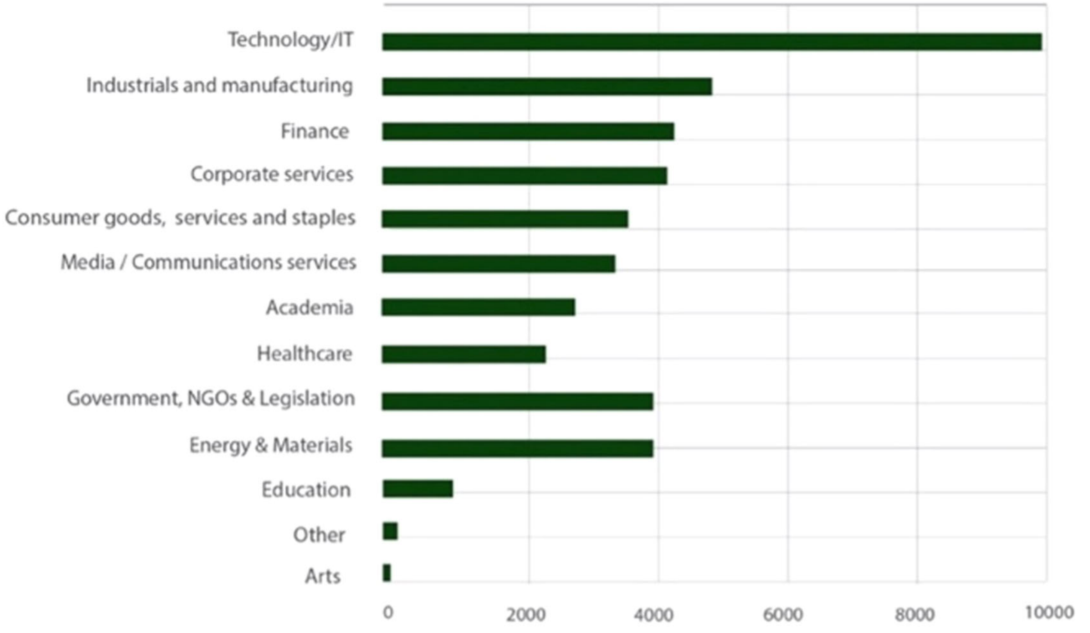

FIGURE 3 Number of profiles who have held at least one job by industry. [Color figure can be viewed at wileyonlinelibrary.com]

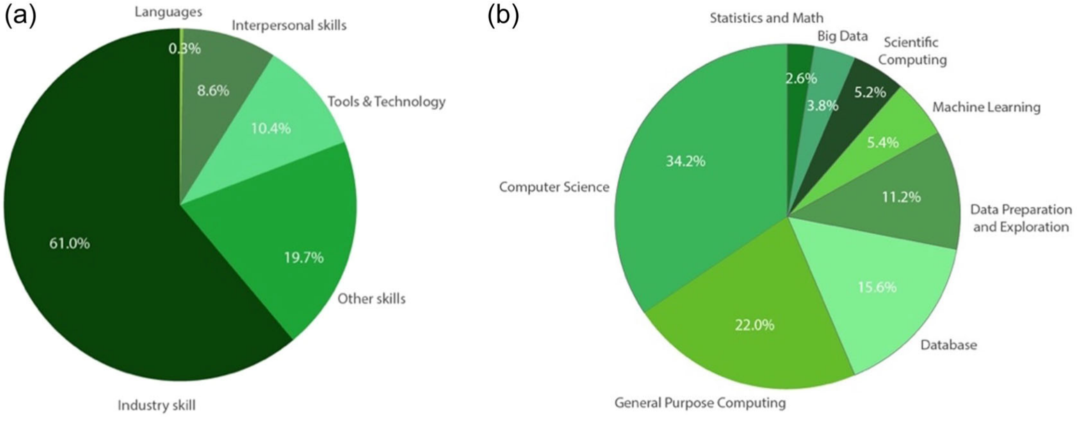

FIGURE 4 (A) Distribution of skills, as classified by LinkedIn across the whole sample. (B) Distribution of data science skills, as classified by Fayyad and Hamutcu (2020). [Color figure can be viewed at wileyonlinelibrary.com]

FINDINGS AND DISCUSSION

Diverging career trajectories

Our research suggests that there is structured gender inequality in the career trajectories of professionals in the data science and AI fields. Women are more likely than men to occupy a job associated with less status and pay in these fields. Figure 5 shows that women have more

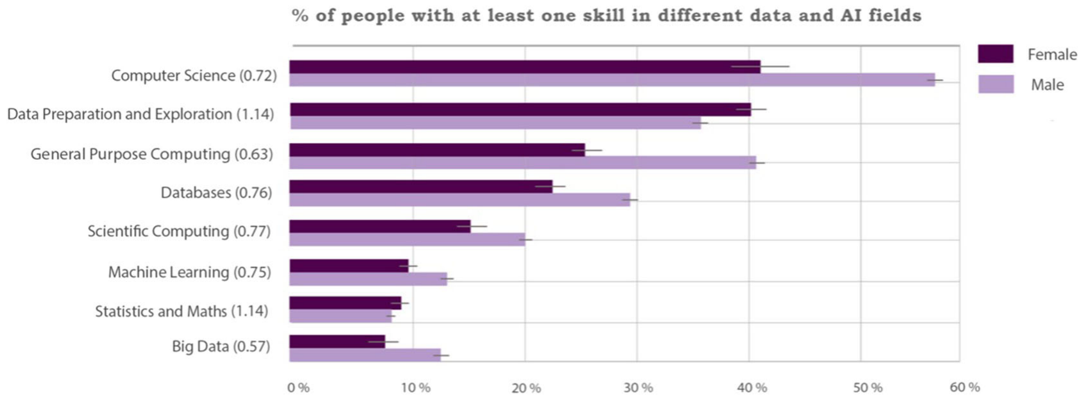

FIGURE 5 Percentage of people with at least one skill in data and AI fields. Numbers in brackets represent the gender gap (female/male) (gender gaps are calculated by dividing % female by % male, for example, for every 100 men with general purpose computing skills, there are only 63 women). [Color figure can be viewed at wileyonlinelibrary.com]

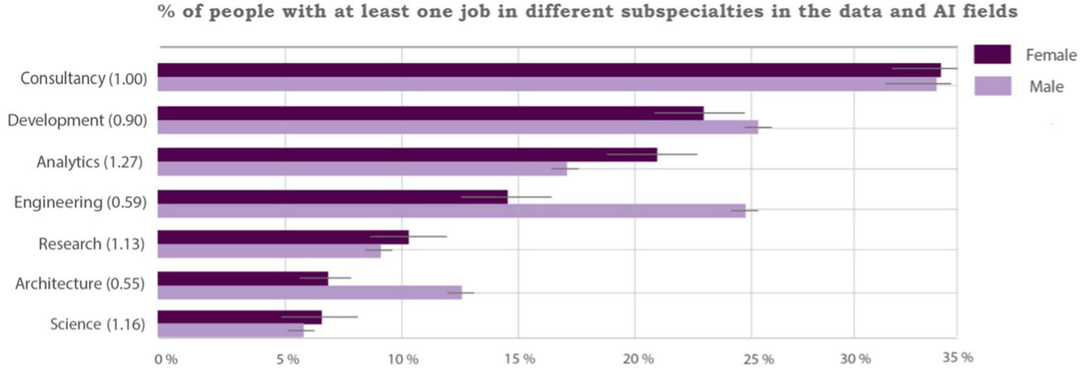

FIGURE 6 Percentage of people with at least one job in different subspecialties in the data and AI fields. Numbers in brackets represent the gender gap (female/male). [Color figure can be viewed at wileyonlinelibrary.com]

data preparation and exploration skills, whereas men have more machine learning, big data, general purpose computing (GPC) and computer science skills (all significant at 95% confidence level with a two-sample z test). The latter are traditionally associated with more prestigious and higher paying careers.

Consistent with our review of skills, we find that men predominate in Engineering, Architecture and Development jobs, while women do so in Analytics and Research (all significant at a 95% confidence level) (see Figure 6). This is interesting to consider alongside the 2021 report 'Quantifying the UK Data Skills gap' which found that women workers believe they are generally strong in 'soft skills', such as 'communication' and 'curiosity', while men claim to be stronger in 'industry/sector expertise' (Department for Digital, Culture, Media and Sport, 2021).

Campero (2020) also found that women are more prevalent among workers in software quality assurance—crucially, lower-paying and perceived as lower-status—than in other software subspecialities. Similarly, Guerrier et al. (2009) note that women are underrepresented in high-skilled IT jobs, with intraoccupational gender segregation: they are located in the less technical project management and customer support roles that require the sorts of skills that women are 'naturally' thought to have. Indeed, as feminist scholars have shown, when women participate in male-dominated occupations, they are often concentrated in the lower-paying and lower-status subfields (Reskin & Roos, 1990). So it would seem that as women have begun to enter certain technological subdomains in recent years, such as front-end development, these fields have started to lose prestige and experience salary drops (Broad, 2019; Posner, 2017). Meanwhile, men are flocking to the new (prestigious and highly remunerated) data science and AI subspecialities. Indeed, the World Economic Forum (2018a) warns about 'emerging gender gaps in Artificial Intelligence-related skills' (see Figure 7). Our results are consistent with their findings that a higher proportion of women than men are data analysts, and higher proportions of men than women are engineers and IT architects. They similarly found that a higher proportion of men have machine learning skills.

Such studies suggest that there is a hardening talent gap that will require focused intervention. In their report proposing elements of a framework on gender equality and AI, UNESCO (2020, p. 27) point out that 'hiring more women is not enough. The real objective is to make sure that women are hired in core roles such as development and coding'. They recommend the need to bring to positions of parity women coders, developers and decision-makers, with intersectionality in mind. 'This is not a matter of numbers, but also a matter of

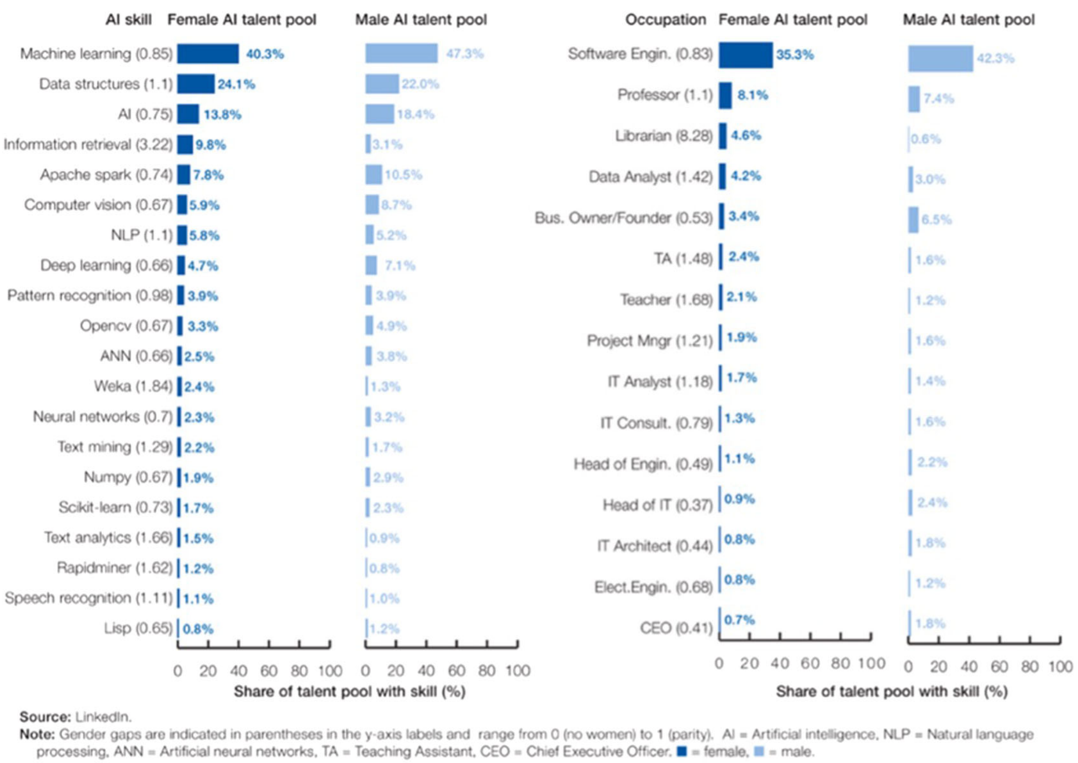

FIGURE 7 'Share of female and male AI talent pool, by AI skill', and 'Share of LinkedIn members with AI skills, by occupation and gender', respectively. Source: World Economic Forum Global Gender Gap report (2018a, p. 31). [Color figure can be viewed at wileyonlinelibrary.com] culture and power, with women actually having the ability to exert influence' (UNESCO, 2020, p. 23).[5]

Industry differences

Women in data and AI are underrepresented in industries which traditionally entail more technical skills (e.g., the Technology/IT sector), and overrepresented in industries which entail fewer technical skills (e.g., the Healthcare sector). Furthermore, there are fewer women than men in C-suite positions across most industries in the data set, and this is even more marked in data and AI jobs in the tech sector.

Our findings suggest that patterns in AI and data science are similar to gender gaps in the overall workforce. Female AI professionals in our sample are more likely to work in 'traditionally feminised' industries such as Healthcare. Figure 8 shows that this is also true for the Corporate Services (e.g., HR and marketing), and Consumer Goods industries. However, women are underrepresented in the Technology/IT and Industrials and Manufacturing sectors.

Notably, women's participation across different industries is inversely correlated with the percentage of 'Tools and Technologies' skills that they hold (Pearson R of −0.7, $p = 0.04$) (Figure 8). Thus, we found that those industries with lower female participation are also the ones with the higher proportion of 'Tools and Technology' skills in women's profiles.

Again, our findings are broadly consistent with Broad (2019): within the AI talent pool, she found more women than men in the Healthcare industry, and more men in the Manufacturing and Software and IT Services sectors.

While our data cannot explain the causation behind this finding, it mirrors what we already know about the reasons for women's underrepresentation in industries which traditionally involve more technical skills. As noted above, stereotypically masculine norms and value systems shape professional practices and career pathways (Muzio & Tomlinson, 2012). These 'masculine defaults' particularly govern technical participation (Cheryan & Markus, 2020), as definitions of technological expertise have been historically framed in a way that renders the feminine as 'incompatible with technological pursuits' (Wajcman, 2010). Such persistent cultural associations around technology drive women away from, and out of, industries which entail more 'frontier' technical skills such as data science and AI.

Finally, it is important to note that men consistently predominate in CXO positions across most industries, regardless of the level of general industry participation (Figure 9). Even in industries where women are overrepresented, there is a lower percentage of women in the C-suite.

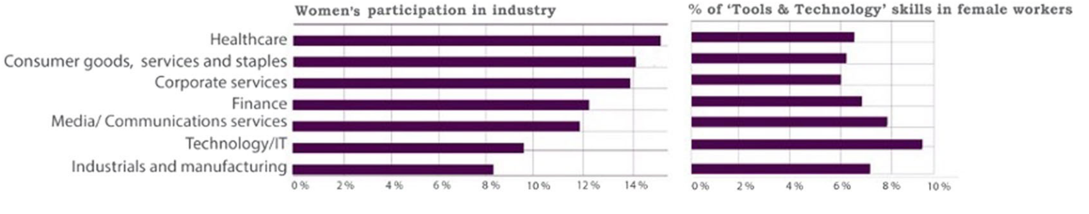

FIGURE 8 Women's participation in industry, and % of 'Tools and Technologies' skills held by women workers by industry, respectively. Only industries with a sample of at least 100 female and male profiles each are shown. [Color figure can be viewed at wileyonlinelibrary.com]

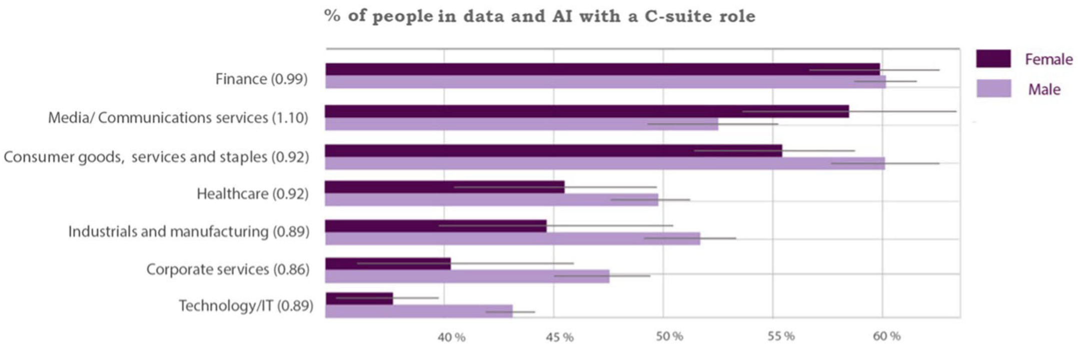

FIGURE 9 Percentage of men and women in data and AI with a C-suite role, by industry. Numbers in brackets represent the gender gap (female/male). [Color figure can be viewed at wileyonlinelibrary.com]

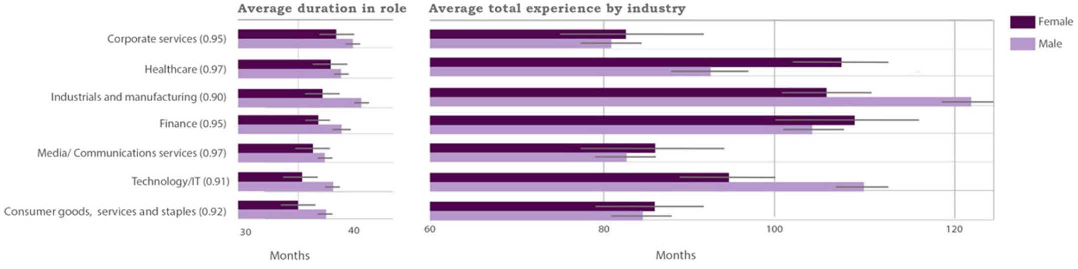

FIGURE 10 Average duration in role by industry, and average total experience in industry by gender, respectively. [Color figure can be viewed at wileyonlinelibrary.com]

Best et al.'s (2019) study of women's participation in leadership among top AI companies found that women only represent 18% of C-level leaders among top AI start-ups across much of the globe. Of the 95 companies they considered, only two have an equal number of women to men in their C-level positions, and none are majority women.

Job turnover and attrition rates

Our study found evidence of persistently high turnover (changing job roles) and attrition rates (leaving the industry altogether) among women as compared to men working in data science and AI in the technology industry. The data shows that, on average, women spend less time in each role than men do (see Figure 10). This holds for every industry, with the biggest gap in the Industrials and Manufacturing, and Technology/IT sectors. Looking at the total years of experience spent in each industry by gender, on average women spend more time than men in every industry except for Industrials and Manufacturing, and crucially, the Technology/IT sector, where they spend almost a year and a half less.

There has been some interesting research on gendered attrition from engineering and technology firms. The US National Center for Women and Information Technology found that women leave technology jobs at twice the rate of men (Ashcraft et al., 2016). Cardador and Hill (2018) comparably show that women (but not men) taking managerial paths in engineering firms may be at the greatest risk of attrition. In a similar vein, Krivkovich et al. (2016) found that women made up 37% of entry-level roles in technology, but only 25% reached senior management roles and 15% made executive level.

Exploring the reasons for women's and minorities' high attrition and turnover rates, the Kapor Center argues that unfairness drives turnover, highlighting that 1 in 10 women in technology has reported experiencing unwanted sexual attention (Scott et al., 2017). Indeed, as much research attests, reasons include unwelcoming environments, workplace discrimination and microaggressions, gendered domestic commitments and cultural associations about who 'fits' in technology fields (Alfrey & Twine, 2017; Margolis & Fisher, 2012; Wu, 2020). According to the State of European Tech Survey, 87% of women have been challenged by gender discrimination compared to 26% of men, and 59% of Black/African/Caribbean women have experienced discrimination in some form (Atomico, 2020).

Self-reported skills

We found that women are more likely to self-report fewer skills than men on LinkedIn. We tested this with both a two-sample *Kolmogorov–Smirnov* and *t test*. We further tested this finding across all industries and countries within our sample and obtained the same results.

This echoes Stanford's AI Index report based on LinkedIn data which found that, across all countries, men tend to report AI skills across more occupations than women (Perrault et al., 2019). Further, the World Economic Forum (2018b) notes that there are '…no signs that this gap is closing: over the past four years, men and women have been adding AI skills to their [LinkedIn] profiles at a similar rate. While women aren't falling further behind, they also aren't catching up'.

Other studies have also found that women are less self-promoting (Lerchenmueller et al., 2019). They also indicate that women are generally less confident in their own abilities, particularly during self-assessment (Cech et al., 2011; Correll, 2001). As touched upon earlier, persistent cultural associations around femininity as 'incompatible' with advanced technological pursuits (e.g., alongside 'brogrammer' stereotypes and 'hustling') affect women's confidence in their technical skills, shaping perceptions of their aptitude and proficiencies (Jacobs, 2018).

Altenburger et al. (2017) take this point further to speculate as to how these gender differences in self-assessment and self-presentation might affect online professional opportunities, for example on LinkedIn. Women's less favourable assessments of their abilities, fit and belonging in male-dominated data science and AI occupations may well be influential in determining women's aspirations in these fields (see also Leslie et al., 2015; Wynn & Correll, 2017).

The qualification gap

Given the above findings, it is striking that our research discovered that women in data and AI have higher formal educational levels than men, and this is the case across all industries. The achievement gap is even higher for those in more senior ranks (e.g., for C-suite roles), and this 'overqualification' aspect is most marked in the Technology/IT sector. This is particularly surprising given that women self-report fewer skills.

We find that 59% of women in our sample hold a graduate (or postgraduate) degree, compared to 55% of men. This trend also holds when the sample is broken down by industry.

Further, when we compared the formal educational levels of our whole sample with a subsample of the most senior profiles, we found that the educational gap is even higher for those at C-Suite level. In fact, the gap is roughly double in every industry; by which we mean that, for instance, in all Technology/IT roles, there is an achievement gap of 6%, but for CXO roles, this shoots up to 13%. In the case of the Technology/IT industry, the leap is mostly explained by an increase in the percentage of graduate women in the C-suite.

This strongly suggests that women are educating themselves to get promoted, while men may not be doing so. The finding is in line with existing evidence that women need to work harder and earn more qualifications than men to progress into senior ranks in the workplace (Scott, 2021).

LIMITATIONS

We acknowledge some limitations in the original data set. These data are skewed towards four countries; although this reflects where many data jobs are located (Bughin et al., 2019), this limits the generalisability of our findings. The data set's bias towards senior professionals also limits the findings around more junior practitioners, and does not include, for example, the disproportionate numbers of women who prematurely leave such professions. Analysis is bound by the limits of the classified, binary data as collected (D'Ignazio & Klein, 2020), but gender intersects with multiple characteristics including race and age (Collins, 1998; Crenshaw et al., 1995). We decided not to computationally infer other characteristics to avoid ethical and statistical errors that might arise from inference methods such as using names or image recognition (Karimi et al., 2016; Raji et al., 2020). Additionally, limitations arise through LinkedIn data. Users' ‘self-declaration’ of their own skills, jobs and education history is inherently biased. Not only can individuals mislead or exaggerate, either consciously or unconsciously, but as our research here has shown, self-presentation on LinkedIn is gendered. Finally, the lack of standardisation of job titles in the emerging data science and AI professions brings with it data and thus methodological limitations, as noted above. Titles across all seniority levels differ between companies and across sectors. Outdated occupational classification systems limit this kind of research, and we need better categories to accurately reflect the changing nature of digital work (see Section 3.1.2).[6]

CONCLUSION

Debates about automation and the future of work generally focus on job losses and deskilling (Wajcman, 2017). When questions about the gender relations of digital technology are canvassed in this context, the research tends to highlight women's exclusion from scientific and technical occupations. This article aims to enrich these discussions by examining the terms of women's inclusion in new technical fields that are increasingly central to the digital society. Specifically, we ask how women are faring in these professions which are still very much in their formative stages.

This article thus contributes to the literature on gender inequality in data science and AI careers, a subject about which much more needs to be known. First, as emerging fields, there is a significant gender data gap and even a problem of definition as to what data and AI occupational labels designate. Our research provides a unique, curated data set of data science and AI professionals and analyses it through innovative methodology. Second, on the basis of this data, we describe, in more detail than hitherto, the persistent structural inequality in these fields associated with extensive disparities in jobs, qualifications, seniority, industry, attrition, skills and even self-confidence levels between men and women. The gendered career trajectories in data science and AI that we identify reflect the historical association of masculinity with computing, in particular, frontier technical skills. Our work is original in examining these new prestigious high-tech professions as they are being formed, and as the gendering processes within them are being configured.

If women are to fully participate in the data science and AI workforce, the gender job gap needs rectifying. This is particularly important in the leadership and technical roles which drive innovation and the development of AI systems. As we among others argue, inequality in these fields is not only about ethical issues of social justice and economic opportunity, but also crucially about how technologies such as AI are designed. Research has shown how a lack of diversity can lead to poor outcomes across many different fields, including healthcare (Koning et al., 2019), scientific research (Nielsen et al., 2018) and the private sector (Noland et al., 2016) and AI is no exception. Indeed, mounting research suggests that gender bias is being built into, and amplified by AI systems. This makes closing the gender gap in data science and AI ever more urgent.

ACKNOWLEDGEMENTS

The authors would like to thank *Quotacom*, an international executive search and consulting firm in digital transformation and data, for providing the seed data set for our research.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

ENDNOTES

[1] 'Generative AI' refers to deep learning models that can generate new content, including audio, code, images, text, simulations and videos, based on the data on which they were trained.

[2] Handel reports that figures for data scientists are not available by gender (Personal communication 11 September 2022).

[3] To deal with the language variations of our international data set, we did iterative matching rounds, reviewing a sample of unmatched job titles and adding unknown terms and categories.

[4] Code for the LinkedIn scraping is available at github.com/sprejerlaila/linkedInScraping/.

[5] The AI industry must avoid participation-washing; somebody merely participating does not lend ethical legitimacy (Mitchell et al., 2020; Sloane et al., 2022).

[6] Promising initiatives include the 'Data Science Professional Profiles' (DSPP) report, which provides a list of occupations for inclusion in the ESCO taxonomy (Cuadrado-Gallego & Demchenko, 2020). Fayyad and Hamutcu (2020) also note the Initiative for Analytics and Data Science Standards (IADSS), (www.iadss.org) a knowledge framework for professional standards in data science.

REFERENCES

Abbate, J. (2012) *Recoding gender: women's changing participation in computing*. Cambridge: MIT Press.
Abbott, A. (1988) *The system of professions*. Chicago: University of Chicago Press.
Acker, J. (1990) Hierarchies, jobs, bodies: a theory of gendered organizations. *Gender & Society*, 4(2), 139–158.

Alfrey, L. & Twine, F.W. (2017) Gender-fluid geek girls: negotiating inequality regimes in the tech industry. *Gender & Society*, 31(1), 28–50. Available from: https://doi.org/10.1177/0891243216680590

Altenburger, K.M., De, R., Frazier, K., Nikolai, A. & Hamilton, J. (2017) Are there gender differences in professional self-promotion? An empirical case study of LinkedIn profiles among recent MBA graduates. *Proceedings of the International AAAI Conference on Web and Social Media*, 11(1), 460–463. Available from: https://doi.org/10.1609/icwsm.v11i1.14929

Ashcraft, C., Mclain, B. & Eger, E. (2016). *Women in tech: the facts*. Colorado, USA: National Center for Women and Information Technology (NCWIT), p. 76. Available from: https://www.ncwit.org/sites/default/files/resources/womenintech_facts_fullreport_05132016.pdf [Accessed 2nd September 2022].

Atomico. (2020). *The state of European Tech 2020*. Available at: https://2020.stateofeuropeantech.com/ [Accessed 15th August 2022].

Avnoon, N. (2021) Data scientists' identity work: omnivorous symbolic boundaries in skills acquisition. *Work, Employment and Society*, 35(2), 332–349. Available from: https://doi.org/10.1177/0950017020977306

Berman, F.D. & Bourne, P.E. (2015) Let's make gender diversity in data science a priority right from the start. *PLoS Biology*, 13(7), e1002206. Available from: https://doi.org/10.1371/journal.pbio.1002206

Best, M.L. & Modi, D. (2019) Case study 4: where are the women? Gender disparities in AI research and development. In: Sey, A. & Hafkin, N. (Eds.) *Taking stock: data and evidence on gender equality in digital access, skills, and leadership*. Preliminary findings of a review by the EQUALS Research Group. Macau SAR: EQUALS Global Partnership, United Nations University.

Blackburn, H. (2017) The status of women in STEM in higher education: a review of the literature 2007–2017. *Science & Technology Libraries*, 36(3), 235–273. Available from: https://doi.org/10.1080/0194262X.2017.1371658

Brandt, P.S. (2016) *The emergence of the data science profession*. PhD dissertation, University of Columbia.

Broad, E. (2019) Computer says no. *Inside Story*. Available from: https://insidestory.org.au/computer-says-no/ [Accessed 5th October 2022].

Bughin, J., Herring, L., Mayhew, H., Seong, J. & Allas, T. (2019) *Artificial intelligence in the United Kingdom: prospects and challenges*. London: McKinsey Global Institute. Available from: https://www.mckinsey.com/featured-insights/artificial-intelligence/artificial-intelligence-in-the-united-kingdom-prospects-and-challenges [Accessed 1st November 2022].

Burtch, L. (2018). *The Burtch works study salaries of data scientists*. Available from: www.burtchworks.com [Accessed 1st November 2022].

Campero, S. (2020) Hiring and intra-occupational gender segregation in software engineering. *American Sociological Review*, 86(1), 60–92. Available from: https://doi.org/10.1177/0003122420971805

Cardador, M.T. & Hill, P.L. (2018) Career paths in engineering firms: gendered patterns and implications. *Journal of Career Assessment*, 26(1), 95–110. Available from: https://doi.org/10.1177/1069072716679987

Case, T., Gardiner, A., Rutner, P. & Dyer, J. (2012) A LinkedIn analysis of career paths of information systems alumni. *Journal of the Southern Association for Information Systems*, 1(1), 1–13.

Cech, E., Rubineau, B., Silbey, S. & Seron, C. (2011) Professional role confidence and gendered persistence in engineering. *American Sociological Review*, 76(5), 641–666. Available from: https://doi.org/10.1177/0003122411420815

Charles, M. & Grusky, D. (2004) *Occupational ghettos: the worldwide segregation of women and men*. Stanford: Stanford University Press.

Cheryan, S. & Markus, H.R. (2020) Masculine defaults: identifying and mitigating hidden cultural biases. *Psychological Review*, 127(6), 1022–1052. Available from: https://doi.org/10.1037/rev0000209

Cockburn, C. (1985) *Machinery of dominance: women, men and technical know-how*. London: Pluto Press.

Collins, P.H. (1998) It's all in the family: intersections of gender, race, and nation. Hypatia, 13(3), 62–82. Available from: https://www.cambridge.org/core/journals/hypatia/article/abs/its-all-in-the-family-intersections-of-gender-race-and-nation/5021A3F35D94158AE0AD18D2B2746A42 [Accessed 8th August 2010].

Correll, S.J. (2001) Gender and the career choice process: the role of biased self-assessments. *American Journal of Sociology*, 106(6), 1691–1730. Available from: https://doi.org/10.1086/321299

Crenshaw, K.W., Gotanda, N., Peller, G. & Thomas, K. (1995) *Critical race theory: the key writings that formed the movement*. New York: The New Press.

Cuadrado-Gallego, J.J. & Demchenko, Y. (2020) *The data science framework: a view from the EDISON project*. Cham: Springer. Available from: https://doi.org/10.1007/978-3-030-51023-7

D'Ignazio, C. & Bhargava, R. (2020) Data visualization literacy: a feminist starting point. In: Engebretsen, M. & Kennedy, H. (Eds.) *Data visualization in society*. University of Amsterdam Press, pp. 207–222. Available from: https://doi.org/10.5117/9789463722902 [Accessed 2nd july 2022].

D'Ignazio, C. & Klein, L.F. (2020) *Data feminism*. Cambridge, MA: MIT Press.

Davenport, T. (2020) Beyond unicorns: educating, classifying, and certifying business data scientists. *Harvard Data Science Review*, 2(2). Available from: https://doi.org/10.1162/99608f92.55546b4a

Davenport, T.H. & Patil, D.J. (2012) Data scientist. *Harvard Business Review*, 90(5), 70–76.

Department for Digital, Culture, Media and Sport. (2021). *Quantifying the UK data skills gap—full report*. London: GOV.UK. Available from: https://www.gov.uk/government/publications/quantifying-the-uk-data-skills-gap/quantifying-the-uk-data-skills-gap-full-report [Accessed 22nd October 2022].

Dorschel, R. (2021) Discovering needs for digital capitalism: the hybrid profession of data science. *Big Data & Society*, 8(2), 205395172110407. Available from: https://doi.org/10.1177/20539517211040760

Ensmenger, N. (2012) *The computer boys take over: computers, programmers, and the politics of technical expertise*. Cambridge: MIT Press.

Fayyad, U. & Hamutcu, H. (2020) Toward foundations for data science and analytics: a knowledge framework for professional standards. *Harvard Data Science Review*, 2(2). Available from: https://doi.org/10.1162/99608f92.1a99e67a

Fayyad, U. & Hamutcu, H. (2022) From unicorn data scientist to key roles in data science: standardizing roles. *Harvard Data Science Review*, 4(3). Available from: https://doi.org/10.1162/99608f92.008b5006

Freire, A., Porcaro, L. & Gómez, E. (2021) Measuring diversity of artificial intelligence conferences. Proceedings of 2nd Workshop on Diversity in Artificial Intelligence (AIDBEI), Proceedings of Machine Learning Research, 142, 39–50. Available from: https://proceedings.mlr.press/v142/freire21a.html

Garten, Y. (2018) The kinds of data scientist. *Harvard Business Review*. Available from: https://hbr.org/2018/11/the-kinds-of-data-scientist [Accessed 10th October 2022].

Gheorghe, L. (2022) Zoom on gender in the Kaggle World. *Kaggle*. Available from: https://www.kaggle.com/code/loredanagheorghe/zoom-on-gender-in-the-kaggle-world [Accessed 27th November 2022].

Guerrier, Y., Evans, C., Glover, J. & Wilson, C. (2009) 'Technical, but not very.': constructing gendered identities in IT-related employment. *Work, Employment and Society*, 23(3), 494–511. Available from: https://doi.org/10.1177/0950017009337072

Handel, M. (2022) *Growth trends for selected occupations considered at risk from automation*. Washington DC, USA: Monthly Labor Review, U.S. Bureau of Labor Statistics. Available from: https://doi.org/10.21916/mlr.2022.21

Hicks, M. (2017) *Programmed inequality: how britain discarded women technologists and lost its edge in computing*. Cambridge: MIT Press.

Howcroft, D. & Rubery, J. (2019) 'Bias in, bias out': gender equality and the future of work debate. *Labour and Industry*, 29(2), 213–227. Available from: https://doi.org/10.1080/10301763.2019.1619986

Howcroft, D. & Taylor, P. (2023) Automation and the future of work: a social shaping of technology approach. *New Technology, Work and Employment*, 38, 351–370.

Jacobs, J. (2018) Macho 'brogrammer' culture still nudging women out of tech. *Financial Times*, 9 December, pp. 6–9.

Karimi, F., Wagner, C., Lemmerich, F., Jadidi, M. & Strohmaier, M. (2016) Inferring gender from names on the web: a comparative evaluation of gender detection methods. In: Bordeau, J. (Ed.) *Proceedings of the 25th International Conference Companion on World Wide Web (WWW '16 Companion)*. Montreal, Quebec, Canada: International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, CHE, pp. 53–54. Available from: https://doi.org/10.1145/2872518.2889385

Koning, R., Samila, S. & Ferguson, J.P. (2019) *Female inventors and inventions*. Harvard Business School.

Krivkovich, A., Kutcher, E. & Yee, L. (2016) *Breaking down the gender challenge*. McKinsey. Available from: https://www.mckinsey.com/business-functions/organization/our-insights/breaking-down-the-gender-challenge [Accessed 27th October 2022].

Kuhlman, C., Jackson, L. & Chunara, R. (2020) No computation without representation: avoiding data and algorithm biases through diversity. Proceedings of the 26th ACM SIGKDD international conference on knowedge discovery and data mining, p. 3593. Available from: https://arXiv.org/abs/2002.11836

Leavy, S. (2018) Gender bias in artificial intelligence: the need for diversity and gender theory in machine learning. In: Abraham, E., Di Nitto, E. and Mirandola, R. (Eds.) *Proceedings of the 1st International Workshop on Gender Equality in Software Engineering (GE' 18)*. New York: Association for Computing Machinery, pp. 14–16.

Lerchenmueller, M.J., Sorenson, O. & Jena, A.B. (2019) Gender differences in how scientists present the importance of their research: observational study. *BMJ*, 367, l6573. Available from: https://doi.org/10.1136/bmj.l6573

Leslie, S.-J., Cimpian, A., Meyer, M. & Freeland, E. (2015) Expectations of brilliance underlie gender distributions across academic disciplines. *Science*, 347(6219), 262–265. Available from: https://doi.org/10.1126/science.1261375

Li, L., Jing, H., Tong, H., Yang, J., He, Q. & Chen, B. (2019) *NEMO: Next career move prediction with contextual embedding. Proceedings of the 26th International Conference on World Wide Web companion*. Perth, Australia. pp. 505–513. Available from: https://doi.org/10.1145/3041021.3054200

Loukides, M. (2021) *2021 data/ai salary survey*. O'Reilly Media. Available from: https://www.oreilly.com/radar/2021-data-ai-salary-survey/ [Accessed 3rd November 2022].

Mantha, Y. & Hudson, S. (2018) *Estimating the gender ratio of AI researchers around the world*. Available from: https://medium.com/element-ai-research-lab/estimating-the-gender-ratio-of-ai-researchers-around-the-world-81d2b8dbe9c3 [Accessed 15th September 2022].

Margolis, J. & Fisher, A. (2012) *Unlocking the clubhouse: women in computing*. Cambridge: MIT Press.

Misa, T.J. (2010) Gender codes: defining the problem, *Gender codes: why women are leaving computing*. New Jersey: John Wiley and Sons, pp. 3–23.

Mitchell, M., Baker, D., Moorosi, N., Denton, E., Hutchinson, B., Hanna, A. et al. (2020) Diversity and inclusion metrics in subset selection. *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, 117–123 Available from: https://arxiv.org/abs/2002.03256

Muzio, D. & Tomlinson, J. (2012) Editorial: researching gender, inclusion and diversity in contemporary professions and professional organizations: editorial. *Gender, Work & Organization*, 19(5), 455–466. Available from: https://doi.org/10.1111/j.1468-0432.2012.00608.x

Nielsen, M.W., Bloch, C.W. & Schiebinger, L. (2018) Making gender diversity work for scientific discovery and innovation. *Nature Human Behaviour*, 2, 726–734.

Noland, M., Moran, T. & Kotschwar, B.R. (2016) *Is gender diversity profitable? evidence from a global survey*. Peterson Institute for International Economics Working Paper No: 16-3. https://doi.org/10.2139/ssrn.2729348

O'Neil, C. & Schutt, R. (2013) *Doing data science: straight talk from the frontline*. California, USA: O'Reilly and Associates.

Passi, S. & Sengers, P. (2020) Making data science systems work. *Big Data & Society*, 7(2), 205395172093960.

Perrault, R., Shoham, Y., Brynjolfsson, E., Clark, J., Etchemendy, J., Grosz, B. et al. (2019) *Artificial intelligence index 2019 annual report*. Human-Centered AI InstituteStanford University: 291. Available from: https://hai.stanford.edu/sites/default/files/ai_index_2019_report.pdf

Posner, M. (2017) We can teach women to code, but that just creates another problem. *The Guardian*. Available from: https://www.theguardian.com/technology/2017/mar/14/tech-women-code-workshops-developer-jobs

Raji, I.D., Gebru, T., Mitchell, M., Buolamwini, J., Lee, J. & Denton, E. (2020) Saving face: investigating the ethical concerns of facial recognition auditing, *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society (AIES 2020)*. New York, NY, USA: Association for Computing Machinery, pp. 145–151. Available from: https://doi.org/10.1145/3375627.3375820

Reskin, B.F. & Roos, P.A. (1990) Job queues, gender queues: explaining women's inroads into male occupations. Philadelphia: Temple University Press.

Roca, T. (2019) *Identifying AI talents among LinkedIn members: a machine learning approach. LinkedIn Economic Graph/Zenodo*, pp. 1–16. Available from: https://zenodo.org/record/3240963

Scott, A., Kapor, F.K. & Onovakpuri, U. (2017) *Tech leavers study voluntarily left jobs in tech*. Oakland, CA, USA: Kapor Center for Social Impact, pp. 1–27. Available from: https://www.kaporcenter.org/wp-content/uploads/2017/08/TechLeavers2017.pdf

Scott, L. (2021) The Spark podcast: Linda Scott and the double-X economy. *BBC Sounds*, 20 January. Available at: https://www.bbc.co.uk/sounds/play/m000rdlr [Accessed 18th July 2022].

Segovia-Pérez, M., Castro Núñez, R.B., Santero Sánchez, R. & Laguna Sánchez, P. (2019) Being a woman in an ICT job: an analysis of the gender pay gap and discrimination in Spain. *New Technology, Work and Employment*, 35(1), 20–39. Available from: https://doi.org/10.1111/ntwe.12145

Simonite, T. (2018) AI is the future—but where are the women? *Wired*, 17 August. Available at: https://www.wired.com/story/artificial-intelligence-researchers-gender-imbalance [Accessed 22nd August 2022].

Sloane, M., Moss, E., Awomolo, O. & Forlano, L. (2022) Participation is not a design fix for machine learning. *Equity and Access in Algorithms, Mechanisms, and Optimization*, 1–6 Available from: https://arxiv.org/abs/2007.02423

Slota, S.C., Hoffman, A.S., Ribes, D. & Bowker, G.C. (2020) Prospecting (in) the data sciences. *Big Data & Society*, 7(1), 205395172090684.

Stathoulopoulos, K. & Mateos-Garcia, J. (2019) *Gender diversity in AI research*, Nesta, July.

Terrell, J., Kofink, A., Middleton, J., Rainear, C., Murphy-Hill, E., Parnin, C. et al. (2017) Gender differences and bias in open source: pull request acceptance of women versus men. *PeerJ Computer Science*, 2017(5), 1–30. Available from: https://doi.org/10.7717/peerj-cs.111

The Alan Turing Institute. (2021) *Frequently asked questions*. London: The Alan Turing Institute. Available from: https://www.turing.ac.uk/about-us/frequently-asked-questions [Accessed 10th August 2022].

Thompson, C. (2019) The secret history of women in coding. *New York Times*, pp. 1–22. Available from: https://nyti.ms/2UVFlxP [Accessed 12th October 2022].

UNESCO. (2020) *Artificial intelligence and gender equality: key findings of UNESCO's global dialogue*. Paris: UNESCO Publishing, p. 27.

Wagman, K.B. & Parks, L. (2021) Beyond the command: feminist STS research and critical issues for the design of social machines. *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW1), 101. Available from: https://doi.org/10.1145/3449175

Wajcman, J. (1991) Patriarchy, technology, and conceptions of skill. *Work and Occupations*, 18(1), 29–45. Available from: https://doi.org/10.1177/0730888491018001002

Wajcman, J. (2010) Feminist theories of technology. *Cambridge journal of economics*, 34(1), 143–152. Available from: https://doi.org/10.1093/cje/ben057

Wajcman, J. (2017) Automation: is it really different this time?: review essay: automation. *The British Journal of Sociology*, 68(1), 119–127.

Wajcman, J. & Young, E. (2023) Feminism confronts AI: the gender relations of digitalisation. In: Browne, J., Cave, S., Drage, E. & McInerney, K. (Eds.) *Feminist AI: critical perspectives on algorithms, data and intelligent machines*. Oxford: Oxford University Press. forthcoming.

West, M., Kraut, R. & Chew, H.E. (2019) I'd blush if I could: closing gender divides in digital skills through education. *UNESCO Equals*, 306.

West, S.M., Whittaker, M. & Crawford, K. (2019) *Discriminating systems: gender, race, and power in AI*. https://ainowinstitute.org/discriminatingsystems.pdf [Accessed 27th September 2022].

Witz, A. (1992) *Professions and patriarchy*. London: Routledge.

World Economic Forum. (2018a) *Assessing gender gaps in artificial intelligence: global gender gap report 2018*, p. 6. Available from: https://doi.org/10.4081/jlimnol.2003.139 [Accessed 15th September 2022].

World Economic Forum. (2018b) *Will AI make the gender gap in the workplace harder to close?* Available from: https://www.weforum.org/agenda/2018/12/artificial-intelligence-ai-gender-gap-workplace/ [Accessed 15th September 2022].

World Economic Forum. (2020) *Diversity, equity and inclusion 4.0: a toolkit for leaders to accelerate social progress in the future of work*. Available from: https://www3.weforum.org/docs/WEF_NES_DEI4.0_Toolkit_2020.pdf [Accessed 1st October 2022].

World Economic Forum. (2021) *2021 the global gender gap report*. Available from: https://www.weforum.org/reports/global-gender-gap-report-2021%0Afile:///Users/natalienunes/Documents/AshridgeHULTBusinessSchool/ArticlesHULT/WEF_GGGR_2021.pdf (Accessed 17th October 2022).

Wu, T. (2020) The labour of fun: masculinities and the organisation of labour games in a modern workplace. *New Technology, Work and Employment*, 35(3), 336–356. Available from: https://doi.org/10.1111/ntwe.12180

Wynn, A. & Correll, S. (2017) Gendered perceptions of cultural and skill alignment in technology companies. *Social Sciences*, 6(2), 45. Available from: https://doi.org/10.3390/socsci6020045

Wynn, A.T. & Correll, S.J. (2018) Puncturing the pipeline: do technology companies alienate women in recruiting sessions? *Social Studies of Science*, 48(1), 149–164. Available from: https://doi.org/10.1177/0306312718756766

Young, E., Wajcman, J. & Sprejer, L. (2021) *Where are the women? mapping the gender job gap in AI policy briefing—full report*. London: The Alan Turing Institute.

Zhang, D., Mishra, S., Brynjolfsson, E., Etchemendy, J., Ganguli, D., Grosz, B. et al. (2021) *The AI index 2021 annual report*. Human-Centered AI Institute, Stanford University.

How to cite this article: Young, E., Wajcman, J. & Sprejer, L. (2023) Mind the gender gap: inequalities in the emergent professions of artificial intelligence (AI) and data science. *New Technology, Work and Employment*, 38, 391–414. https://doi.org/10.1111/ntwe.12278