Where are the women?

Mapping the gender job gap in AI

Policy Briefing – Summary

Erin Young, Judy Wajcman, Laila Sprejer

The Alan Turing Institute

Public Policy Programme
Women in Data Science and AI project

Authors

Erin Young is a Postdoctoral Research Fellow in the Public Policy Programme at The Alan Turing Institute. https://www.turing.ac.uk/people/researchers/erin-young

Judy Wajcman is the Principal Investigator of the Women in Data Science and AI project at The Alan Turing Institute, and Anthony Giddens Professor of Sociology at the London School of Economics. http://www.lse.ac.uk/sociology/whoswho/academic/wajcman.aspx

Laila Sprejer is a Data Science Research Assistant in the Public Policy Programme at The Alan Turing Institute. https://www.turing.ac.uk/people/researchers/laila-sprejer

Cite as: Young, E., Wajcman, J. and Sprejer, L. (2021). Where are the Women? Mapping the Gender Job Gap in AI. Policy Briefing: Summary. *The Alan Turing Institute*.

Contents

Introduction

There is a troubling and persistent absence of women employed in the Artificial Intelligence (AI) and Data Science fields. Over three-quarters of professionals in these fields globally are male (78%); less than a quarter are women (22%) (World Economic Forum, 2018). In the UK, this drops to 20% women. This stark male dominance results in a feedback loop shaping gender bias in AI and machine learning systems. It is also fundamentally an ethical issue of social and economic justice, as well as one of value-in-diversity.

Nearly 4 years ago, the House of Lords Select Committee on Artificial Intelligence (2018) advocated for increasing gender and ethnic diversity amongst AI developers, and last year the European Commission (2020: 3) noted that it is 'high time to reflect specifically on the interplay between AI and gender equality'. Yet there is still a striking scarcity of quality, disaggregated, intersectional data which is essential to interrogate and tackle inequities in the AI and data science labour force.[1] Indeed, the Royal Society (2019: 51) has noted that 'a significant barrier to improving diversity is the lack of access to data on diversity statistics'. The recent AI Roadmap (UK AI Council, 2021: 4) strongly recommends 'mak[ing] diversity and inclusion a priority [by] forensically tracking levels of diversity to make data-led decisions about where to invest and ensure that underrepresented groups are given equal opportunity'.

As AI becomes ubiquitous in everyday life, closing the gender gap in the data science and AI workforce matters. The fields are fast-moving, so it is particularly important to comprehensively map how these gaps are manifest across different industries, occupations, and skills. Such work has added urgency since Covid-19 is having a disproportionate impact on women, potentially increasing the gender gap in the technology industry.

This policy briefing from The Alan Turing Institute's Women in Data Science and AI project is a contribution to this endeavour, charting women's participation in data science and AI in the UK and other countries.[2] By presenting a new, curated dataset, analysed through innovative data science methodology, we are able to explore in detail the gendered dynamics of data science

[1] Women are a multifaceted and heterogeneous group, with a plurality of experiences, and gender intersects with multiple aspects of difference and disadvantage (Crenshaw, 1995).

[2] https://www.turing.ac.uk/research/research-projects/women-data-science-and-ai

and AI careers. The project would not have been possible without our partnership with Quotacom,[3] who provided the seed dataset for our research.

Our research findings reveal extensive disparities between women and men in skills, status, pay, seniority, industry, job, attrition and educational background. We therefore call for effective policy responses if society is to reap the benefits of technological advances.

[3] Quotacom is an executive search and consulting firm specialising in data science, advanced analytics and AI.

Key findings

1. **Existing data is sparse**: The existing evidence base about gender diversity in the AI and data science workforce is severely limited. The available data is fragmented, incomplete and inadequate for investigating the career trajectories of women and men in the fields. Where datasets are available, they often rely on commercial data produced through proprietary analyses and methodologies. National labour force statistics lack detailed information about job titles and pay levels within ICT, computing, and technology, which is in particular a major barrier to examining the emerging hierarchy between data science and AI, and other subdomains. These omissions are compounded by a severe lack of intersectional data about the global AI workforce, broken down by age, race, geography, (dis)ability, sexual orientation, socioeconomic status as well as gender. This is particularly concerning since it is those at the intersections of multiple marginalised groups who are at the greatest risk of being discriminated against at work and by resulting AI bias.

2. **Diverging career trajectories**: There is evidence of persistent structural inequality in the data science and AI fields, with the career trajectories of data and AI professionals differentiated by gender. Women are more likely than men to occupy a job associated with less status and pay in the data and AI talent pool, usually within analytics, data preparation and exploration, rather than the more prestigious jobs in engineering and machine learning. This gender skill gap risks stalling innovation and exacerbating gender inequality in economic participation.

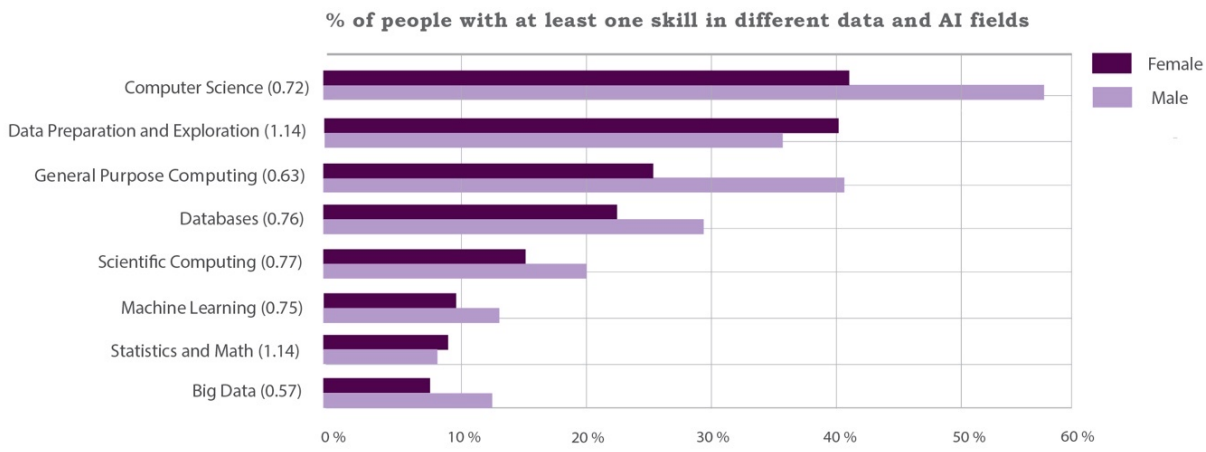

Figure 1: Percentage of people with at least one skill in different data and AI fields. Numbers in brackets represent the gender gap (female/male).

3. **Industry differences**: Women in data and AI are under-represented in industries which traditionally entail more technical skills (for example, the Technology/IT sector), and over-represented in industries which entail fewer technical skills (for example, the Healthcare sector). Furthermore, there are fewer women than men in C-suite positions across most industries, and this is even more marked in data and AI jobs in the technology sector.

4. **Job turnover and attrition rates**: Women working in AI and data science in the tech sector have higher turnover (i.e. changing job roles) and attrition rates (i.e. leaving the industry altogether) than men.

5. **Self-reported skills**: Men routinely self-report having more skills than women on LinkedIn. This is consistent across all industries and countries in our sample. This correlates with existing research into women's lower confidence levels in their own technical abilities.

6. **The qualification gap**: Women in data and AI have higher formal educational levels than men across all industries. The achievement gap is even higher for those in more senior ranks (i.e. for C-suite roles), and this 'over-qualification' aspect is most marked in the Technology/IT sector. This is particularly striking given that Findings 3 and 5 indicate that women are severely under-represented in the C-suite in the technology industry, and that they self-report having fewer data and AI skills.

7. **Participation in online platforms**: Our research indicates that women comprise only about 17% of participants across the online global data science platforms Data Science Central ('DS Central'), Kaggle and OpenML. On StackOverflow, women are a mere 8%. Additionally, we find that only about 20% of UK data and AI researchers on Google Scholar are women. Of the 45 researchers with more than 10,000 citations, only five were women.

Recommendations

1. Reporting standards regarding gender and other workforce characteristics in data science and AI companies urgently need to be developed and implemented. Many of the biggest tech companies provide only headline statistics regarding diversity in their data and AI divisions. Institutions must be more transparent about their workforce and governance diversity. Responsible collection of detailed disaggregated data on women and marginalised groups in these fields must be improved, centrally collated and made available to researchers. This should include data on the proportion, seniority, skills, job tenure, turnover, and remuneration levels of women in the sector, and linked explicitly to issues of bias. The ways in which gender interacts with other sources of inequality such as class, race, ethnicity, religion, disability, age and sexual orientation needs to be a focus of analysis. Governments should apply such reporting requirements to all large tech companies, obliging them to disclose and report on the gender composition of their data science and AI teams.

2. Governments must investigate effective ways to tackle gender data gaps in the AI and data science fields, while maintaining privacy and data protection standards. They should work with national and international organisations to initiate research and advocacy programmes, such as the Inclusive Data Charter (IDC), which promotes more granular data to understand the needs and experiences of the most marginalised in society; the UN Women's Women Count programme, which 'seeks to bring about a radical shift in how gender statistics are used, created and promoted'; and the Data2X project, which aims to improve the 'quality, availability, and use of gender data in order to make a practical difference in the lives of women and girls worldwide'. We recommend working with big technology firms such as LinkedIn that have substantial client databases to begin to build a picture.

3. Countries need to take proactive steps to ensure the inclusion of women and marginalised groups in the design and development of machine learning and AI technologies. For example, the UK government should require companies to scrutinise and disclose the gender composition of their technical, design, management and applied research teams. This must also include mandating responsible gender-sensitive design and implementation of data science research and machine learning. This is an issue of social and economic justice, as well as one of AI ethics and fairness.

4. Given the emerging evidence of biases in AI and discriminatory algorithms, there is an ethical imperative to understand the underlying processes, and to have fair opportunity to challenge

the data, the assumptions, and the metrics employed to mechanise the act of decision-making. We need genuine accountability mechanisms, external to companies and accessible to citizens.

5. Gender inclusive labour market policies, such as paid maternity and parental leave and flexible working hours, must be more effectively implemented and enforced across all industries, and affordable childcare must be provided. These measures are a prerequisite to ensuring that women's disproportionate responsibility for domestic and care work does not inhibit their ability to participate in the digital economy on an equal footing to men. Without them, women will not have equal access to training, re-skilling and job transition pathways, especially in expanding, frontier fields such as data science and AI. This is particularly important given the disproportionate impact of pandemic-related job losses on women.

6. Companies in the tech sector must embed intersectional gender mainstreaming in human resources policy so that women and men are given equal access to well-paid jobs and careers. Actionable incentives, targets and quotas for recruiting, up-skilling, retaining and promoting women at work should be established, as well as ensuring women's equal participation in 'frontier' technical and leadership roles.

Conclusions

Our research, based on a unique dataset of AI professionals, indicates that data science and AI careers in the UK and globally are heavily gendered. There is persistent structural inequality in these fields associated with extensive disparities in skills, status, pay, seniority, industry, attrition rates, educational background, and even self-confidence levels. This gender job gap needs rectifying so that women can fully participate in the AI workforce, including in powerful leadership roles in the design and development of AI.

Our findings are consistent with existing work on the AI gender gap. They require urgent attention given the disproportionate impact of the Covid-19 pandemic on women which risks widening the gender gap in the tech industry (Little, 2020). As Leavy (2018: 16) says: 'advancing women's careers in the area of Artificial Intelligence is not only a right in itself; it is essential to prevent advances in gender equality supported by decades of feminist thought being undone'.

This is not only about issues of economic opportunity and social justice, but also crucially about AI innovation, fairness and ethics. As evidence mounts of gender, race and other social biases embedded in algorithms, there is the risk that AI systems will amplify existing inequities (Wajcman, Young and FitzMaurice, 2020). We cannot even begin to remedy this, let alone take advantage of the huge potential of AI, without first having a data and AI workforce who are representative of the people these systems are meant to serve.

Whilst it is clear that there is a worrying lack of women in the data science and AI fields, there is a scarcity of detailed, intersectional, publicly available demographic information about the data and AI workforce. This is primarily due to the unwillingness of large technology firms to disclose their own diversity data. The lack of transparency has serious implications for Government policymaking around technological advancement and equity, and for labour market policies. It is crucial that we develop a better understanding of the dynamics of the problem. This policy report, in both summary and full form, provides a first step in building a robust evidence base to comprehend the dearth of women working in such fields, and its relationship with biased AI. In our future work, The Alan Turing Institute's Women in Data Science and AI project will build upon this research in order to explore the factors driving the AI gender gap.

Acknowledgements

We would like to thank **Quotacom** for providing the seed dataset for our research. Quotacom is an international executive search and consulting firm with expertise in the digital transformation and data domains. With offices in Europe and the USA, Quotacom is recognised as the leading recruitment specialist within decision science. Quotacom prides itself on diversity and inclusion and has a strong focus on women in technology, with over 60% of the team being female, including a number of members of the senior leadership team. The company hold long-term, strategic partnerships with their clients, ranging from VC backed start-ups, consulting firms and Fortune 500 enterprises, focussing on digital transformation through frontier technologies across Data, Advanced Analytics, AI, Robotics, Machine Learning, Open Source, IOT, Cloud and Blockchain.

References

Crenshaw, K. W. (1995). Mapping the margins: Intersectionality, identity politics, and violence against women of color. In: (Eds.) K. Crenshaw, N. Gotanda, G. Peller and K. Thomas. Critical race theory: The key writings that formed the movement New York: New Press.

Data2X (n.d.). “Important Data about Women and Girls Is Incomplete or Missing.” Retrieved from: https://data2x.org

European Commission (2020). Opinion on Artificial Intelligence – opportunities and challenges for gender equality. *Advisory Committee on Equal Opportunities for Women and Men*. (18 March). Retrieved from: https://ec.europa.eu/info/publications/list-previous-opinions-advisory-committee-equal-opportunities-women-and-men-2014-2016_en

Global Partnership for Sustainable Development Data (n.d.). “Inclusive Data Charter.” Retrieved from: https://www.data4sdgs.org/inclusivedatacharter

House of Lords (2018). AI in the UK: ready, willing and able? Report of Session 2017–19. Select Committee on Artificial Intelligence. HL Paper 100. Authority of the House of Lords.

Leavy, S. (2018). Gender Bias in Artificial Intelligence: The Need for Diversity and Gender Theory in Machine Learning. In GE ’18: Proceedings of the 1st International Workshop on Gender Equality in Software Engineering, 14-16. New York, NY: Association for Computing Machinery.

Little, J. (2020). 'Everything has been pushed back': how Covid-19 is dampening tech's drive for gender parity. The Guardian, 18 June. Retrieved from: https://www.theguardian.com/careers/2020/jun/18/everything-has-been-pushed-back-how-covid-19-is-dampening-techs-drive-for-gender-parity

Quotacom (n.d.). “Quotacom”. Retrieved from: https://www.quotacom.com/

The Alan Turing Institute (n.d.). Women in Data Science and AI Hub. Retrieved from: https://www.turing.ac.uk/about-us/equality-diversity-and-inclusion/women-data-science-and-ai

The Royal Society (2019). Dynamics of data science skills How can all sectors benefit.

UK AI Council. (2021). *AI Roadmap.* Retrieved from: https://www.gov.uk/government/publications/ai-roadmap

UN Women/Women Count (n.d.). "Women Count." Retrieved from: https://data.unwomen.org/women-count

Wajcman, J., Young, E. and FitzMaurice, A. (2020). The Digital Revolution: Implications for Gender Equality and Women's Rights 25 Years after Beijing. Discussion Paper No. 36 (August), UN Women.

World Economic Forum (2018). Global Gender Gap Report 2018. Retrieved from: https://www.weforum.org/reports/the-global-gender-gap-report-2018