Gender Diversity in AI Research

Kostas Stathoulopoulos and Juan Mateos-Garcia

July 2019

nesta

About Nesta

Nesta is an innovation foundation. For us, innovation means turning bold ideas into reality and changing lives for the better.

We use our expertise, skills and funding in areas where there are big challenges facing society.

Nesta is based in the UK and supported by a financial endowment. We work with partners around the globe to bring bold ideas to life to change the world for good.

www.nesta.org.uk

If you'd like this publication in an alternative format such as Braille or large print, please contact us at: **information@nesta.org.uk**

Design: **Green Doe Graphic Design Ltd**

nesta

Gender Diversity in AI Research

Summary

Lack of gender diversity in the Artificial Intelligence (AI) workforce is raising growing concerns, but the evidence base about this problem has until now been based on statistics about the workforce of large technology companies or submissions to a small number of prestigious conferences.

We build on this literature with a large-scale analysis of gender diversity in AI research using publications from arXiv, a widely-used preprints repository where we have identified AI papers through an expanded keyword analysis, and predicted author gender using a name-to-gender inference service. We study the evolution of gender diversity in various disciplines, countries and institutions, finding that while the share of female co-authors in AI papers is increasing, it has stagnated in disciplines related to computer science. We also find that geography plays an important role in determining the share of female authors in AI papers and that there is a severe gender gap in the top research institutions. We also study the link between female authorship in papers and the citations it receives, finding a strong, positive correlation in research domains related to the impact of information technology on society. Having done this, we examine the semantic differences between AI papers with and without female co-authors. Our results suggest that there are significant differences in machine learning and computer ethics between the United States and the United Kingdom as well as differences in the research focus of papers with female co-authors. We conclude by reporting the results of interviews with female AI researchers and other important stakeholders aimed at interpreting our findings and identifying policies to improve diversity and inclusion in the AI research workforce.

1 Introduction

Artificial intelligence (AI) is a general purpose technology that increasingly mediates our social, cultural, economic and political interactions.[1] From improved medical applications to self-driving cars and smart cities, AI has the potential to transform our digital, physical and social environments in unprecedented ways and at an unprecedented speed.[2] However, the same technologies can be used for mass surveillance, computational propaganda and biased, discriminating decision-making.[3, 4] It is generally believed that increasing the diversity of the workforce developing AI systems will reduce the risk that they generate discriminatory and unfair outcomes, thus ensuring that their benefits are more widely shared.

But how diverse is the workforce of the AI sector?

There is mounting evidence of serious gaps in the gender and ethnic diversity of the AI research and industrial workforce. Recently, the AI Index (2018) reported that 80 per cent of AI professors in prestigious US universities were men, while just over a quarter of the students in undergraduate AI classes at Stanford and University of California Berkeley were women.[5] Meanwhile, Element AI found that only 18 per cent of paper authors at 21 leading AI conferences were women.[6]

The situation is similar in industry. AI Index used online job advertisement data and found that 71 per cent of applicants for AI roles in the United States in 2017 were men. The World Economic Forum highlighted in its Global Gender Gap Report (2018) that only 22 per cent of AI professionals on LinkedIn were women with no evidence of improvement in recent years.[7] The report also showed a 'persistent structural gender gap among AI professionals' with career trajectories being differentiated by gender. For example, women were better represented in roles such as data analysis and information management while men tended to fill software engineering and senior level roles.

This lack of gender diversity in AI R&D creates the risk that AI systems 'perpetuate existing forms of structural inequality even when working as intended'.[8] The reason for this is that R&D teams lacking diversity will be insufficiently aware of, or sensitive to, the risks of the technologies that they develop for other social (vulnerable) groups. Avoiding lock-in to discriminatory trajectories of AI deployment is an urgent task, and one that needs to be informed by robust evidence.[9]

The existing evidence base about gender diversity in the AI workforce is, however, not without its limitations: It is mostly based on small samples that although highly relevant (technology industry workforce, papers presented in prestigious conferences) are not necessarily representative of the wider AI research workforce. They also tend to ignore the extent to which the situation of AI is the same, better or worse than in other STEM disciplines, and do not consider variation in the situation between countries that might help to identify practices and policies that could improve the situation. They also tend to assume that increasing gender diversity will directly change the nature of the AI research that is produced in ways that increase the inclusiveness of its benefits and reduces its risks, yet this assumption remains untested. In some cases, it is reliant on commercial data with analyses that are hard to reproduce. As the AI Index report notes, 'a significant barrier to improving diversity is the lack of access to data on diversity statistics in industry and in academia'.

Here, we use a larger dataset from arXiv, an online preprints repository widely adopted by AI researchers, enriched with geographical, discipline and gender information, to address some of these questions, thus improving the evidence base about gender diversity in AI research. Moreover, we conduct a small number of interviews with researchers and university representatives in order to get a qualitative interpretation of our findings, identify promising diversity and inclusion policies in education and academia and inform our future work stream. After describing data collection and processing in Section 2, in Section 3 we present the findings of our analysis of the state and evolution of gender diversity in AI research, its drivers and its links with citations and research content. In Section 4 we report the results of interviews with leading female AI researchers and other important stakeholders that we have identified through our analysis and in Section 5 we concludes by outlining the limitations of our analysis, its implications and issues for further research.

2

Data collection and pre-processing

Our analysis relies on several data collection and processing steps that are described below and can be inspected on GitHub. Table 1 summarises our variables and their sources.

Table 1: Variables

Variable	Source	Description
Title	arXiv	Paper title
Abstract	arXiv	Paper abstract
Citation count	MAG	Paper citations
Year	arXiv	Publication year
Categories	arXiv	arXiv categories
ID	arXiv	Paper ID
Is AI	Own authors	Flag showing if a paper contains AI terms
Communities	Own authors	Clustered disciplines – See Section 2.5
Gender	GenderAPI	Inferred authors gender
Affiliations	MAG	Author affiliations
Country	Google Places API	Country of the affiliations

2.1 arXiv

Arxiv is an online repository providing open access to more than 1.5 million research articles. It contains e-prints on Physics, Mathematics, Computer Science, Quantitative Biology, Quantitative Finance, Statistics, Electrical Engineering and Systems Science, and Economics. ArXiv is widely used by the AI research community to share the findings of their work.[10]

In March 2019, we collected information about all papers in arXiv through its application programming interface.[11] We then removed papers where the abstract was missing, shorter than 300 characters, or indicating that the publication had been withdrawn from arXiv. This left us with 1,372,350 papers which we used in the analysis.

2.1.1 Microsoft Academic Graph (MAG)

Microsoft Academic Graph (MAG)[12] is an academic knowledge base compiled by Microsoft as part of its Cognitive Services that can be accessed programmatically through an API and is increasingly used in scientometric research.[13] It contains more than 140 million academic papers and documents. In order to enrich our arXiv corpus with relevant information from MAG, such as the institutional affiliation of paper authors and their citations, we matched both datasets using the strategy described in Klinger, et al. (2018) [1]. 87 per cent of the arXiv preprints were matched with MAG. We believe that most of the mismatches are due to titles on arXiv being significantly different from those on MAG or MAG not containing the publication.

2.2 Geocoding affiliations

The Google Places API is a commercial cloud service that 'provides names, addresses, and other rich details like ratings, reviews, or contact information for over 150 million places.'[14] Here, we queried the institutional affiliations of the authors in our corpus to determine their location.

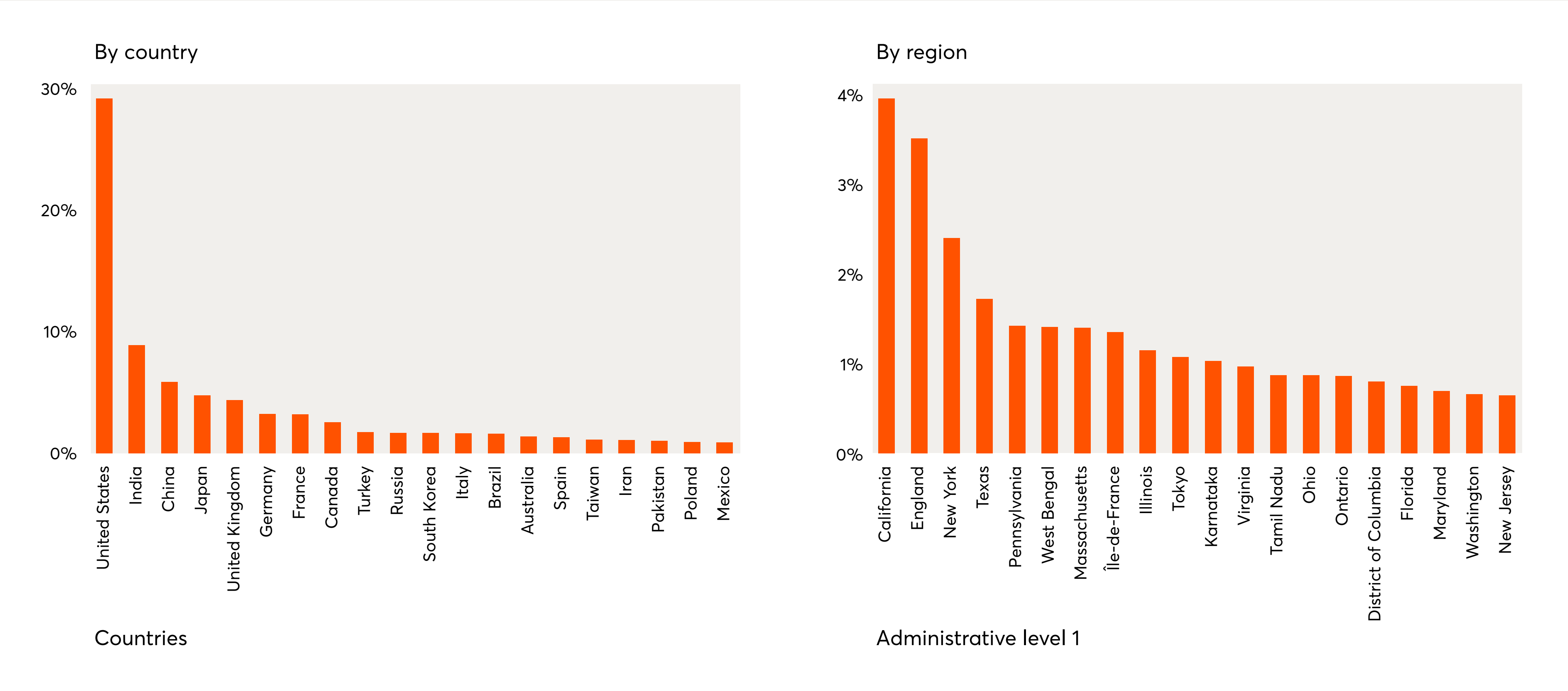
We used three API endpoints for the matching:

- **Place search**: Search for places either by proximity or a text string. The text input can be any kind of location data such as name, address, or phone number. It returns basic information for a given place such as its name, address, longitude and latitude.
- **Place autocomplete**: Provides an autocomplete functionality for text-based geographic searches. It returns place predictions.
- **Place details**: Search for a place using its Place ID.[15] It returns comprehensive information about the queried place such as its complete address, phone number, user rating and reviews.

We queried the affiliations to the **Place search** endpoint and successfully geocoded 88 per cent of them. We assumed that those not matched to any location had a slightly different name to the ones contained in Google Maps. We queried them to the **Place autocomplete** endpoint, selected their most probable match and gathered their Place IDs. Finally, we queried Place IDs to the **Place details** endpoint to geocode the affiliations.

This way, we geocoded 93 per cent of the 8,351 affiliations in our data.

Figure 1: Geocoded affiliations



2.3 Gender classification

In our analysis, we use author names to infer their gender.[16] There are various name-to-gender inference services but we decided to use Gender API, the biggest platform on the internet to determine gender by a first name, a full name or an email address.[17]

Its database contains 1,877,874 validated names from 178 different countries,[18] that are collected from publicly available governmental sources and combined with data crawled from social networks. In addition, each name has to be verified by different sources to be incorporated and the API provides two confidence parameters, number of samples and accuracy. The former shows the number of database records matching the request and the latter determines the reliability of the assignment. A recent comparative study showed that the Gender API exhibits very high accuracy (92.1 per cent) and classifies 97 per cent of the queried names.[19]

We infer the gender from author names in our corpus using the following approach:

- Query the Gender API with full names. The last name is used to improve results on gender-neutral names. Every full name was provided as a text string, was pre-processed by the API and used in inference.
- 2.3.1 Exclude results where the first name field contained only an initial
- 2.3.2 Remove results with less than 80 per cent accuracy
- 2.3.3 Remove any papers where less than 50 per cent of the authors had gender information

Following this procedure, we labelled ~480K of the ~772K author names in arXiv.

It should be mentioned that as with all other inference systems, Gender API has limitations. It may underestimate the number of female names[20] and its performance degrades with Asian and especially South-East Asian names.[21] Lastly, inferred genderisation assumes that gender identity is both a fixed and binary concept. We acknowledge that this limitation restricts the scope of our analysis to binary genders.

2.4 AI labelling

There are many potential approaches to identify papers related to AI in our corpus. Some options include using specific arXiv categories such as cs.AI or cs.NE (respectively referring to AI and neural networks), using an expert-curated list of keywords,[22] or topic modelling approaches.[23] Here, we decided to identify papers related to AI by developing an information retrieval system that uses a query expansion method based on word embeddings, a machine learning technique that projects words into a vector space where it is possible to measure similarities between them. This makes it possible to expand an initial seed term in the query to also include synonyms and related terms, thus improving the comprehensiveness of the vocabulary used in the query and the recall of results.[24]

Our decision to use this approach was motivated by our interest in identifying applications of AI in research fields outside of computer science and by our interest in AI research applications beyond deep learning (the specific subfield of AI that was identified using topic modelling in[25]), while ensuring that our results were robust to changes in the composition of our initial keyword list.

We implemented our approach in the following way: first, we lowercased, tokenised and removed stop words, punctuation and numeric characters from all of the published abstracts. We also created bigrams and trigrams. Then, we applied two models to the data:

- 2.4.1 Word2Vec with the Continuous Bag-of-Words (CBOW) architecture[26]
- 2.4.2 Term frequency, Inverse document frequency (TF-IDF)

To search for AI publications, we started with an initial list of keywords, namely Artificial Intelligence, Machine Learning, Deep Learning and Data Science, and used the trained Word2Vec to find semantically similar tokens. We retrieved the 250 most similar tokens of each keyword, repeated the process and collected the 50 most similar terms of each token on the expanded query list. Lastly, we removed tokens with an IDF weight lower than the 5th percentile or higher than the 95th percentile of the IDF frequency distribution.

Figure 2: Number of publications of AI papers in arXiv

Through the query expansion, we identified 2,250 AI related keywords. Then, we searched for them in the processed publication abstracts and labelled as 'AI' those that contained at least one of the keywords. We identified 74,407 AI papers in arXiv.

We evaluated our approach in multiple ways. We measured its precision and recall. For the former, we randomly sampled papers labelled as AI and manually investigated them for mismatches. We report a precision of 96 per cent. For the latter, we focused on the cs.LG topic which contains the Machine Learning papers in Computer Science, which is assumed to contain only AI publications and we report a recall of 75.24 per cent.[27]

We also evaluated our results qualitatively. As Figure 2 shows, we find most of the AI papers in the arXiv categories with relevant subjects such as Machine Learning, Computer Vision, Artificial Intelligence and Computation and Language. Lastly, we show that the publication of AI papers has been increasing dramatically from 2011, which is consistent with our findings in.[28]

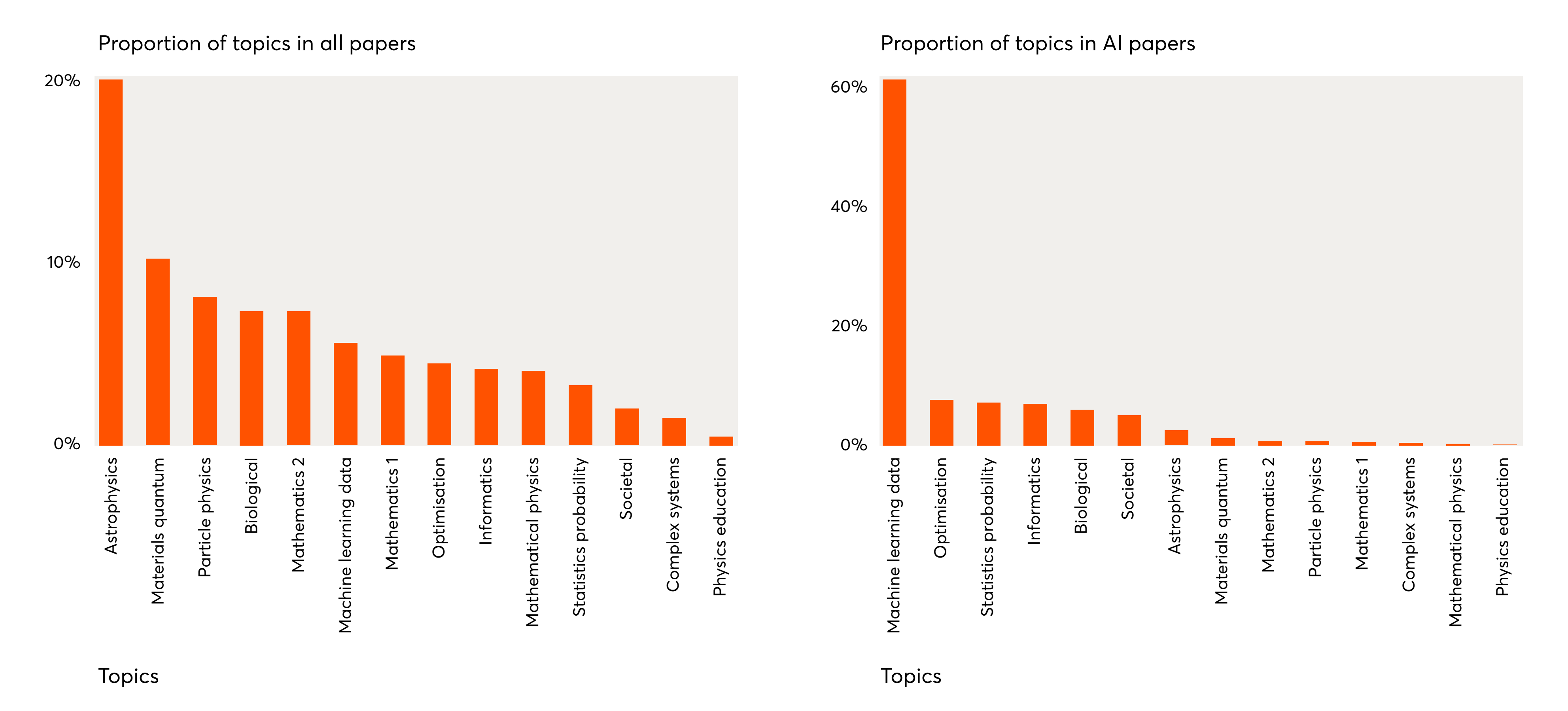
2.5 Discipline clustering

As mentioned in the introduction, we are interested in understanding differences in gender diversity in AI research across research disciplines. The reason for this is that different disciplines could display variation in their research culture and levels of inclusion, thus encouraging or discouraging female participation to different degrees. It might also be the case that disciplines 'feeding' talent into industries could experience different levels of gender diversity, perhaps because those industries are perceived to offer fewer opportunities for women.[29] In order to explore these questions, we need a way to classify papers into disciplines.

Since the arXiv taxonomy includes 175 categories, which is too finely grained and potentially noisy for reporting, we have clustered them into broader 'research domains' by creating a co-occurrence network of the categories used in the AI subset of the data where the edge weight between two nodes shows their Jaccard similarity (roughly, the extent to which they occur together to a greater degree than if they were co-occurring randomly). We then apply the Louvain method for community detection to extract clusters from this category network. Overall, this leads us to identify 15 'research domains' in the data which we use to tag the papers in our corpus (here we note that a paper can be tagged with more than one discipline community).

Lastly, as Figure 3 shows, the distribution of research domains in all arXiv and AI papers differs. We find that 61 per cent of the AI papers fall within the Machine_Learning_Data domain while each of the Optimisation, Statistics_Probability and Informatics domains are found in approximately 7 per cent of the papers.

Figure 3: Proportion of research domains in all arXiv (left) and AI papers (right)



3

Analysis

Having described how we collected and processed our data, here we present our findings focusing on complete years.[30]

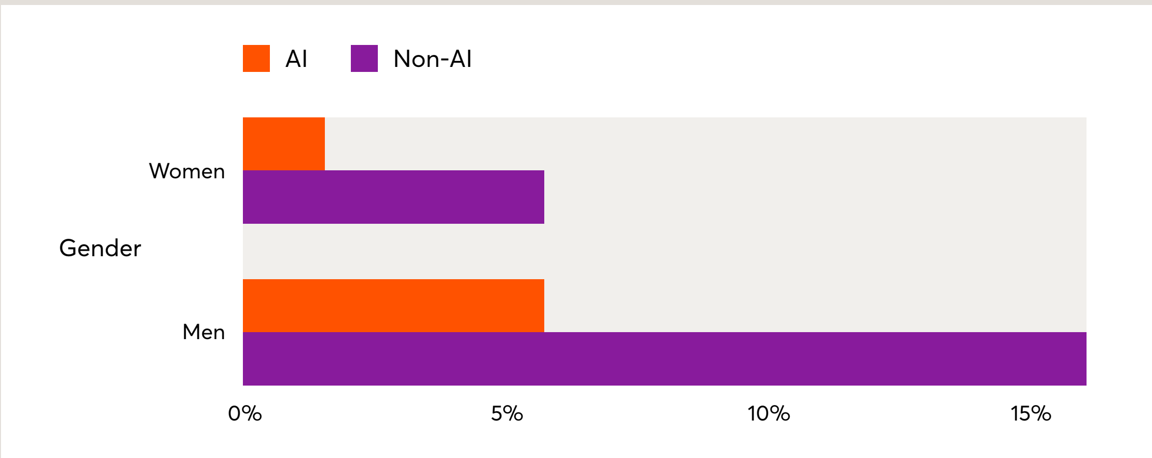
3.1 Descriptive analysis

3.1.1 The state of gender diversity

Our findings confirm that there is a severe gender diversity gap in AI research, with only 13.83 per cent of authors in arXiv being women.[31] This is consistent with the results reported in West et al. (2019),[32] who note that the diversity issues in AI are systemic, with women being underrepresented in most fields related to Computer Science. When examining the non-AI papers in arXiv, we find that 15.51 per cent of the authors with inferred gender are women. Despite the low number of women in AI, we report that 25.4 per cent of the AI publications have been co-authored by a woman, while only 21.04 per cent of the non-AI arXiv papers has a female co-author.

We have also examined gender diversity in single-author papers and find that only 6.72 per cent of the AI publications and 7.3 per cent of the non-AI papers were written by women. Moreover, when looking at the female single-authorship as a proportion of all AI papers with a female author, we find that women are less likely to to single-author a paper in comparison to men.[33] We find a statistically significant difference with the proportion of male single-author AI papers. We show this difference in Figure 4.

Figure 4: Proportion of AI and non-AI single-author papers written by women and men



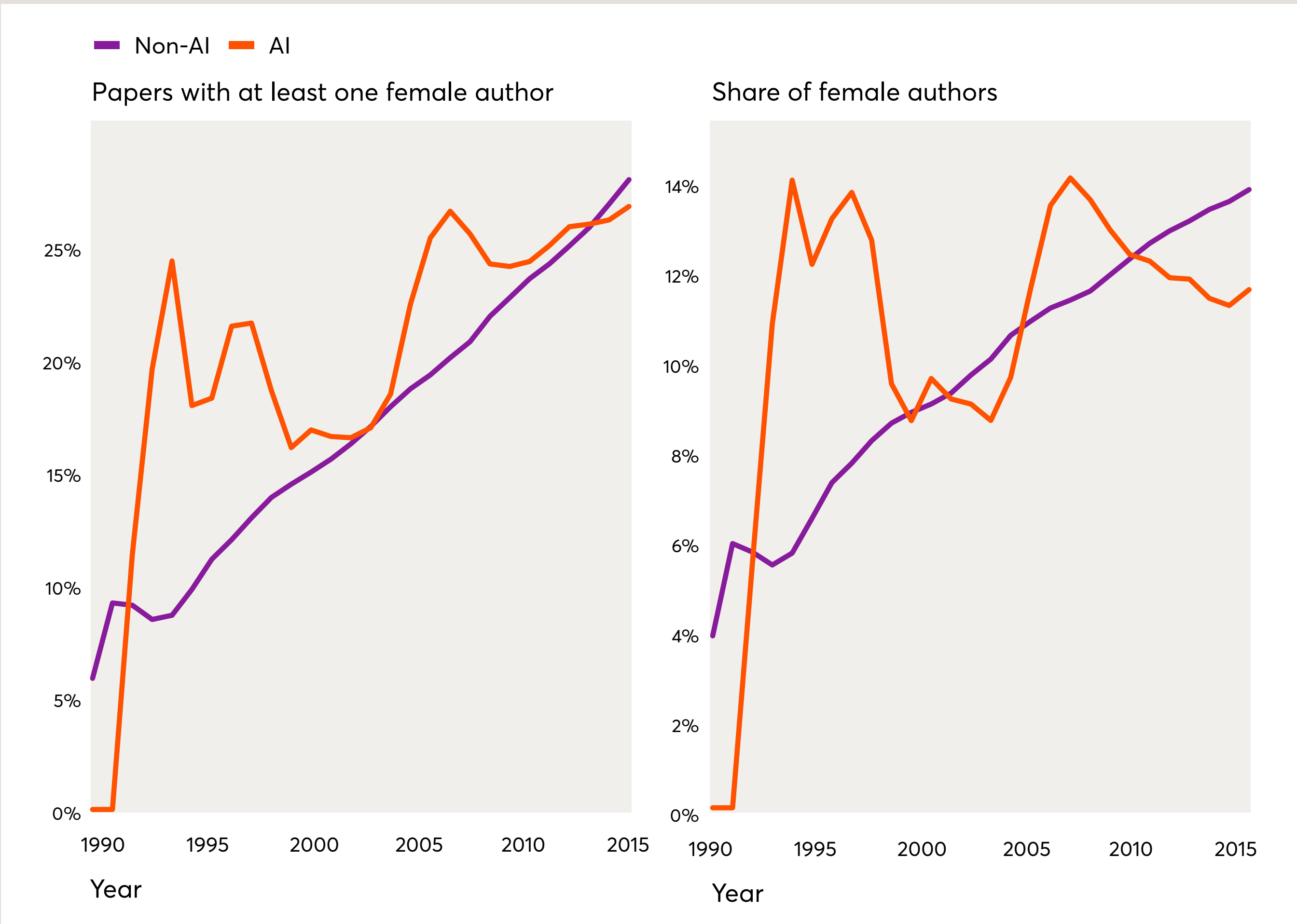
3.1.2 Trends

Here, we focus on how gender diversity has evolved over time and how it changes when looking at particular research domains and geographies.

As Figure 5 shows, the proportion of AI papers co-authored by at least one woman has been increasing from 2004. However, in recent times this growth appears to have stagnated. Looking further back, we see that gender diversity today is not much better than in the 1990s (although it is worth noting that our statistics for the 1990s are based on small sample sizes).

When looking at the share of AI female researchers in the total number of AI researchers, we find stagnation and even decline after some growth between 2005 and 2009. This contrasts with the overall trend in non-AI publications on arXiv where we see a steady increase in the share of female authors. Lastly, it should be mentioned that these results hold when examining the proportion of unique female authors publishing AI research.

Figure 5: Female authorship in AI and non-AI arXiv preprints



The aggregate statistics above mask significant differences between research domains.[34] As shown in Figure 6, we find that the proportion of papers in Machine Learning, Robotics and other data related topics with at least one female author has remained stable, around 25 per cent, throughout the time frame of our analysis. This also holds true for Informatics where approximately 20 per cent of the papers have a female author. On the contrary, in other quantitative disciplines that are not closely related to Computer Science, the share of papers with female authors has been steadily increasing. For example, approximately 40 per cent of the AI publications of 2018 in Astrophysics, 35 per cent in Biology and 28 per cent in Statistics were co-authored by a woman. Lastly, roughly the same trends are observed when examining the number of unique female authors in AI research.

3.1.3 Geographic differences

Having analysed differences between domains in gender diversity, we move on to consider national differences. Here, we use author affiliations at the date of publication as a proxy of their location and focus on countries with at least 5,000 publications and more than 50 per cent of the authors gender-labelled with a high degree of confidence (this unfortunately means excluding China, one of the world leaders in AI research, from the analysis).

Our analysis shows that there are important international differences in the gender diversity gap in AI research. More specifically, we find that 30 per cent of the AI papers in the Netherlands had at least one female co-author. By contrast, only 10 per cent and 16 per cent of those with Japanese and Singaporean affiliations had a female co-author.

We also find differences between the share of female authors in AI and non-AI papers within countries. Most countries of Figure 7 (left) have a higher share of AI papers with female authors, however, this is not observed in Figure 7 (right) where we show the proportion of unique female authors.[35] Nevertheless, there are countries such as Malaysia, Denmark, Norway and Israel that show a stronger presence of women in AI research than outside, according to both variables.

Figure 6: Share of papers with at least one female co-author (split by research domain)

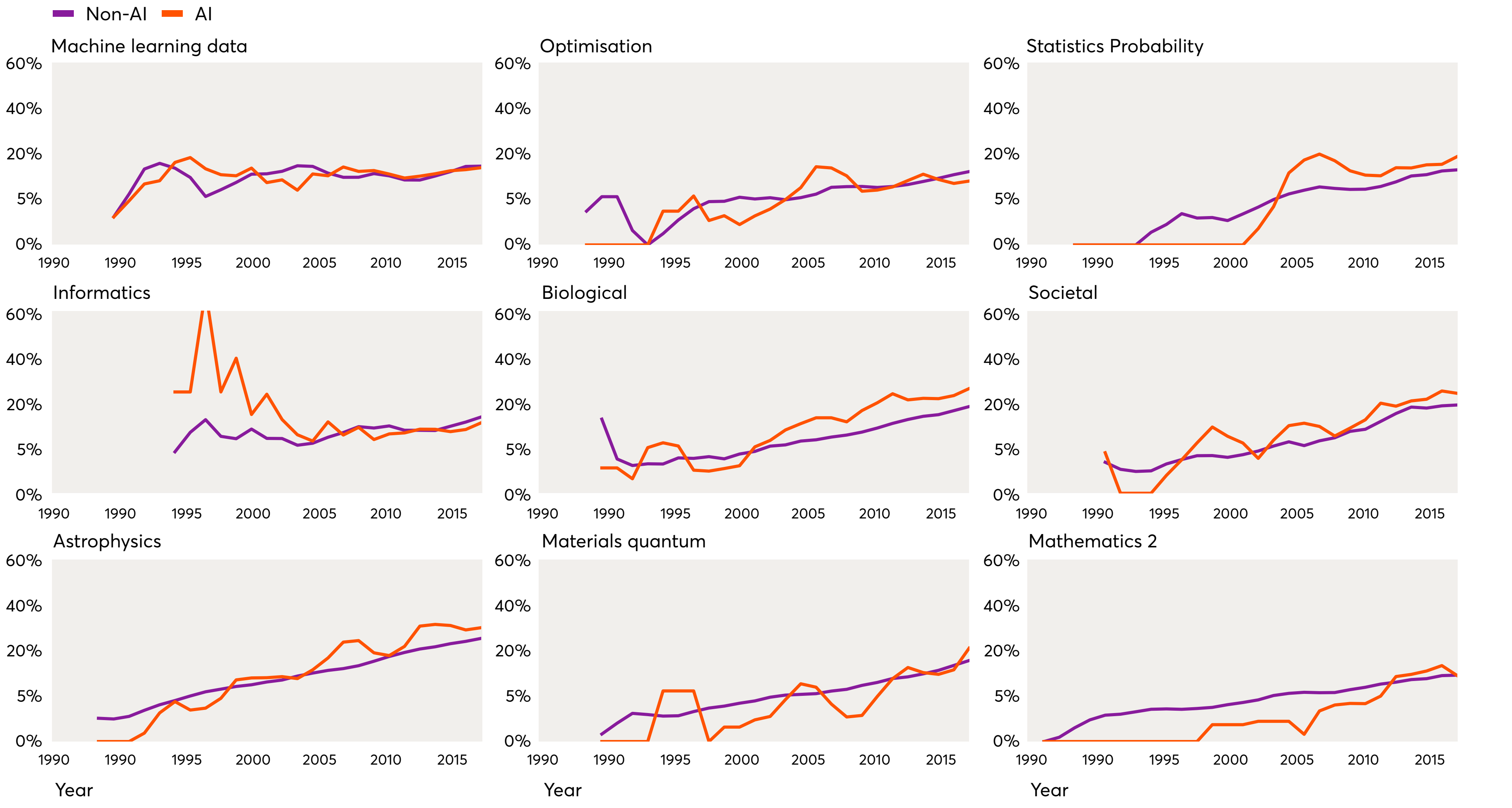


Figure 7: Share of papers with at least one female author (left). Unique female authors in AI and non-AI research (right). China is excluded from the analysis

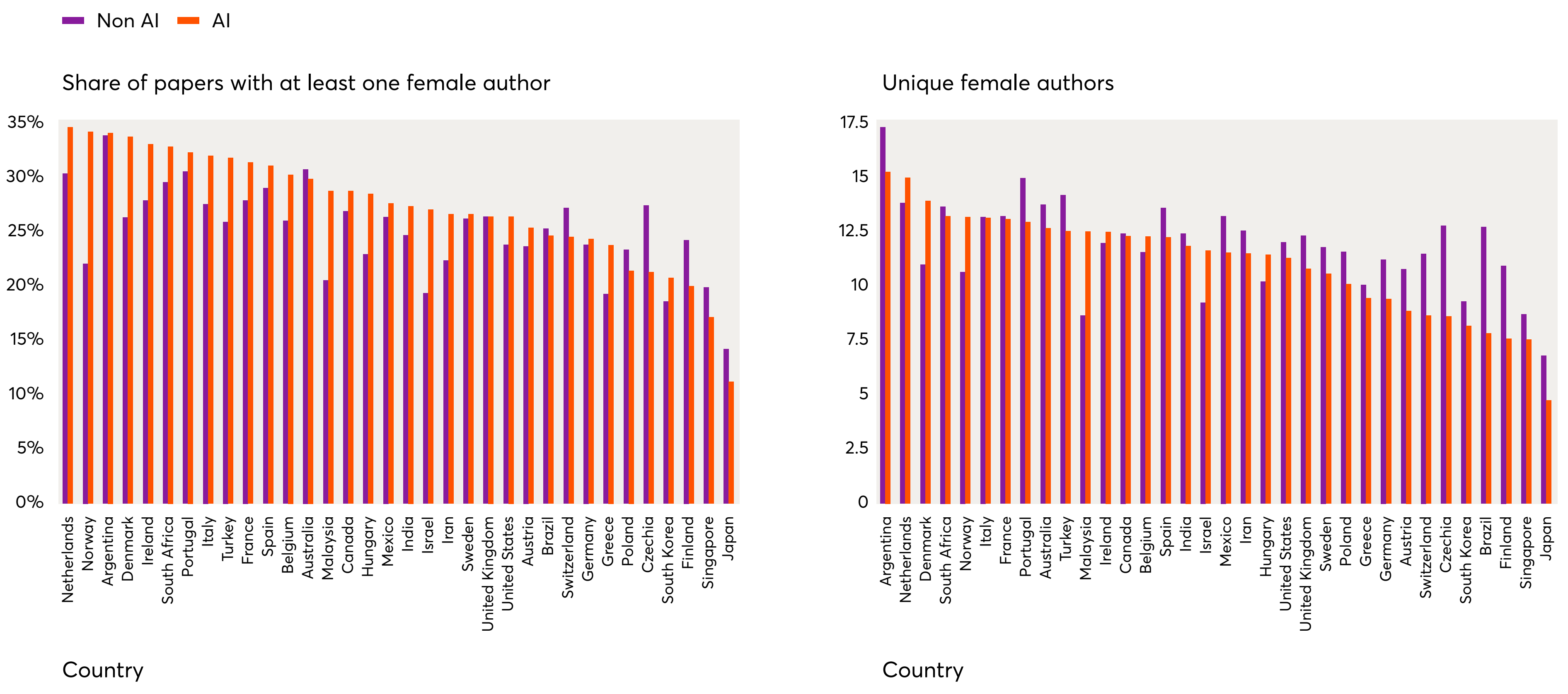
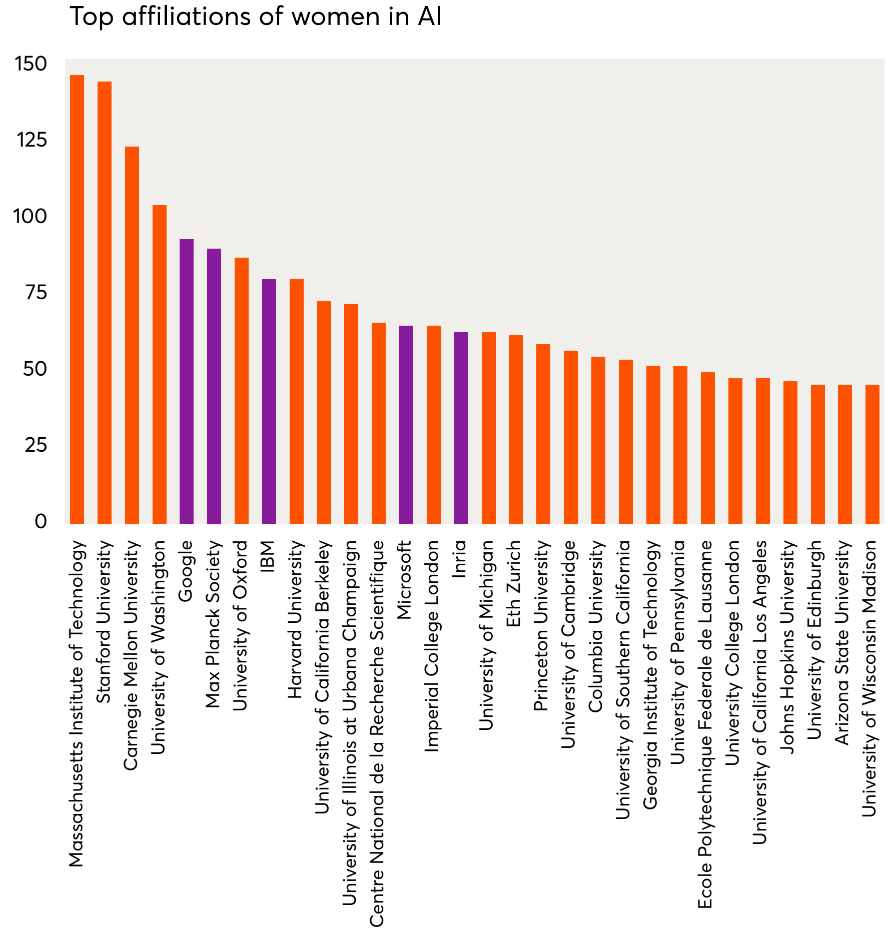
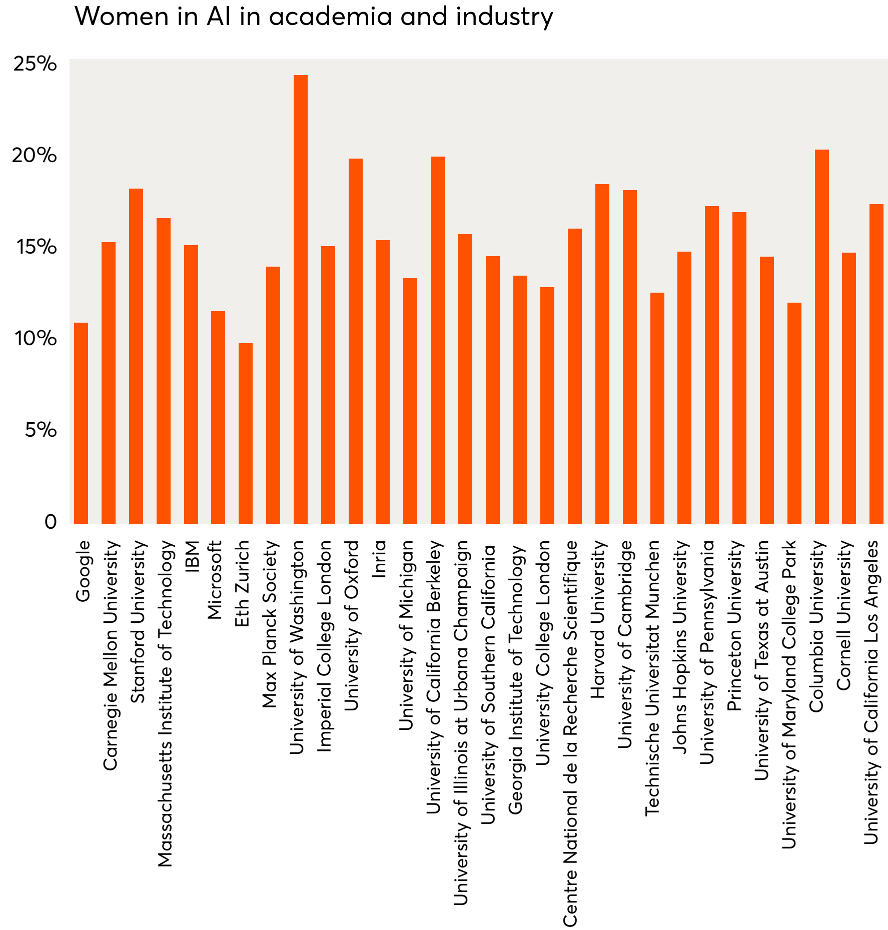


Figure 8: Top affiliations for women in AI (left). Proportion of women in AI in the top research institutions and companies, ordered by the number of publications they have on arXiv (right)





3.1.4 Affiliation differences

We also examined the affiliations of the women in AI research. We find that 79 per cent of the women are affiliated with a university while the proportion decreases to 77 per cent for men. As Figure 8 shows, only six non-academic institutions are in the top 30 affiliations of female authors in AI, while our findings suggest an important gender diversity gap in well-known companies and universities (Figure 8).

For example, only 11.3 per cent of Google's employees who have published their AI research on arXiv are women, while the proportion is similar for Microsoft (11.95 per cent) and IBM (15.66 per cent). When looking at the universities, the ETH Zurich with 10.15 per cent has the lowest share of women authors in AI research in arXiv. It is striking that with the exception of the University of Washington, the share of female AI researchers in the academic institutions and organisations of Figure 8 is never above 25 per cent.

3.2 Drivers of gender diversity

Having studied the evolution of gender diversity in AI research, here we consider its drivers. We are in particular interested in determining whether the disciplinary and geographical differences that we outlined before are significant, and what are their differential contributions to the likelihood that a paper will have a female co-author, taking into account differences between countries in the disciplinary composition of AI research.

First, we have performed z-tests of whether the proportion of women in AI papers is significantly different from the proportion of women in all papers by country and research domain. Figure 9 shows that the share of women in AI is significantly higher than outside in countries such as Netherlands and Norway, while it is lower in Asian and Eastern European countries. When looking at research domains, we find a higher proportion of women working on AI in Physics Education, Astrophysics, Biology and Societal, while the opposite is found for Mathematics and Complex Systems.

In order to understand the association between a factor (countries and research domains) while controlling for con- founders, we estimated a logit model where we regress whether a paper has at least one female author with country and research domain dummies and years, as well as an interaction between whether a paper has been classified as AI and those variables. We present the estimated coefficients and standard errors in Figure 10.

Our analysis shows that, other things being equal, women working in countries such as Ireland, Norway, Malaysia or Netherlands, or in particular domains (Physics and Education and Societal) have a higher probability of publishing work related to Artificial Intelligence. We note that AI papers in computer science research domains such as Machine Learning and Data and Informatics have a significantly lower probability of containing at least one female author after controlling for other factors, consistent with the idea that computer science fields face particularly strong issues with gender diversity in AI research.

Figure 9: Relative representation of female authors in AI. The y-axis shows if women in AI are over-represented (positive values) in a country or domain. Colour shows if the finding is statistically important (orange) or not (green).

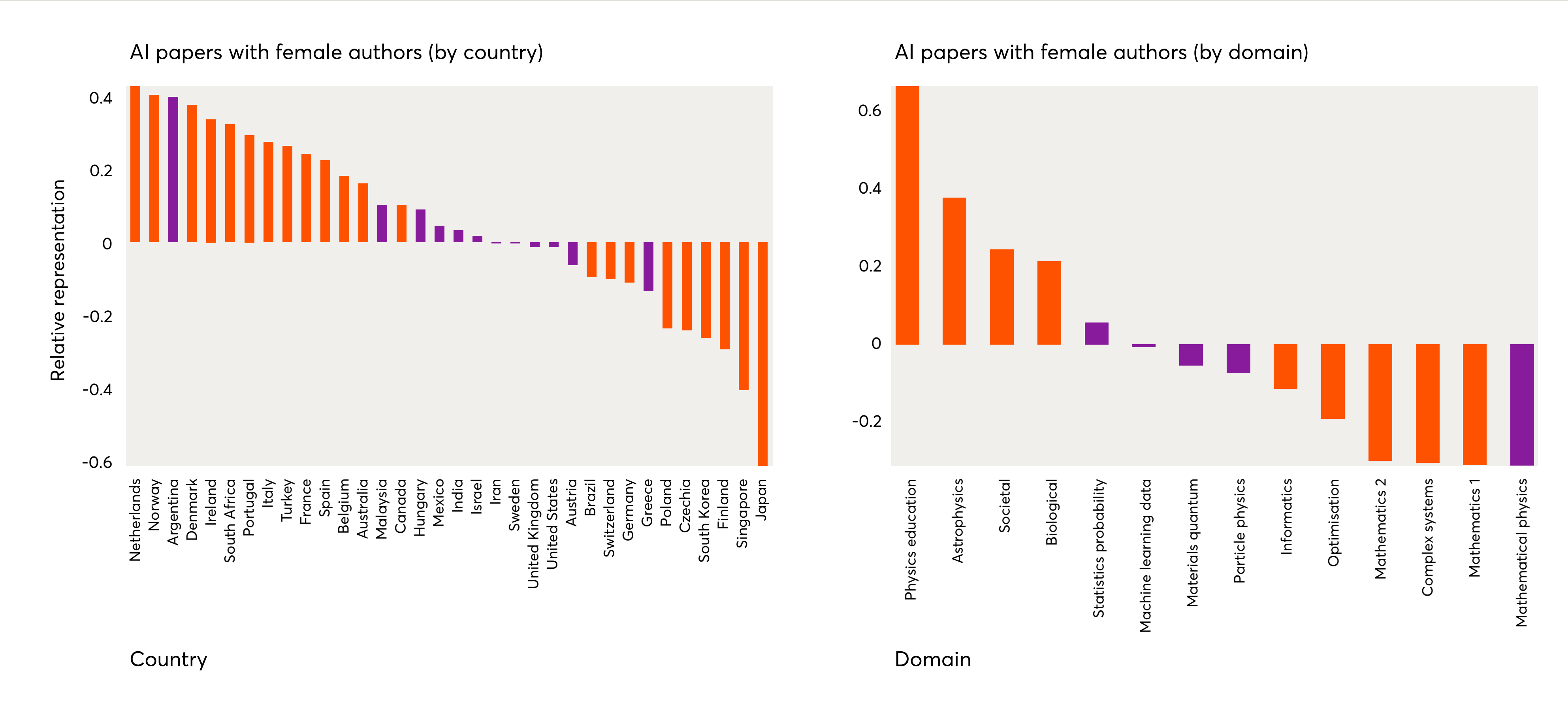


Figure 10: Predicting the presence of female authors in AI publications.
The black lines show the standard deviation of the features

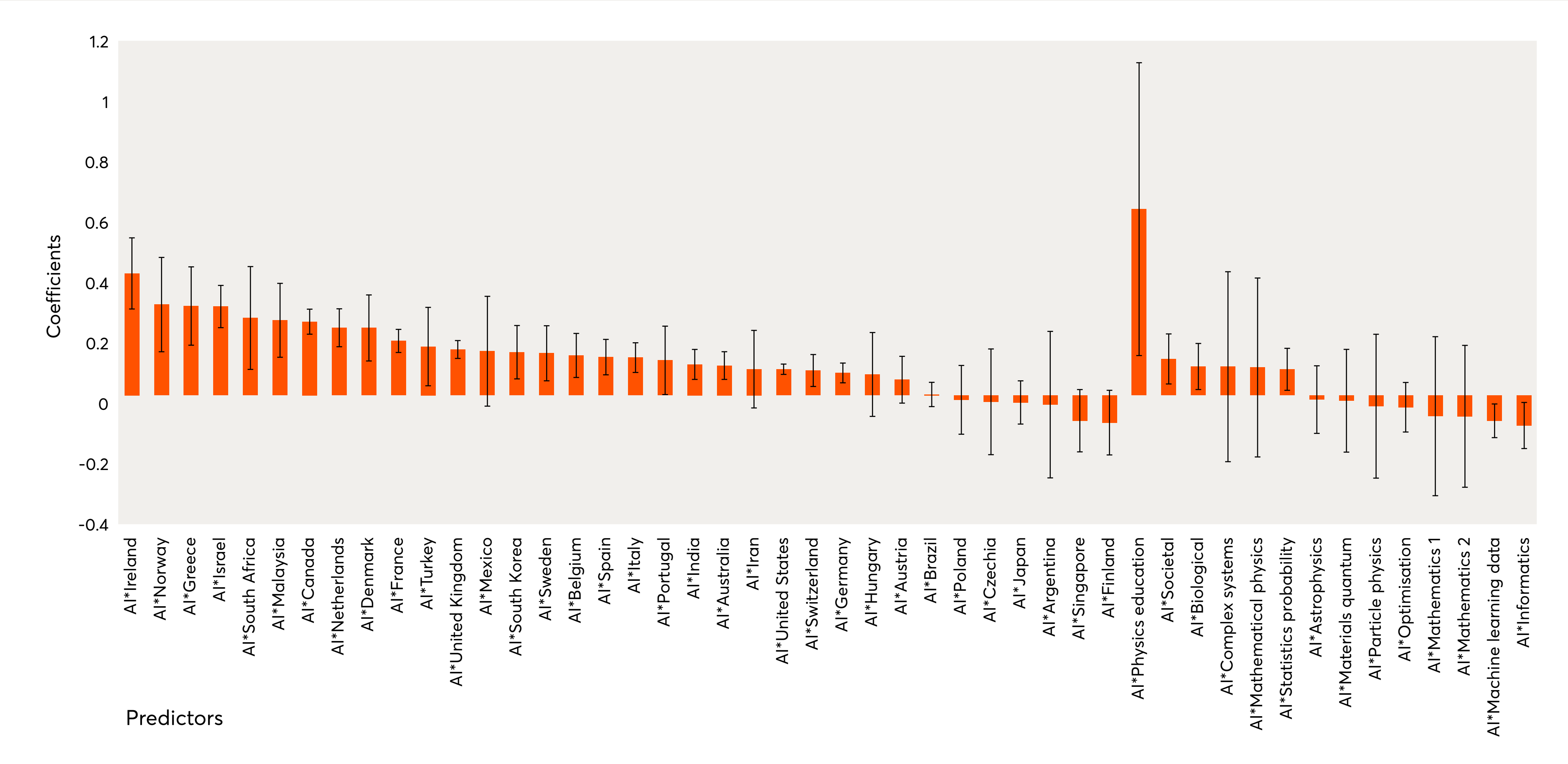
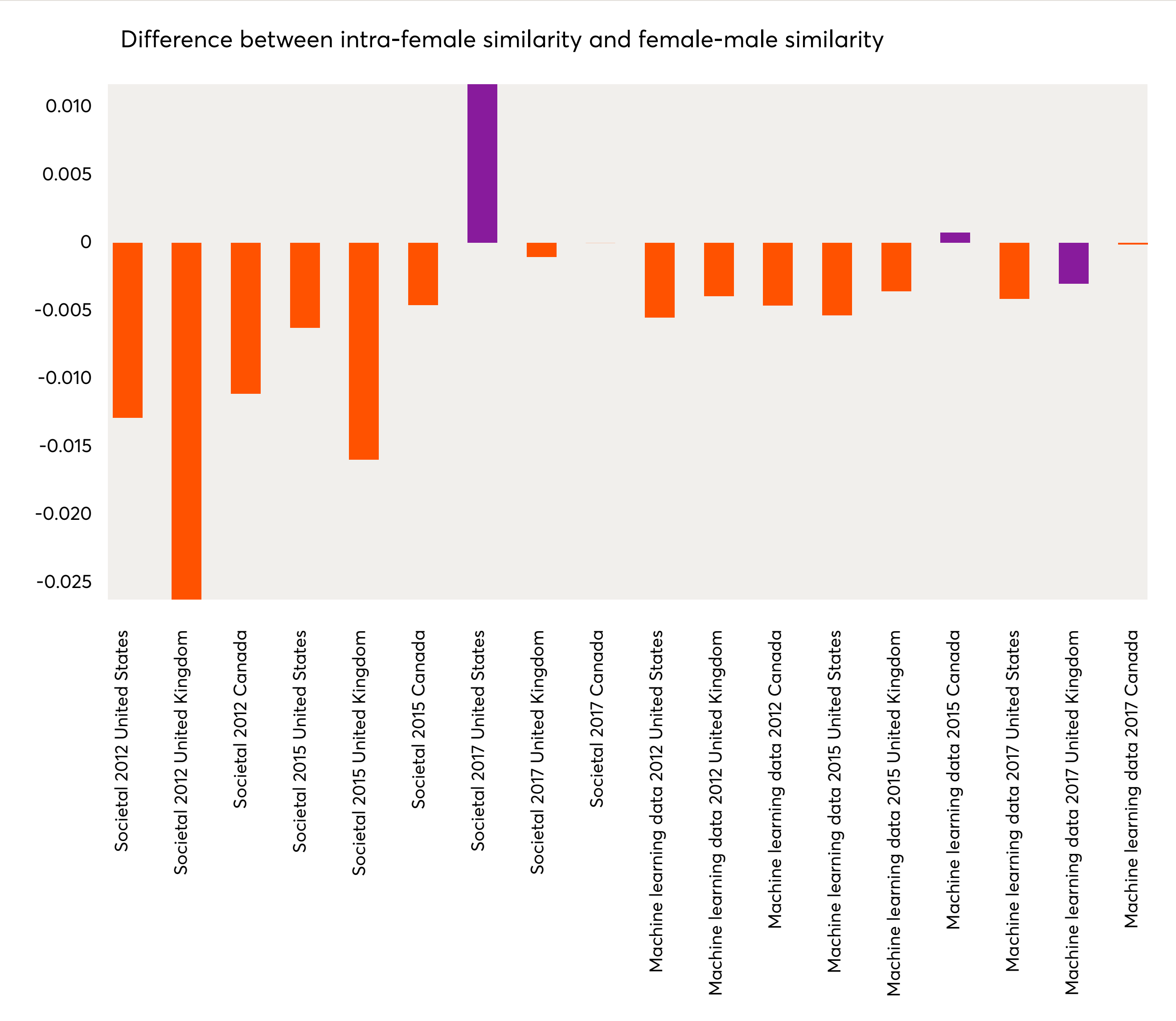


Figure 11: Importance of semantic differences between AI papers co-authored by at least one woman and male-only publications. Colour shows if the finding is statistically significant (orange) or not (blue)



3.3 Effects of gender diversity

3.3.1 Semantic differences

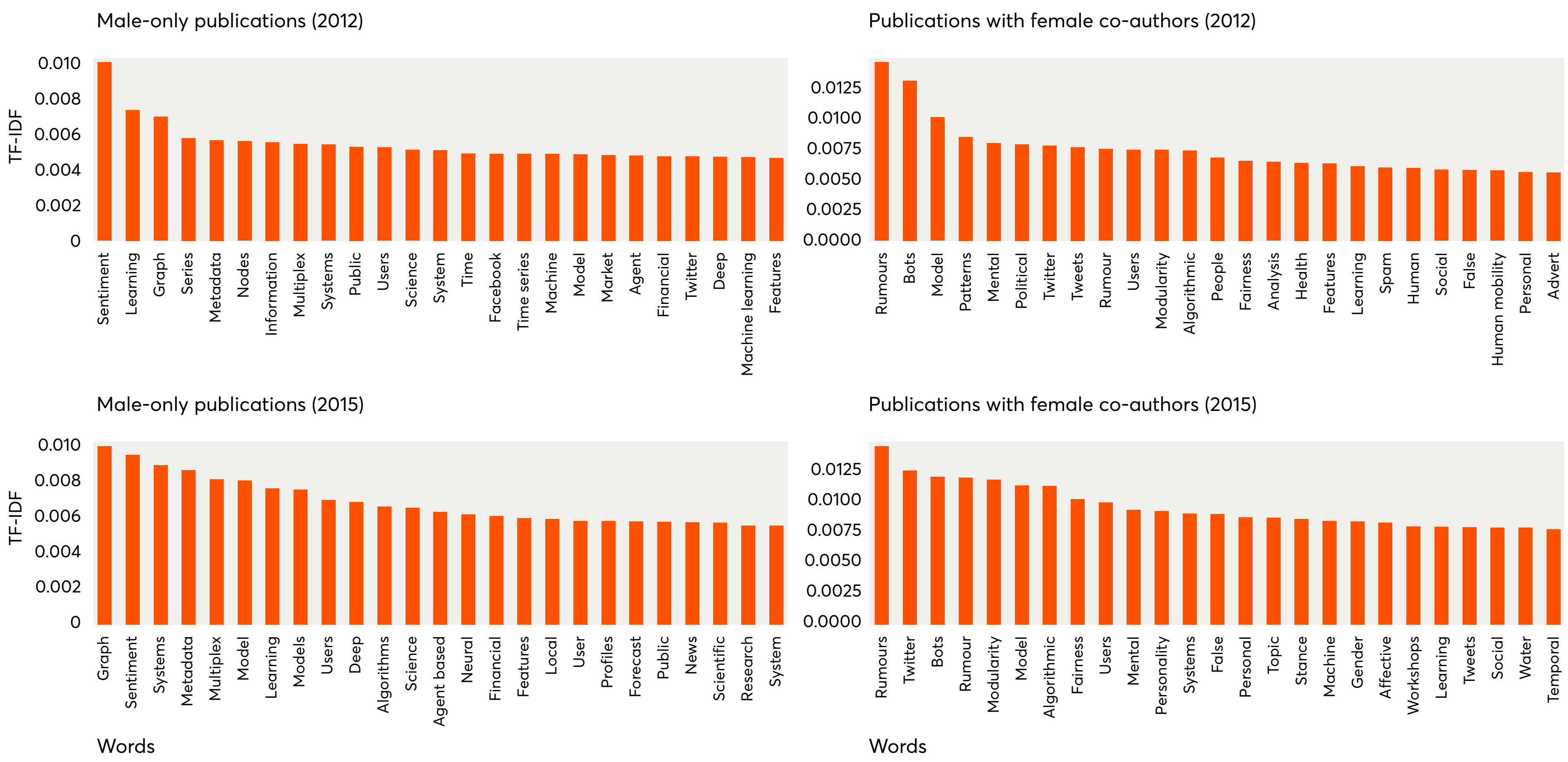
To conclude, we report the findings of an experimental analysis of semantic differences between AI papers involving at least one female co-author and those without any women. Our goal here is to explore whether papers involving women tend to focus on different issues, consistent with the idea that gender diversity might lead to the consideration of a wider set of perspectives and concerns, making the kind of AI research that is undertaken, and the systems that are developed, more diverse and inclusive.

To do this work, we used the Word2Vec and the TF-IDF models of the Section 2.4. We weighted the word vectors of every abstract by their TF-IDF value and averaged the word vectors to create document vectors. Then, we created a matrix with the cosine distance of the document vectors and split it into two parts, those co-authored by at least one female researcher and the rest. Lastly, for every domain, year and country, we ran a t test to evaluate if the mean differences in similarities between both groups are significant (that is, if semantic differences between papers with at least one female author and papers with no female authors are significantly higher/lower than semantic differences inside the group of papers with at least one female author). We perform the analysis inside country and domain cells to control for differences in language that might be brought about by those factors.

Our results suggest that there are statistically significant semantic differences between AI papers with at least one female author and male-only publications when looking at the Societal and Machine Learning and Data topics in Canada, United States and United Kingdom in 2012 and 2015. In general, papers involving at least one woman tend to be more semantically similar to each other than to papers without any female authors.

We further investigated our findings by comparing the words with the highest TF-IDF weight across the corpus for the subsets shown in Figure 11. For example, in Figure 12 we compare the Societal category in the United Kingdom in 2012 and 2015. We show that the 25 most salient terms of papers co-authored by women are more applied and socially aware, with terms such as fairness, human mobility, mental, health, gender and personality being among the top ones.

Figure 12: Comparison of the most insightful terms used in AI papers with and without a female co-author. We show the terms with the highest TF-IDF weight for the Societal category in the United Kingdom in 2012 and 2015.



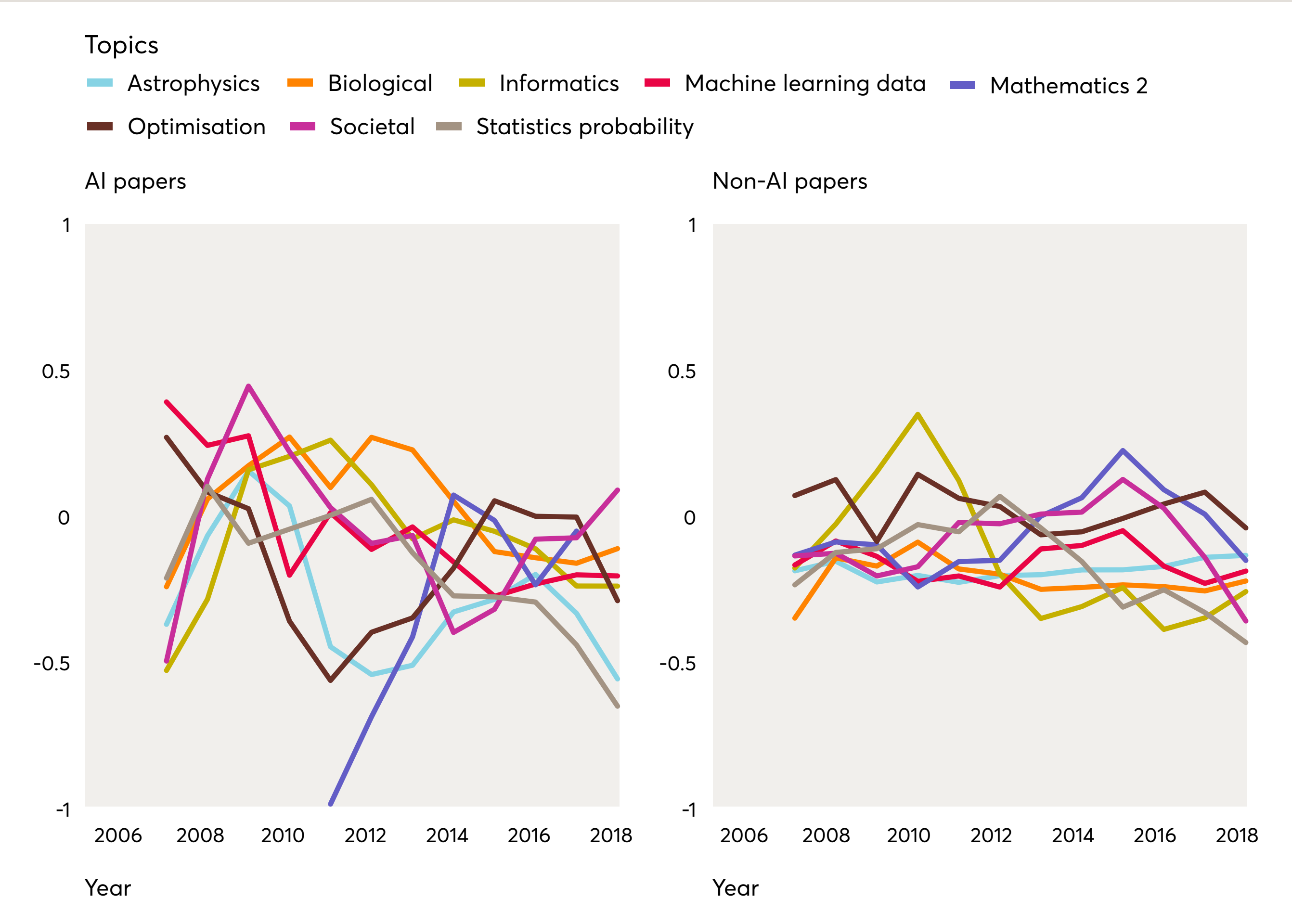
3.3.2 Citations

Finally, we examined how the proportion of women in AI and non-AI publications is related to their number of citations. In detail, we split our analysis by country and topic and removed papers with fewer citations than the median (seven citations). We measured the average number of citations of papers with a different share of female co-authors. Then, we calculated the Pearson correlation between the female author share and average number of citations and tested the statistical significance of our results.

We find a strong and statistically significant correlation of 0.62 between the share of female co-authors and the number of citation in AI papers in the Societal research domain. When examining the results for the countries with the most AI papers on arXiv, we find statistically significant correlations in the United States (0.64) in Societal; in Germany (0.64) and United Kingdom (-0.7) in Optimisation; and in Japan (0.99) in Statistics_Probability.

Lastly, when looking at year-to-year differences, we find that the proportion of female co-authors and the number of ci- tations are not significantly correlated with the exception of Astrophysics (2008, 2018), Societal (2017), Mathematics_2 (2005, 2009, 2010, 2011, 2017) and Biology (2008) in AI papers. This suggests that in general, papers involving female authors are not of lower quality than those involving only male authors.

Figure 13: Pearson correlation: Proportion of female authors in AI and non-AI papers.



4

Interview results

In this section, we complement the quantitative findings that we have reported so far with the results of a small number of interviews conducted with leading women AI researchers and institutions with comparatively high representation of women in their AI workforce, as well as key stakeholders within the UK AI landscape. This way, we seek to contextualise and interpret our findings, consider how they resonate with the personal experience of people in the field and consider their policy implications.[36]

Mihaela van der Schaar is the female AI researcher based in the UK with the most publications in our data. She is the John Humphrey Plummer Professor of Machine Learning, Artificial Intelligence and Medicine at the University of Cambridge and a Turing Fellow at The Alan Turing Institute in London, where she leads the effort on data science and machine learning for personalised medicine. In the interview, she mentioned that even though she began working in AI 16 years ago, her presence in the field and much of her early work has only been recognised recently. She noted that the *"Disparity of recognition between men and women is slowly changing but there is a lot more that needs to be done."* She also highlighted the importance of an open discussion about the challenges that women face in the AI sector and that workplace changes, such as flexible hours are needed to enable researchers to participate in a fast-paced sector without sacrificing their family life. She also highlighted the importance of having a good mentor, people to champion gender diversity in the AI sector as well as individuals at the top of institutions to push forward policies and drive change. Finally, she pointed out that *"Publicising the interdisciplinary scope of possibilities and career paths that studying AI can lead to will help to inspire a more diverse group of people to pursue it. In parallel, the industry will benefit from a pipeline of people who are motivated by combining a variety of ideas and applying them across domains."*

Petia Radeva is another female AI researcher among those with the highest number of publications on arXiv. She is a Professor at the Department of Mathematics and Computer Science at the University of Barcelona, Icrea Academia and she is also the University's Head of Computer Vision and Machine Learning research group as well as Senior Researcher at Computer Vision Center (CVC). She noted that lack of diversity in AI research requires policy interventions to tackle the issue at its root, in higher education and universities. She underlined that prestigious AI conferences have initiatives, from parallel sessions and workshops to policies supporting diversity and inclusion, that are trying to address the lack of women in the field and create an inclusive environment. She also highlighted international differences in gender diversity in research, with Mediterranean countries, such as Spain, having much fewer women studying Informatics than in northern Europe. Lastly, she emphasised that it is important for young researchers to work on topics that motivate them while she was positive that the broad domains of application and the potential impact of this technology will attract more women into the sector. In Section 3.1.4, we showed that, among the institutions with the highest number of AI publications on arXiv, only the University of Washington had more than 25 per cent women researchers. We interviewed **Ed Lazowska** and **Eve Riskin** the schemes they have put in place to achieve this and what still needs to be done.

Ed Lazowska is the Bill & Melinda Gates Chair in the Paul G. Allen School of Computer Science & Engineering at the University of Washington as well as a Senior Data Science Fellow in the University of Washington eScience Institute. He underscored that *"Diversity in AI and in all of computer science matters because these are inherently creative subjects focused on solving problems. Individuals bring their own backgrounds to tackling problems and the more perspectives included, the better the solution."* He also stressed out that this is a multifaceted problem, requiring a variety of interventions; creating a friendly and inclusive environment, actively encouraging women to apply for positions on the faculty and reaching out to potential students. Lastly, he mentioned the need to *"Transform the whole culture of an organisation (in a way that is) driven from the top, with leadership embracing it as a priority."*

Eve Riskin is the Associate Dean of Diversity and Access in the College of Engineering, Professor of Electrical & Computer Engineering and Faculty Director of the ADVANCE Center for Institutional Change where she works on mentoring and leadership development programs for women faculty in STEM. Eve commented that the environment in male dominated subjects can become 'toxic' and that the University of Washington is working on schemes with the department chairs on cultural change while providing professional development assistance to female faculty in STEM. As she highlighted, *"Research has shown that female undergrads achieve more under female faculty, and so you need a two-pronged approach."* At the same time, commitment, time and funding are needed to drive change. She concluded by pointing out that *"It is important to make under-represented groups more visible but this has to be done thoughtfully. If communications teams are highlighting the stories of women and minorities this must be done as part of a broader programme of activity. These groups have to be genuinely welcomed and nurtured, not just used for photo opps."*

We conclude our summary of interview results with **Sir Alan Wilson**, Executive Chair of the Ada Lovelace Institute and Director, Special Projects of The Alan Turing Institute, where he was CEO between 2016-2018. Sir Alan highlighted that 'precise and targeted interventions' that are focused on increasing diversity are needed in order to reduce the gender imbalance. He also commented that *"Further research into the kinds of interventions that can inspire young women and minority groups to pursue study all subjects that lead to careers in AI – computer science, but also maths, statistics and engineering – is needed."* Lastly, he stressed out the significance of creating an evidence base, as our work is trying to do, to benchmark the diversity gap in AI research and evaluate the impact of policies to address it.

5
Discussion

We have examined gender diversity in AI research, its drivers and links with paper content and citations. Our analysis confirms the idea that there is a gender diversity gap in AI research, in a larger and more comprehensive corpus than those which have been used to study this important issue before. While the share of papers involving female co-authors and the share of female authors in some fields are increasing, the situation in other fields, particularly those related to computer science where AI research is most important, has stagnated in recent years.

When considering international differences, we find countries such as Netherlands, Norway and Denmark have a much higher than average share of papers with female co-authors. We also find that women publishing on arXiv are often affiliated with universities, while there is a significant gender diversity gap in the top companies and research institutions. Furthermore, our findings suggest that both nationality and research domains play a role in influencing participation of women in AI publications. This means that national policies and institutions and social norms in research communities will both need to play a role in increasing female participation in AI research. Our experimental analysis of semantic differences between papers with or without female participation suggests that arXiv papers in specific fields and countries can be semantically different depending on the presence of female co-authors, and that papers with female co-authors have higher salience of terms related to social and political issues. This is consistent with out finding that women AI researchers are particularly active in the Societal research domain (which considers the social implications of new technologies).

Our qualitative interviews suggest that system-wide changes are required in order to reduce the AI gender gap. This includes interventions that encourage women to study and work in AI and Computer Science, the creation of safe and inclusive spaces that support and promotes researchers from underrepresented groups, and communicating more widely the transformative potential of AI in many domains and sectors. All this will require leadership, funding and changes in organisational cultures and attitudes.

Our work is not without limitations. Unfortunately, we do not capture non-binary gender categories, and we have excluded authors affiliated to institutions from some countries (most notably China) due to the relatively low accuracy with Chinese names of the name-to-gender inference system we used. Furthermore, our AI labelling system does not capture all the AI papers on arXiv, meaning that we are under-estimating their true number. Regarding the qualitative component of our research, our sample sizes are small and additional interviews would be needed to further validate our quantitative findings, understand their drivers and develop policy implications.

In future work, we aim to incorporate Chinese researchers in the analysis by triangulating their gender using a variety of name-to-gender inference APIs. We also plan to refine our AI labelling pipeline as well as our model for determining the drivers of diversity in AI research. Furthermore, we will examine the authorship network of AI papers to identify influential individuals and map their trajectory using their publications.We also intend to carry out more interviews and potentially survey the AI researchers on arXiv. We hope that these extensions of the analysis will allow us to understand in more detail the social and institutional determinants of gender diversity in AI, and to identify suitable interventions to improve it, making the development of this important technology and its outcomes more inclusive.

References and endnotes

1. Joel Klinger, Juan C Mateos-Garcia, and Konstantinos Stathoulopoulos. Deep learning, deep change? mapping the development of the artificial intelligence general purpose technology. Mapping the Development of the Artificial Intelligence General Purpose Technology (August 17, 2018), 2018.
2. Spyros Makridakis. The forthcoming artificial intelligence (ai) revolution: Its impact on society and firms. Futures, 90:46–60, 2017.
3. Whittaker M. West, S.M. and K. Crawford. Discriminating systems: Gender, race and power in ai. 2019.
4. Philip N Howard and Bence Kollanyi. Bots,# strongerin, and# brexit: computational propaganda during the uk-eu referendum. Available at SSRN 2798311, 2016.
5. Perrault Raymond Brynjolfsson Erik Clark Jack Manyika James Carlos Niebles Juan Lyons Terah Etchemendy John Grosz Barbara Shoham, Yoav and Zoe Bauer. The ai index 2018 annual report. AI Index Steering Committee, Human-Centered AI Initiative, Stanford University, Stanford, CA, 2018.
6. Element AI. Global ai talent report. 2019.
7. World Economic Forum. Global gender gap report. 2018.
8. Whittaker M. West, S.M. and K. Crawford. Discriminating systems: Gender, race and power in ai. 2019.
9. Juan C Mateos-Garcia. The complex economics of artificial intelligence. Available at SSRN 3294552, 2018.
10. Joel Klinger, Juan C Mateos-Garcia, and Konstantinos Stathoulopoulos. Deep learning, deep change? mapping the development of the artificial intelligence general purpose technology. Mapping the Development of the Artificial Intelligence General Purpose Technology (August 17, 2018), 2018.
11. https://arxiv.org/help/api
12. Arnab Sinha, Zhihong Shen, Yang Song, Hao Ma, Darrin Eide, Bo-june Paul Hsu, and Kuansan Wang. An overview of microsoft academic service (mas) and applications. In Proceedings of the 24th international conference on world wide web, pages 243–246. ACM, 2015.
13. Anne-Wil Harzing and Satu Alakangas. Microsoft academic: is the phoenix getting wings? Scientometrics, 110(1):371–383, 2017.
14. https://cloud.google.com/maps-platform/places/
15. Place IDs uniquely identify a place in the Google Places database and on Google Maps.
16. Wendy Liu and Derek Ruths. What's in a name? using first names as features for gender inference in twitter. In 2013 AAAI Spring Symposium Series, 2013.
17. https://gender-api.com/en/
18. https://gender-api.com/en/frequently-asked-questions/which-countries-are-supported
19. Lucía Santamaría and Helena Mihaljevic´. Comparison and benchmark of name-to-gender inference services. PeerJ Computer Science, 4:e156, 2018.
20. Ibid.
21. To address this issue, in our international analysis below we exclude those countries where less than 50 per cent of the gender inferences are 'good quality' in the sense that they are based on full author names and have accuracy above 80 per cent.
22. Iain M Cockburn, Rebecca Henderson, and Scott Stern. The impact of artificial intelligence on innovation. Technical report, National Bureau of Economic Research, 2018.
23. Joel Klinger, Juan C Mateos-Garcia, and Konstantinos Stathoulopoulos. Deep learning, deep change? mapping the development of the artificial intelligence general purpose technology. Mapping the Development of the Artificial Intelligence General Purpose Technology (August 17, 2018), 2018.
24. Bhaskar Mitra, Nick Craswell, et al. An introduction to neural information retrieval. Foundations and Trends R in Information Retrieval, 13(1):1–126, 2018.
25. Joel Klinger, Juan C Mateos-Garcia, and Konstantinos Stathoulopoulos. Deep learning, deep change? mapping the development of the artificial intelligence general purpose technology. Mapping the Development of the Artificial Intelligence General Purpose Technology (August 17, 2018), 2018.
26. Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781, 2013.
27. cs.LG is a sub-category of Computer Science that contains Machine Learning papers. It is the most popular AI related category on arXiv.
28. Joel Klinger, Juan C Mateos-Garcia, and Konstantinos Stathoulopoulos. Deep learning, deep change? mapping the development of the artificial intelligence general purpose technology. Mapping the Development of the Artificial Intelligence General Purpose Technology (August 17, 2018), 2018.
29. Whittaker M. West, S.M. and K. Crawford. Discriminating systems: Gender, race and power in ai. 2019.
30. We removed 2019 from the analysis as we have collected only January and February for that year.
31. We report figures only for the authors with inferred gender.
32. Whittaker M. West, S.M. and K. Crawford. Discriminating systems: Gender, race and power in ai. 2019.
33. This might be caused by the fact that research tends to be more collaborative in recent years.
34. Note that the population of papers on arXiv before and after 2000 is significantly different. For example, Astrophysics, Maths and Physics were introduced in the early 1990s while Statistics in 2007 and Electrical Engineering in 2017.
35. One explanation for this finding is that there are outlier non-AI papers in some disciplines such as astrophysics that tend to involve large numbers of unique female co-authors, increasing the average share of unique female authors in the total.
36. All interviews were conducted between 27 June 2019 and 12 July 2019 via video-conference and lasted one hour. All interviewees have given us authorisation to identify them in this section.

nesta

58 Victoria Embankment
London EC4Y 0DS

+44 (0)20 7438 2500

information@nesta.org.uk

@nesta_uk

www.facebook.com/nesta.uk

www.nesta.org.uk