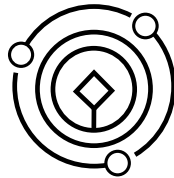


Chapter Preview

Journal Publications: Elsevier	14
Papers on arXiv	21
Microsoft Academic Graph	
Journals	24
Conferences	27
Patents	30
Github Stars	33
Women in AI Research	34
Measurement Questions	36

Chapter 1: Research and Development



Introduction

This chapter presents bibliometrics data, including volume of journal, conference and patent publications and their citation impacts by world regions. The chapter also presents Github Stars for key AI software libraries followed by societal considerations and gender diversity of AI researchers based on arXiv.

The Report has used different datasets to comprehensively assess the state of AI R&D activities around the world. The MAG dataset covers more publications than Elsevier's Scopus, which is mostly limited to peer-reviewed publications, but there are also publications on Scopus that are not in MAG.² arXiv, an online repository of electronic preprints, reflects the growing tendency of certain parts of the field of AI, particularly those depending on machine learning, to post papers before peer review, so reflects recent work more quickly than the other sources. Our [arXiv Monitor](#) tool uses full-text papers to quickly identify new results.

²see these studies by [Anne-Wil Harzing](#) and [Martijn Visser](#).



Published Papers: AI Papers in All Publications

Elsevier's [Scopus](#) is the world's largest abstract and citation database of peer-reviewed literature with over 22,800 titles from more than 5,000 international publishers. The graph below (Figure 1.1) shows the percentage of AI publications in peer-reviewed publications (conferences, reviews, and articles) between 1998-2018. Here, AI papers correspond to all publications in AI, including journal publications

and conference publications in the Scopus database. In the late 1990's AI papers accounted for less than 1% of articles and around 3% of conference publications. By 2018, the share of published AI papers in total papers has grown three-fold in 20 years, accounting for 3% of peer reviewed journal publications and 9% of published conference papers (see [Appendix Graph](#)).

AI Publications in All Publications

Source: Scopus, 2019.

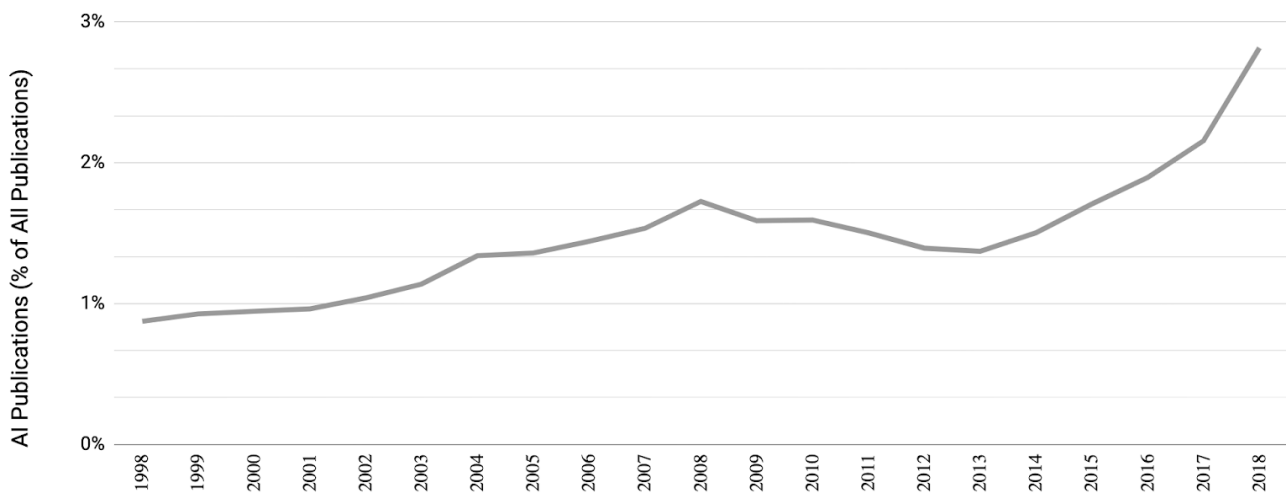


Fig. 1.1.

Between 1998 and 2018, the share of AI papers among all papers published worldwide has grown three-fold, now accounting for 3% of peer reviewed journal publications and 9% of published conference papers.



Published Papers: AI Papers By Region

Which regions witnessed the fastest growth in peer-reviewed AI publications? The graphs below show the number of AI papers published annually by region (Figure 1.2a), and the growth in AI papers published by region (Figure 1.2b). Europe has consistently been the largest publisher of AI papers — rising to over

27% of AI publications tracked by Scopus in 2018. Papers published from Chinese entities increased from 10% of global AI publications in 2000 to 28% in 2018 (see [Appendix Graph](#)). See [Technical Appendix](#) for data and methodology.

Annual Number of AI Papers on Scopus

Source : Elsevier, 2019.

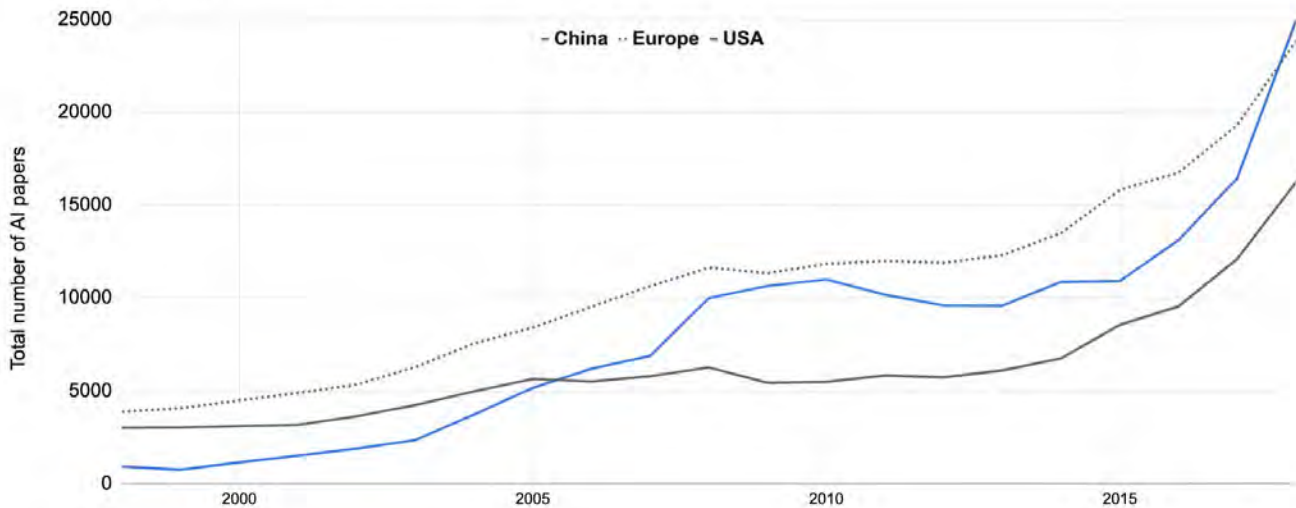


Fig. 1.2a.

Annual Growth in AI papers on Scopus

Source: Elsevier, 2019.

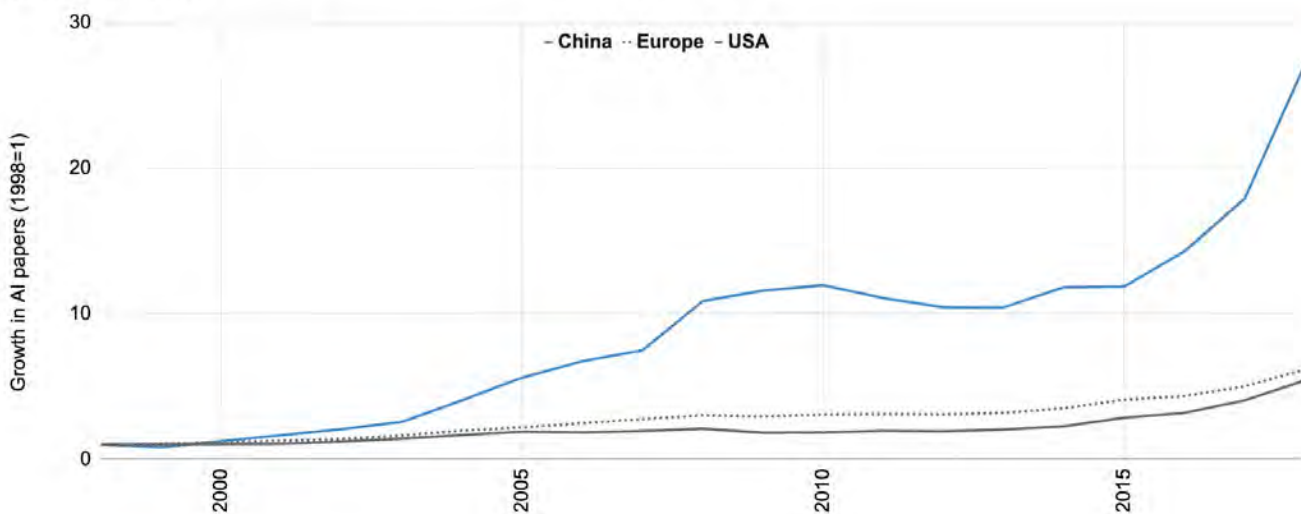


Fig. 1.2b.



Published Papers: Institutional Affiliation

The following graphs show the number of Scopus papers affiliated with government, corporate, medical, and other organizations in China (Figure 1.3a), the United States (Figure 1.3b), and Europe (Figure 1.3c). Excluding academia, the graphs show that government-affiliated institutions contribute the highest number of AI publications in China and Europe, whereas, corporate-affiliated AI papers make up a higher proportion in the US

In 2018, Chinese government institutions produced nearly three times more AI papers than Chinese corporations. China has also seen a 300-fold increase in government-affiliated AI papers since 1998, while corporate AI papers increased by 66-fold in the same period.

In the US., a relatively large proportion of AI papers are affiliated with corporations. In 2018, the number of corporate-affiliated AI papers in the US was over seven times the proportion of corporate AI papers in China, and almost twice that of Europe.

Note that in all three regions, academic papers (not shown) outweigh government, corporate, and medical papers by a large margin, making up 92% of AI publications from China, 90% from Europe, and 85% from the US [Growth trends of institutional affiliation dynamics are available in the Appendix.](#)

Total number of papers by institutional affiliation, China (1998-2018)

Source: Elsevier, 2019.

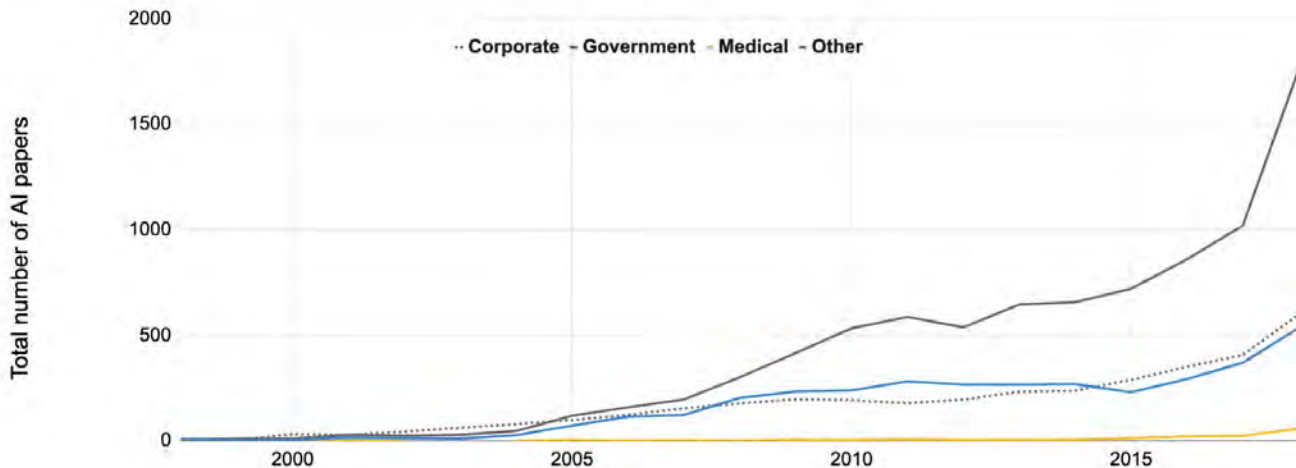


Fig. 1.3a.



Published Papers: Institutional Affiliation

Total number of papers by institutional affiliation, USA (1998-2018)

Source: Elsevier, 2019.

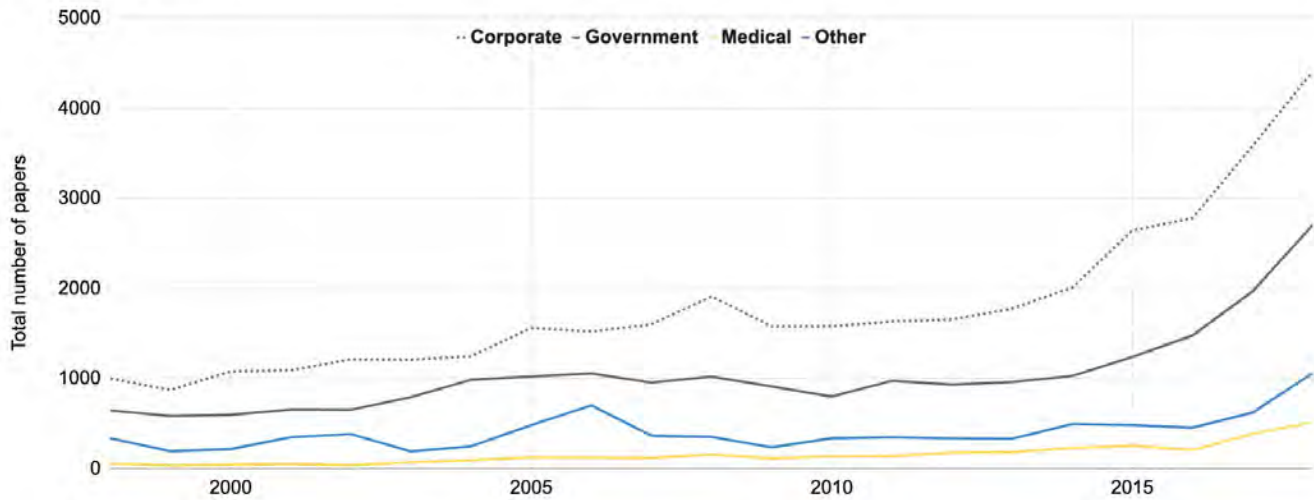


Fig. 1.3b.

Total number of papers by institutional affiliation, Europe (1998-2018)

Source: Elsevier, 2019.

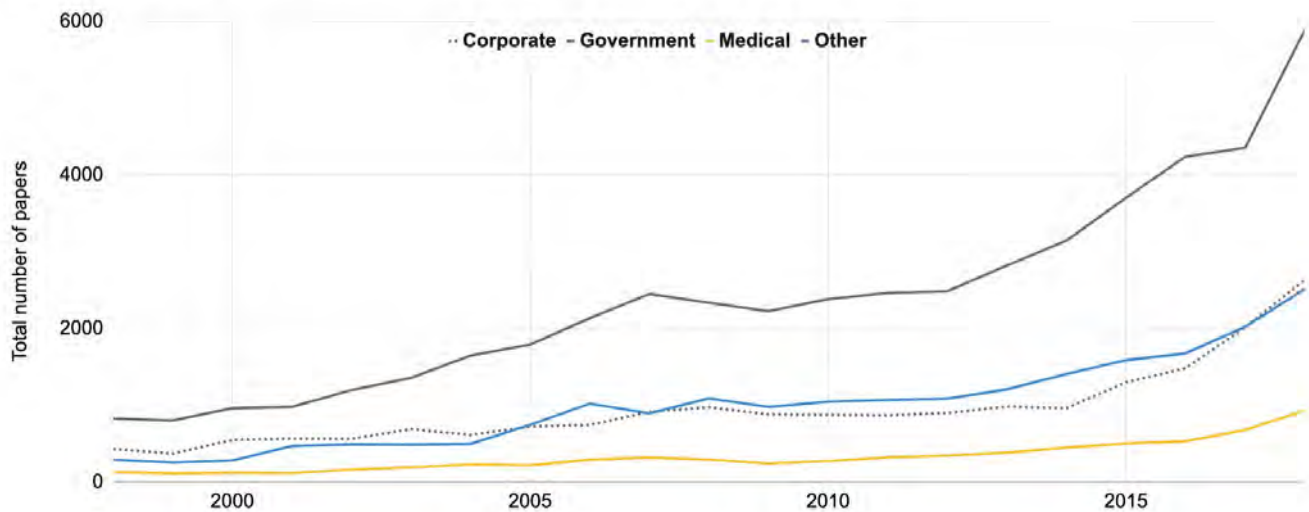


Fig. 1.3c.

Note: Europe refers to EU44.



Published Papers: Citation impact by region

The graph below (Figure 1.4) shows the average field-weighted citation impact of AI authors by region. A region's **Field-Weighted Citation Impact (FWCI)** is the average number of citations received by AI publications originating from that region divided by the average number of citations by all AI publications worldwide in the same publication year, subject area, and document type.

In this visual, the citation impacts are shown relative to the world average for AI, whose FWCI is normalized at 1. A re-based FWCI of 1 indicates that the publications have been cited on par with the world average for AI. A re-based FWCI of 0.85 indicates that the papers are 15% less cited than the world average for AI.

While Europe has the largest number of annually published AI papers in Scopus, Europe's FWCI has remained relatively flat and on-par with the world average. In contrast, China has increased its FWCI considerably. Still, the US outperforms other regions in total citations. Authors from the US are cited 40% more than the global average. See [Technical Appendix](#) for data and definitions. Both the US and China are gaining in prominence in Field-Weighted Download Impact (FWDI) of AI publications (see [Appendix Graph](#)).

Field-Weighted Citation Impact of AI authors by region, (1998-2018)

Source: Elsevier, 2019.

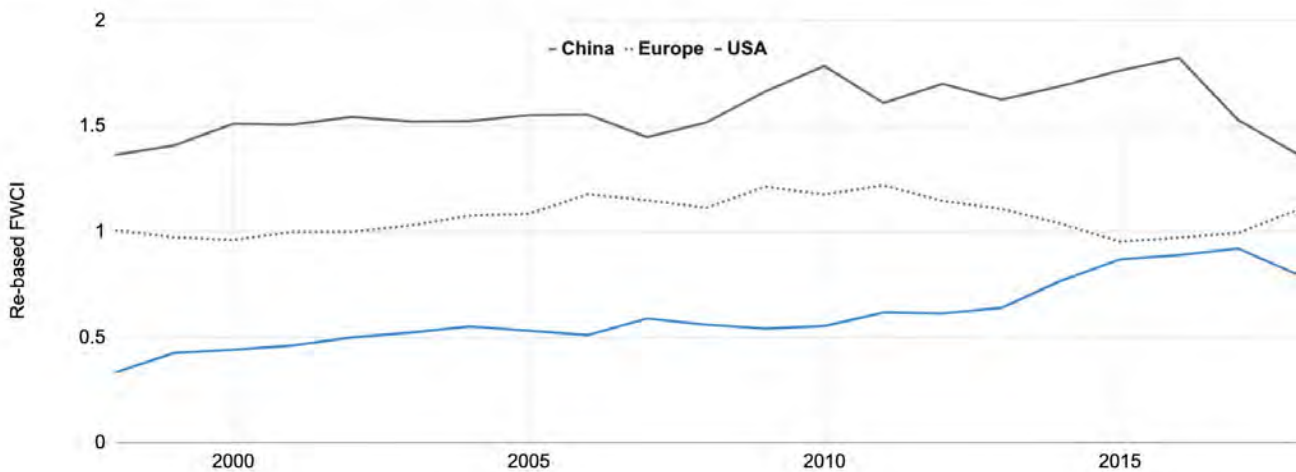


Fig. 1.4.



"China has consistently increased its footprint in AI research, both in terms of volume and quality. Their advance is truly remarkable."

— Maria de Kleijn, SVP Analytical Services, Elsevier



Cross Country Trends in Impact and Academic-Corporate Collaboration

In recent years it's increasingly common for AI-focused companies to conduct research in partnership with colleagues in academia. This map (Figure 1.5a) shows the quantity of academic-corporate collaborations in different countries around the world. Academic-corporate collaborations are

identified through publications with at least one author with an academic affiliation and at least one author with a corporate affiliation. Academic-corporate AI collaborations are largely prevalent in the US, China, Japan, France, Germany, and the UK.

World Map of Academic-Corporate Collaboration: Total Number of AI papers

Source: Scopus, 2019.

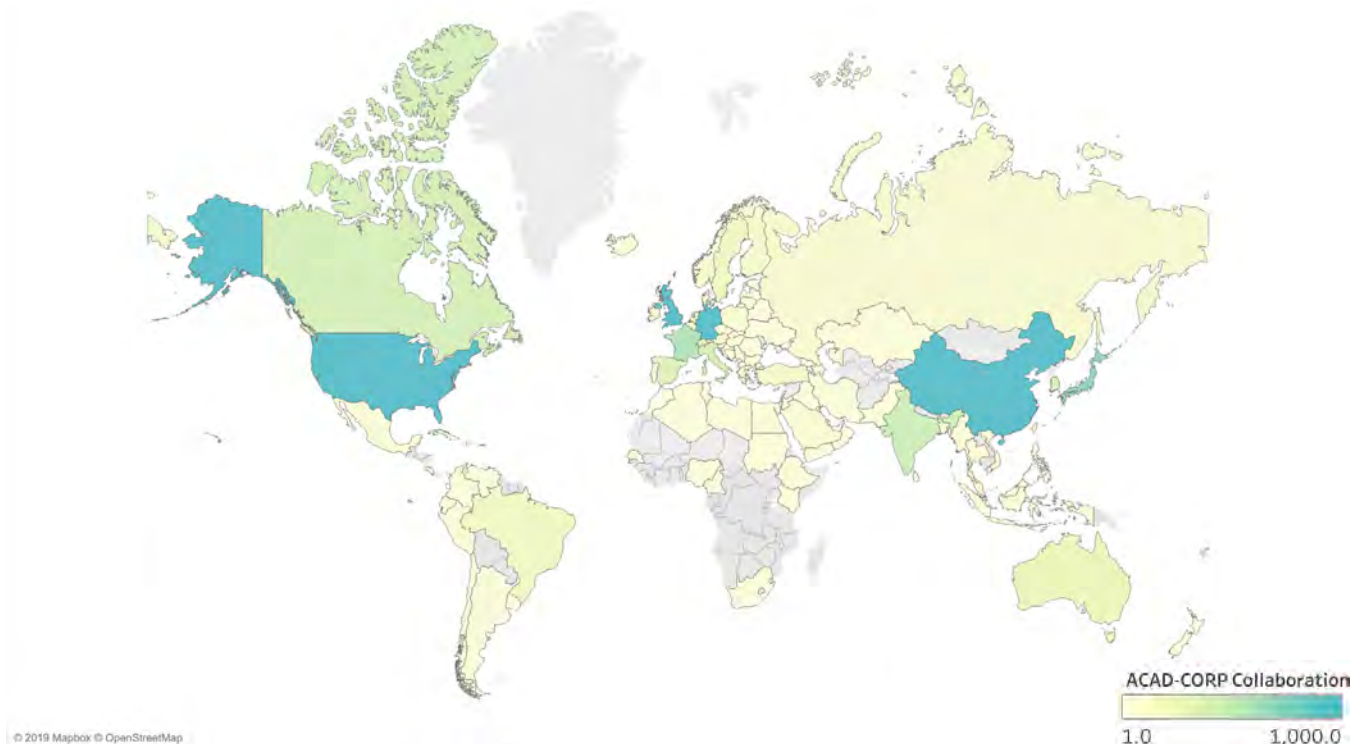


Fig. 1.5a.

Academic-corporate AI collaborations are prevalent in the US, China, France, Hong Kong, Switzerland, Canada, Japan, Germany, and the UK. US Authors are cited 40% more than the global average.



Cross Country Trends in Impact and Academic-Corporate Collaboration

How do academic-corporate collaborations impact the overall FWCI of AI research publications from different countries? This graph (Figure 1.5b) shows the FWCI (for all AI papers) on the y-axis and the total number of AI papers based on academic-corporate collaborations on the x-axis. The chart can be split into four quadrants: high degree of

collaboration and high degree of impact (*top right quadrant*); low degree of collaboration but high impact (*top left quadrant*); low degree of collaboration and low impact (*bottom left quadrant*); high degree of collaboration but low impact (*top left quadrant*); Chart for countries across scholarly output metrics is available in the [Appendix](#).

Four Quadrants for Overall AI Citation Impact (vertical axis) and the Total number of Academic-Corporate AI Papers (horizontal axis)

Source: Scopus, 2019.

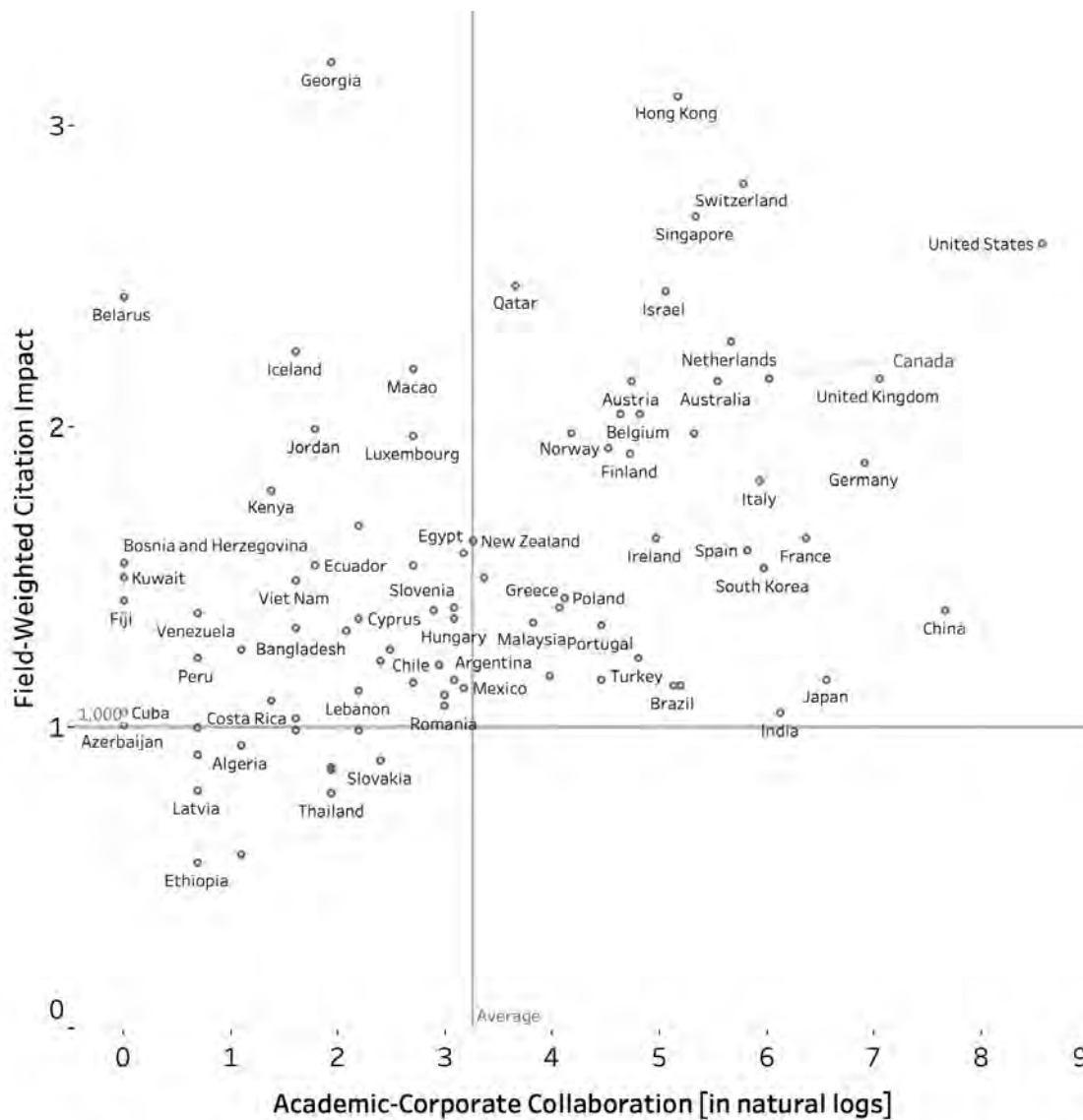


Fig. 1.5b.



"Counter to common assumptions, working together with Corporate institutions is beneficial to the academic impact of universities."

Maria de Kleijn, SVP Analytical Services, Elsevier



AI papers on arXiv

In recent years, AI researchers have adopted the practice of publishing paper pre-prints (frequently before peer-review) on arXiv, an online repository of electronic preprints. The graph below shows the number of AI papers on arXiv by each paper's primary subcategory (Figure 1.6).

The number of AI papers on arXiv is increasing overall and in a number of subcategories, reflecting a broader growth in AI researchers publishing preprints of their research. Between 2010 and 2019, the total number of AI papers on arXiv increased over twenty-fold. Submissions to the Computation & Language arXiv sub-category have grown almost sixty-fold since 2010.

In terms of volume, *Computer Vision (CV)* and *Pattern Recognition* had been the largest AI subcategory on arXiv since 2014 but *Machine Learning* has become the largest category of AI papers in 2019. In addition to showing a growing interest in *Computer Vision* and *Machine Learning* (and its general applied applications), this chart also indicates growth in other AI application areas, such as *Robotics* growing over thirty-fold between 2010 and 2019. See [Technical Appendix](#) for data and methodology.

Number of AI papers on arXiv, 2010-2019

Source: arXiv, 2019.

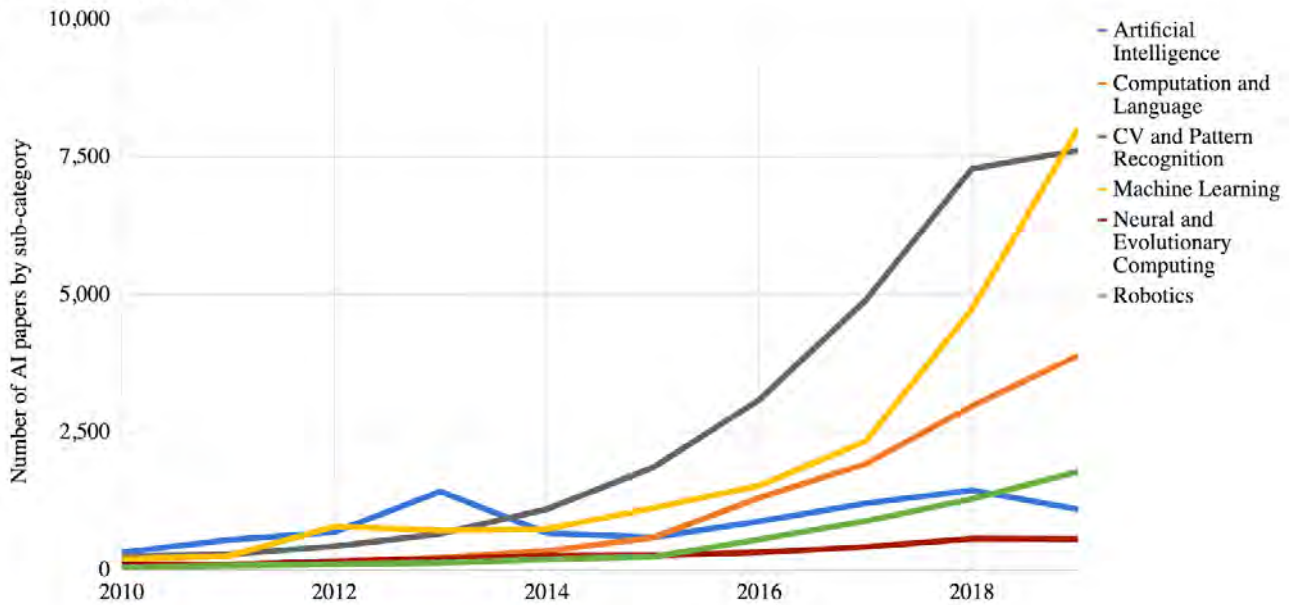


Fig. 1.6.



Deep Learning Papers on arXiv

Machine Learning (ML) is one of the most active research areas in AI. Within ML, Deep Learning (DL) approaches have become increasingly popular in recent years. The number of deep learning (DL) papers published on arXiv is increasing across regions. The first chart (Figure 1.7a) shows that North America published the largest volume of DL papers, followed by Europe in 2018. The volume of DL papers from East Asia reached the same level as Europe in 2018.

The following graphs show the ranking of countries with the largest volume of DL papers (Figure 1.7b) as well as the associated per capita DL papers (Figure 1.7c). Singapore, Switzerland, Australia, Israel, Netherlands, and Luxembourg have relatively high per capita DL papers published on arXiv. More details on methodology (see [Technical Appendix](#)) and detailed country chart (see [Appendix Graph](#)).

Number of Deep Learning Papers on arXiv

Source: arXiv, NESTA, 2019.

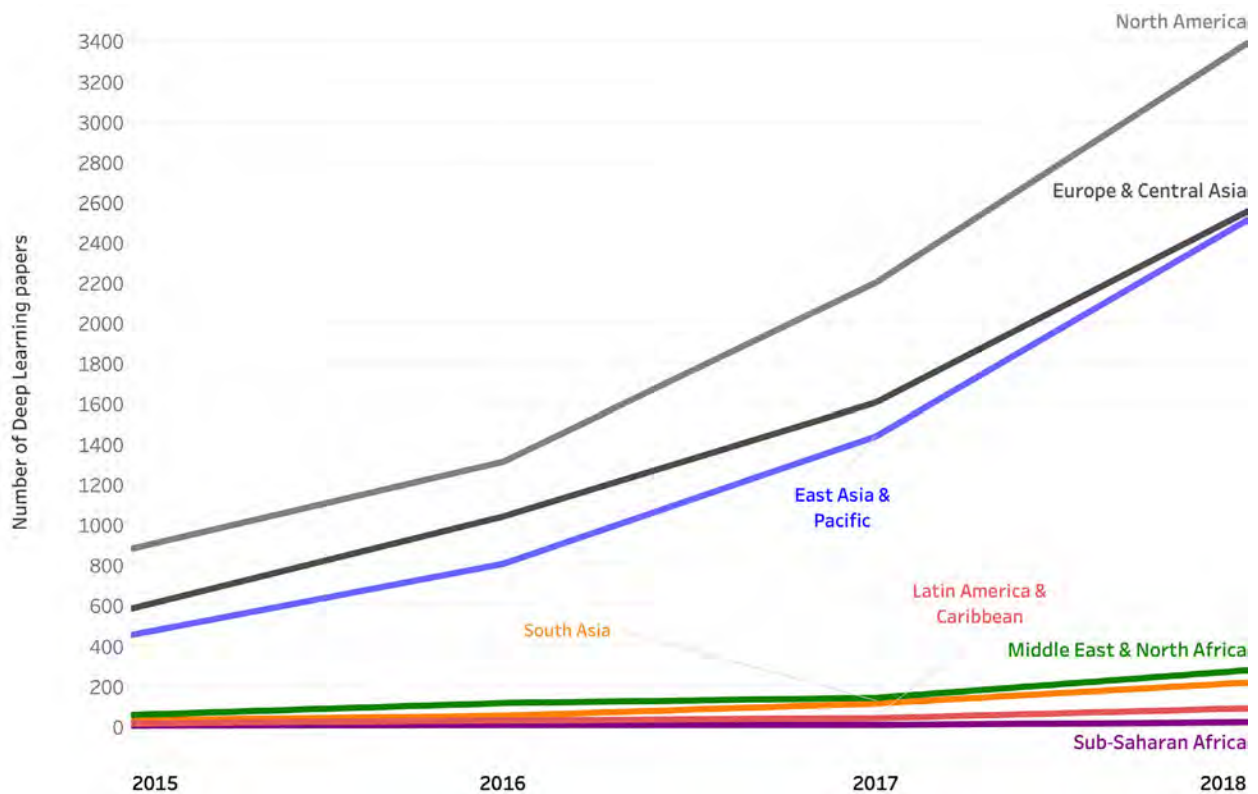


Fig. 1.7a.

Notes on World Regions:

In the following sections, cross-country bibliometrics analysis may correspond to World Bank region codes where explicitly stated. The regions include: East Asia & Pacific, Europe & Central Asia, Latin America & Caribbean, Middle East & North Africa, North America, South Asia, and Sub-Saharan Africa. "East Asia" can be referred to East Asia & Pacific and "Europe" to Europe & Central Asia. The [country codes](#) and [API](#) are available.

Singapore, Switzerland, Australia, Israel, Netherlands, and Luxembourg have relatively high per capita DL papers published on arXiv.



Deep Learning Papers on arXiv

Ranking Countries based on Total Number of Deep Learning Papers on arXiv, 2015-18

Source: arXiv, NESTA, 2019.

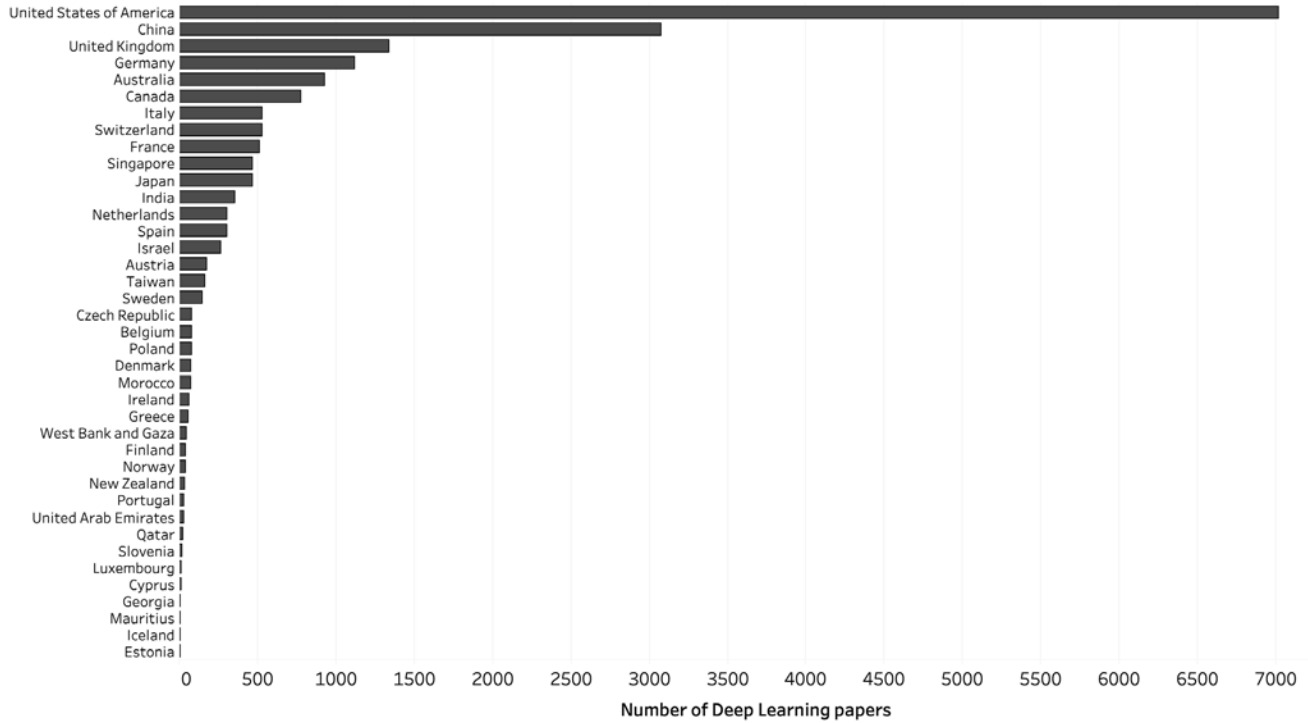


Fig. 1.7b.

Ranking Countries based on Number of Deep Learning Papers per capita on arXiv, 2015-18

Source: arXiv, NESTA, 2019.

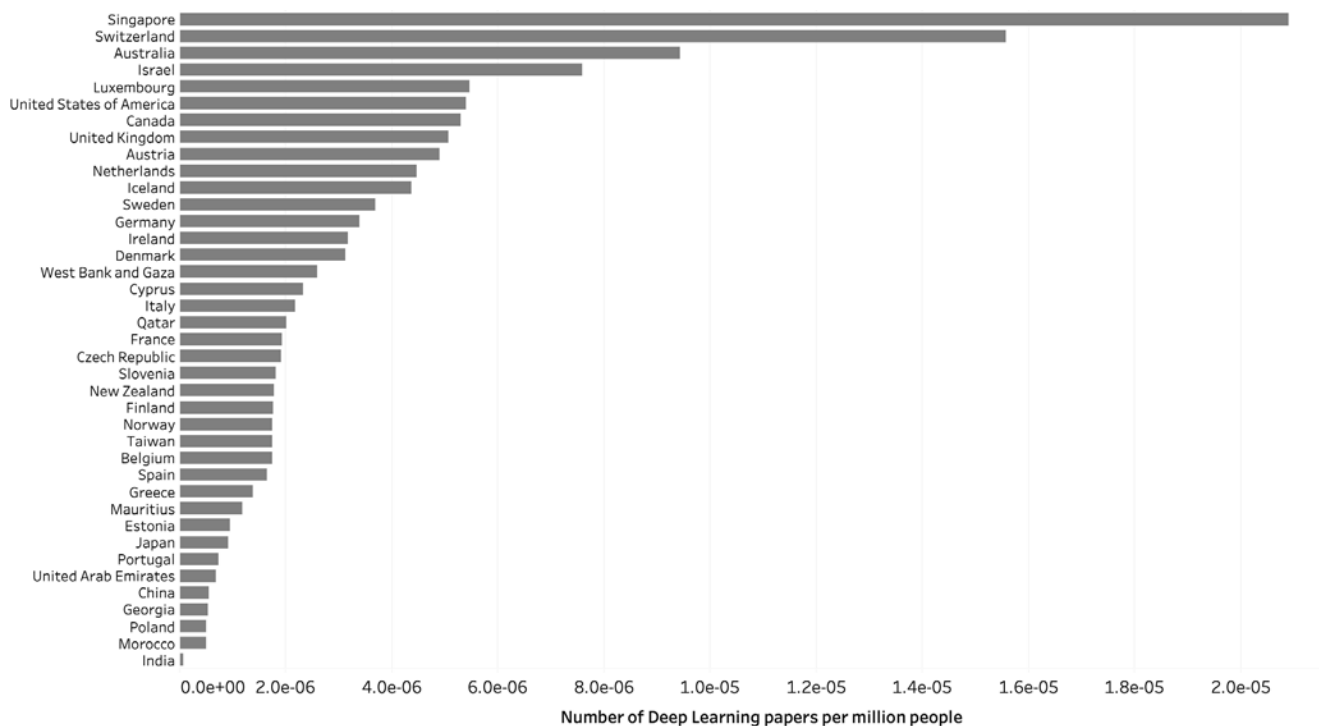


Fig. 1.7c.



Published Papers: AI Journal Publications

The graph below (Figure 1.8a) shows the share of AI journal papers on Microsoft Academic Graph (MAG) by world regions between 1990-2018. 37% of published journal papers are attributed to East Asia and Pacific (herein referred to as East Asia), 24% to Europe and Central Asia (herein referred to as Europe), and 22% to the North America in 2018. The share of South Asia in world AI journal publications has risen steadily to almost 8% in 2018.

The following graph (Figure 1.8b) shows the total number of journal publications and average journal publications per million people between 2015-18. China had the highest volume of AI papers, followed by the US, India, UK, and Germany. East Asia has the highest volume of AI journal papers on MAG (see [Appendix Graph](#)).

Share of World AI Journal Publications (%), 1990-2018

Source: MAG, 2019.

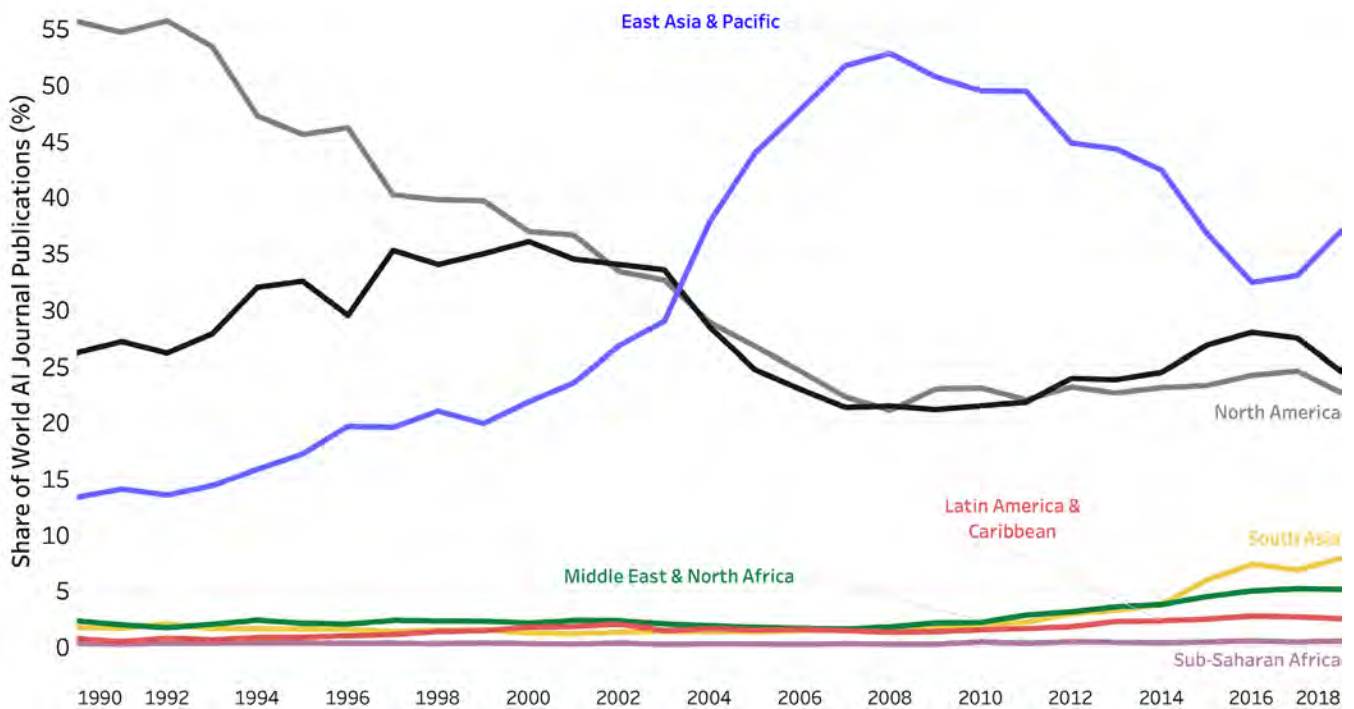


Fig. 1.8a.



Published Papers: AI Journal Publications

Total Volume and average annual per capita AI Journal Publications, 2015-2018

Source: MAG, 2019.

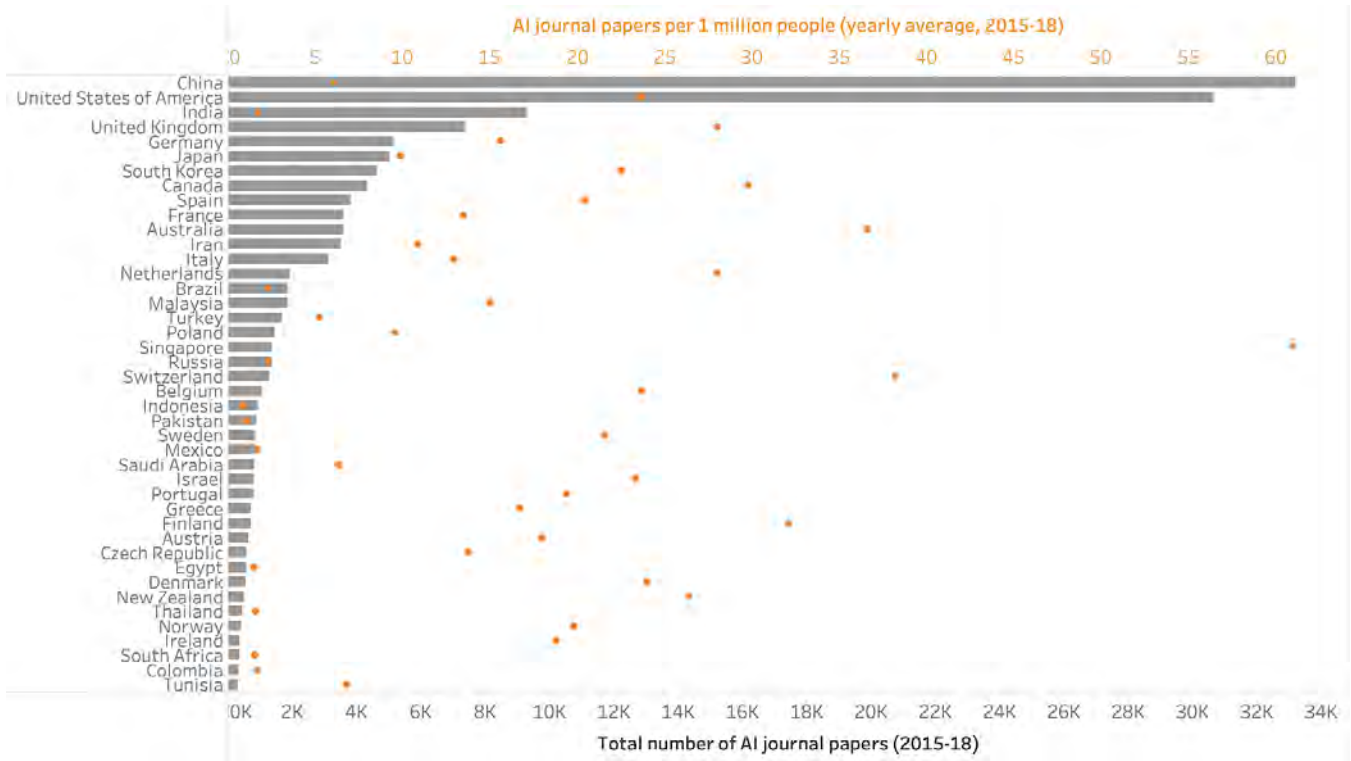


Fig. 1.8b.

In 2018, China had the highest volume of AI journal papers, followed by US, India, UK, and Germany.



Published Papers: AI Journal Citation

AI journal citation provides a signal for AI R&D impact. The share of world AI journal citation from **all** journal papers in MAG data is presented (see Box 1.1). North American papers were most cited by East Asian authors over 220k times, followed by European authors over 191k times. The interactive graphs are available on [the web](#). Methodology paper [A Century of Science: Globalization of Scientific Collaborations, Citations, and Innovations](#).

AI journal citations to East Asia journal papers account for over 32% of world citations; followed by Europe accounting for over 31%, and North America over 27% of world AI journal citations (Figure 1.9).

Box 1.1

- Between 2014-18, 17% of world citation was self-citation with East Asia; 15% was self-citation within Europe; 9% was self-citation within North America.
- Between regions, 8% of world citations were East Asian journals papers citing North American journal papers and 7% papers citing North American papers.
- 7% of world citations were East Asian journal papers citing European papers. The share of European and North American journal papers citing East Asian journals was 5% of world citation each.

Note: Percentage of journal citations to unknown country is 19.1%. Self-citation in these sections is referred to citation from one region to the same, not the more conventional author-cites-self interpretation.

AI Journal Citation Attributed to Region (% of world journal citations), 2014-18

Source: MAG, 2019.

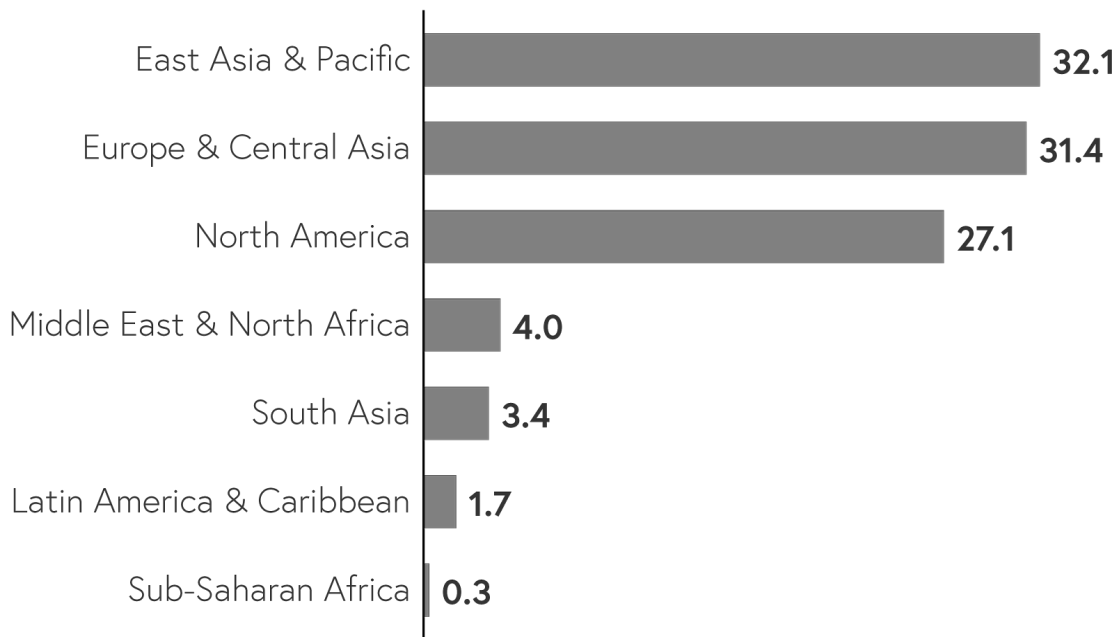


Fig. 1.9.

Note: Percentage of journal citations to unknown country is 19.1%.

AI journal citations to East Asia journal papers account for over 32% of world citations; followed by Europe accounting for over 31%, and North America over 27%



Published Papers: AI Conference Publications

The graph below (Figure 1.10a) shows the share of AI conference papers on MAG by world regions between 1990 and 2018. 33% of published AI conference papers are attributed to East Asia, 27% to North America, 26% to Europe in 2018. The share of South Asia in world AI conference publications has risen steadily to almost 6% in 2018.

The following graph (Figure 1.10b) shows the total number of AI conference publications and number of AI conference publications per million people between 2015-18. The US followed by China, India, Japan, and Germany had the highest volume of published AI conference papers. See [Technical Appendix](#) for data and methodology.

Share of World AI Conference Publications (%), 1990-2018

Source: MAG, 2019.

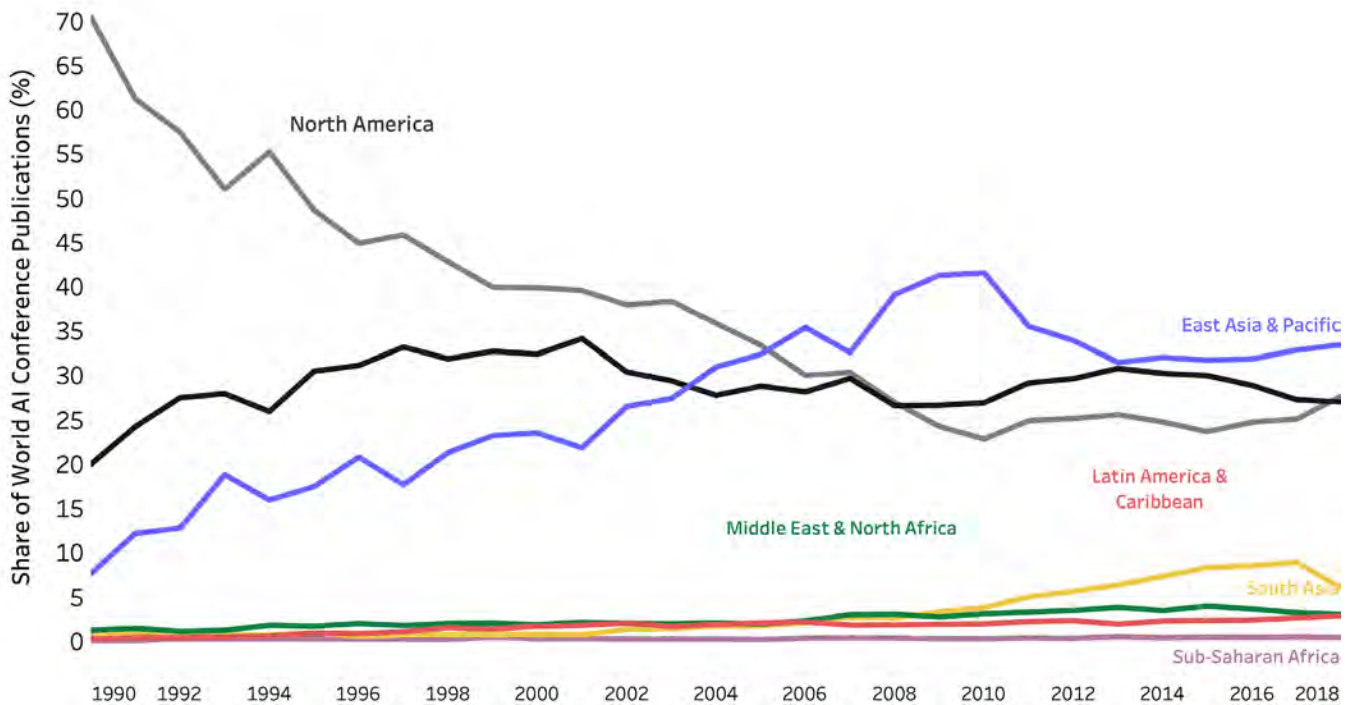


Fig. 1.10a.



Published Papers: AI Conference Publications

Total Volume and average annual per capita AI Conference Publications, 2015-2018

Source: MAG, 2019.

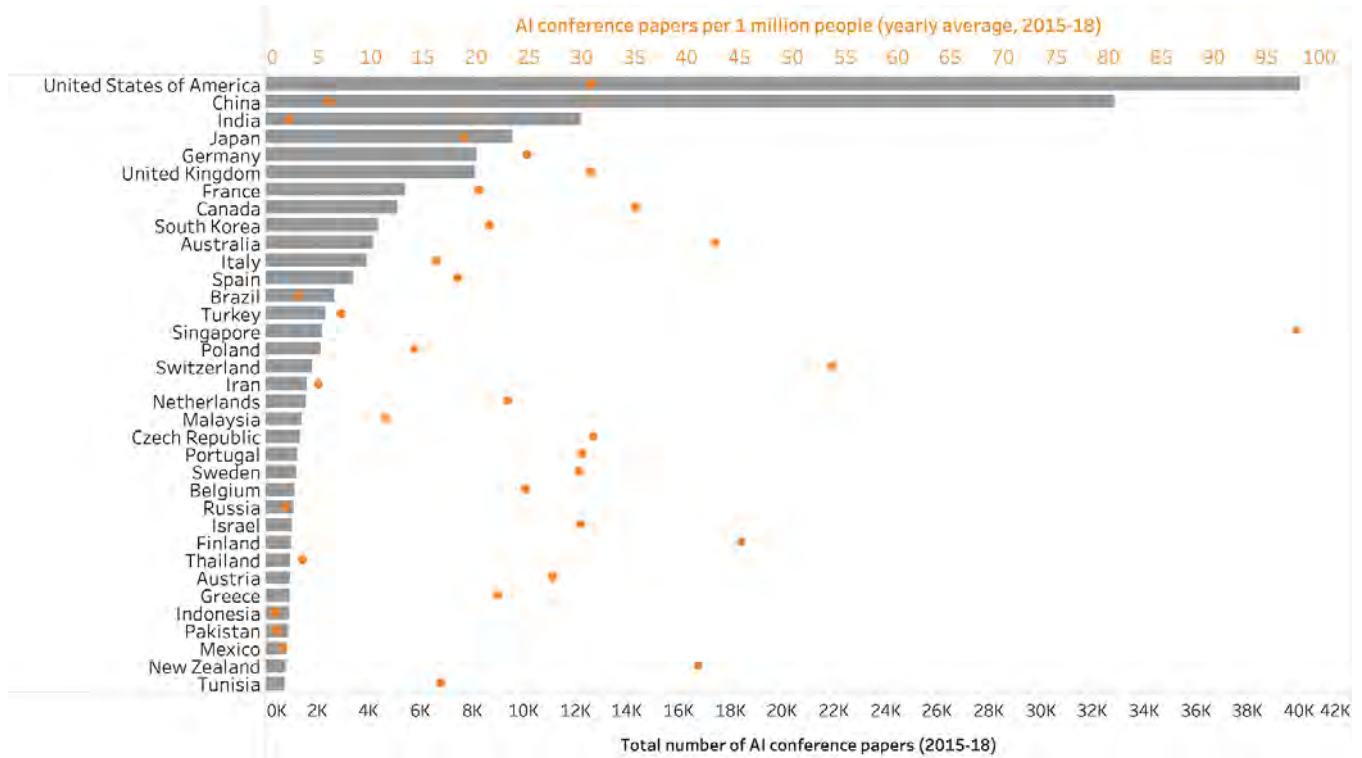


Fig. 1.10b.



Published Papers: AI Conference Citation

Over 40% of world conference paper citations are attributed to North America (self citation - 17%, East Asia - 13%, Europe - 10% of world citation). Self-citation in Europe accounted for 13% and self-citation in East Asia accounted for 11% of world conference publication citation. Box 1.2. presents the highlights for conference citation and the interactive graphs are available on [the web](#).

Almost 43% of world conference citations in AI papers is attributed to North American conference papers. The share of world citation in AI conference papers to European papers was over 28%, and to East Asian papers was over 22% of world AI conference citation activity (Figure 1.11).

Box 1.2.

- Citations to European conference papers by North America and East Asia accounted for 7% and 6% respectively of world conference citation.
- Citation to East Asian papers by North America and Europe accounted for 6% and 4% respectively of world conference citation

Note: Percentage of conference citations to unknown country is 12.7%. Self-citation in these sections is referred to citation from one region to the same, not the more conventional author-cites-self interpretation.

AI Conference Citation Attributed to Region (% of world journal citations), 2014-18

Source: MAG, 2019.

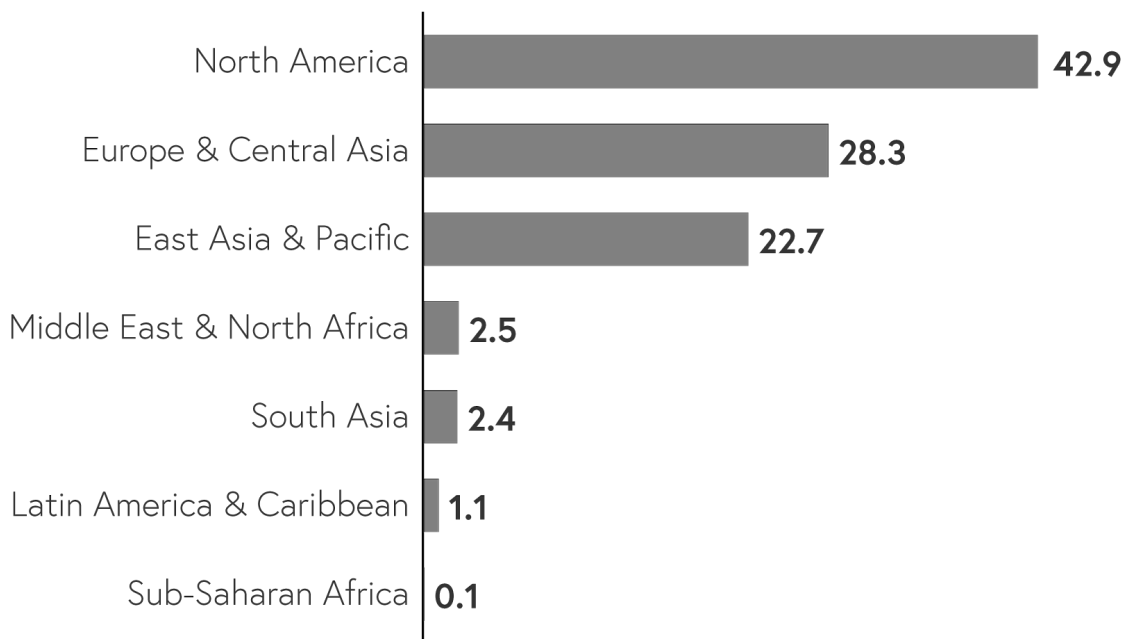


Fig. 1.11.

Note: Percentage of conference citations to unknown country is 12.7%.

Over 40% of world AI conference paper citations are attributed to North America (regional self citation - 17%, from East Asia - 13%, from Europe - 10% of world citation).



AI Patents

Patents on AI technology provide a measurement of AI activity in industry and its potential impact on products. The graph below (Figure 1.12a) shows the share of AI patents on MAG by world regions between 1990-2018. The graph for total number of AI patents published by regions can be found in the Appendix. Over 51% of published AI patents are attributed to the North America, with the share of Europe and Central Asia declining to 23%, close to East Asia & Pacific.

The following graph (Figure 1.12b) shows the total number of AI patents and average per capita AI patent publications between 2015-18. The US published three-folds the number of AI patents of the next country, Japan. Over 94% of AI patents are filed in high income countries, with the share of upper middle-income countries rising to 4% in 2018 (see [Appendix Graph](#)). See [Technical Appendix](#) for data and methodology.

Share of World AI Published Patents (%), 1990-2018

Source: MAG, 2019.

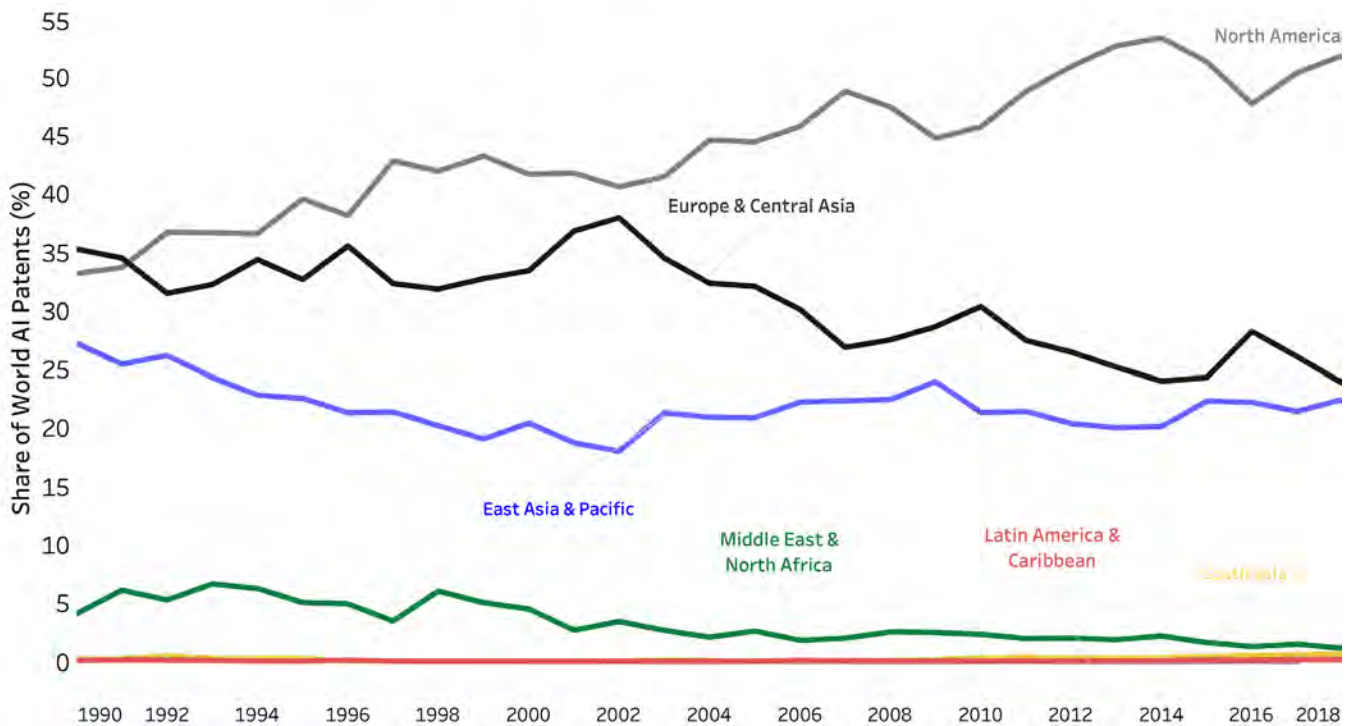


Fig. 1.12a.



AI Patents

Total Volume and average annual per capita AI Published Patents, 2015-2018

Source: MAG, 2019.

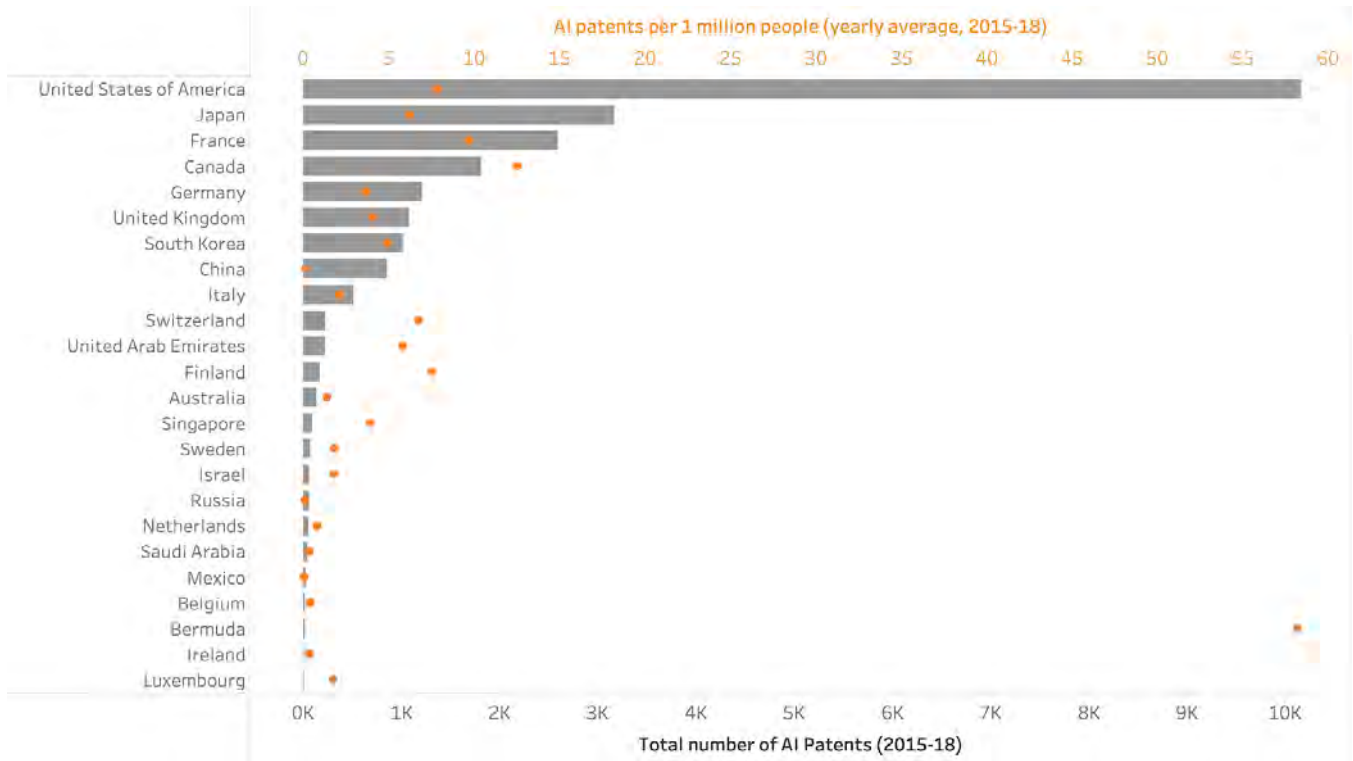


Fig. 1.12b.

Over 94% of AI patents are filed in high income countries, with the share of upper middle-income countries rising to 4% in 2018.



AI Patents Citations

The box below (Box 1.3) presents highlights AI patent citation from all patents. The insights on patent citation is revealing. Majority of world AI patent flow is dependent on North America. The interactive graphs are available on [the web](#).

Citations to North American AI patents accounted for over 60% of world patent citation activity; followed by East Asia with over 22%, and Europe with over 17% of AI patent citation (Figure 1.13).

Box 1.3.

- Over 60% of AI patent citation activity is related to North America, with almost 45% (of world AI patent citation) self-citation, 9% from East Asia patents, and 7% from European patents

- North American patents cited European and East Asian patents around 6,000 times between 2015-18, with the individual regions accounting for 6-7% each of world patent citations

note: Percentage of patent citations to unknown country is 37.2%

AI Patent Citation Attributed to Region (% of world journal citations), 2014-18

Source: MAG, 2019.

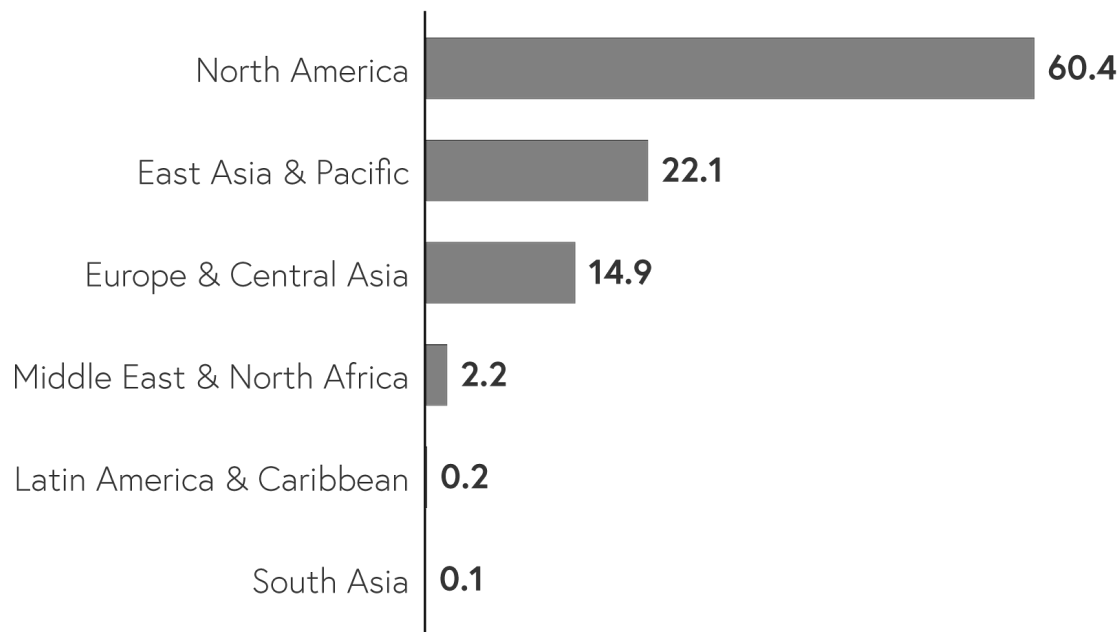


Fig. 1.13.

Note: Percentage of patent citations to unknown country is 37.2%.

North America accounts for over 60% of global AI patent citation activity between 2014 and 2018.



GitHub stars

GitHub is a website where developers upload, comment on, and download software code. Stars indicate a person has expressed interest in a particular piece of code and/or project on GitHub, similar to how 'likes' on social media services like Twitter and Facebook can indicate popularity of a given post. GitHub Stars therefore provide a rough measure of the popularity of various AI-programming frameworks. The graphs below show the number of times various AI and ML software packages have been starred on GitHub (Figure 1.14a and 1.14b).

One noticeable trend is the emergence of corporate-backed research frameworks, like Tensorflow (which was developed predominantly by Google) and PyTorch (which was developed predominantly by Facebook). Note that Keras popularity appears to tail off, but Keras has subsequently been integrated into TensorFlow, so its popularity is partially reflected in that metric. Two non-industry frameworks, sci-kit learn and Caffe, continue to show growing popularity, but their growth trajectories appear lower than those of the corporate frameworks.

Cumulative GitHub stars by AI library (2015—2019)

Source: Github, 2019.

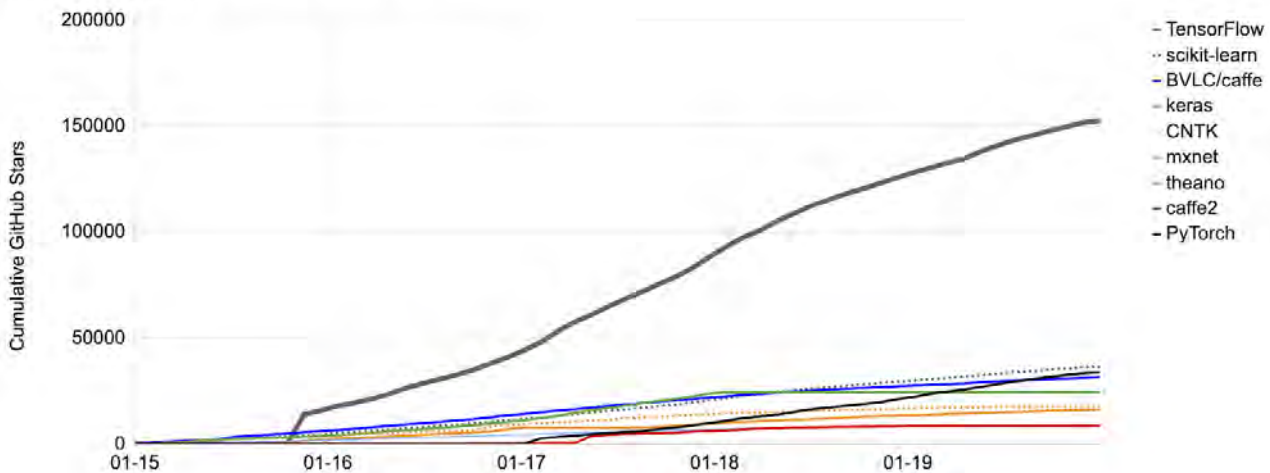


Fig. 1.14a.

Cumulative GitHub stars by AI library, not including TensorFlow (2015—2019)

Source: Github, 2019.

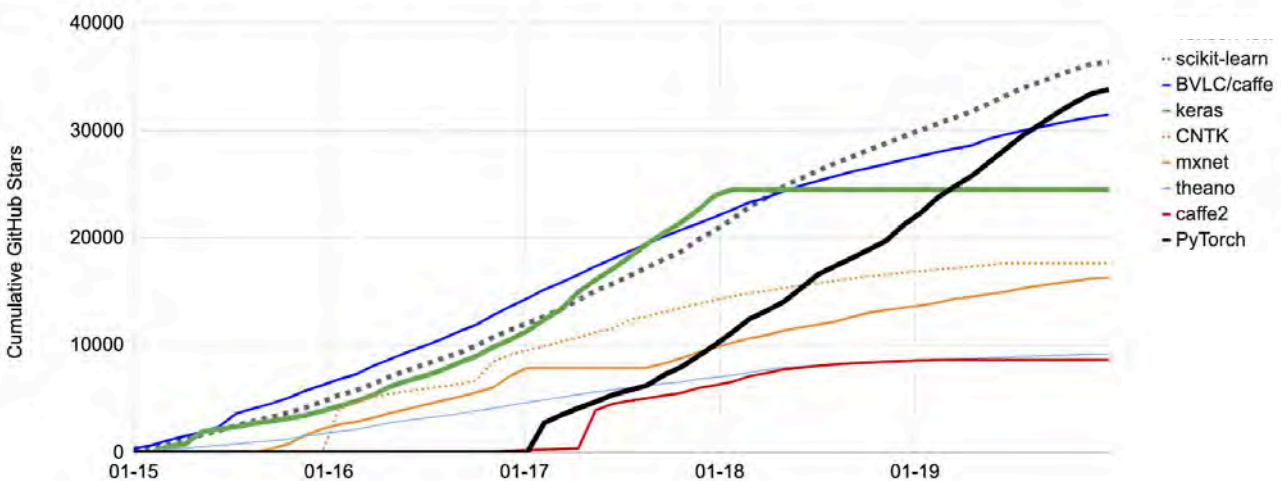


Fig. 1.14b.



Women in AI

There are significant international differences in the gender diversity of AI researchers. Half of the authors could be gender-labelled by first name with a high degree of confidence (China, one of the world leaders in AI research, is excluded from the sample due to a lower confidence in gender-labelling authors by name, and will be included in 2020). Countries with less than 5,000 publications on arXiv are not considered in this analysis. Technical Appendix provides details on data and methodology.

The differences between the share of female authors in AI and non-AI (refers to publications in all fields) papers within countries are presented below (Figure 1.16a). Over 41% of the AI papers in the Netherlands and over 39% of AI papers in Denmark 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.

Countries such as Malaysia, Denmark, Norway and Israel show a stronger presence of women in AI research relative to non-AI papers.

The Women in AI report from NESTA can be found [here](#). The longitudinal country data showing the share of female authors in AI and non-AI publications from NESTA is available [here](#) with the 30 countries with most publications. The change in share of women authors in AI is presented from 2000-2018, showing growth in AI publications with female authors from Europe (Figure 1.16b). Several countries have women as authors of over 30% of AI papers on arXiv including Argentina, Canada, Iran, and many European countries (Portugal, Spain, France, Belgium, Italy, Netherlands, Denmark, Ireland, Hungary). In the United States, the share of women authors in AI has decreased slightly over this period.

Percent of papers with at least one female author.

Source: NESTA, arXiv, 2019.

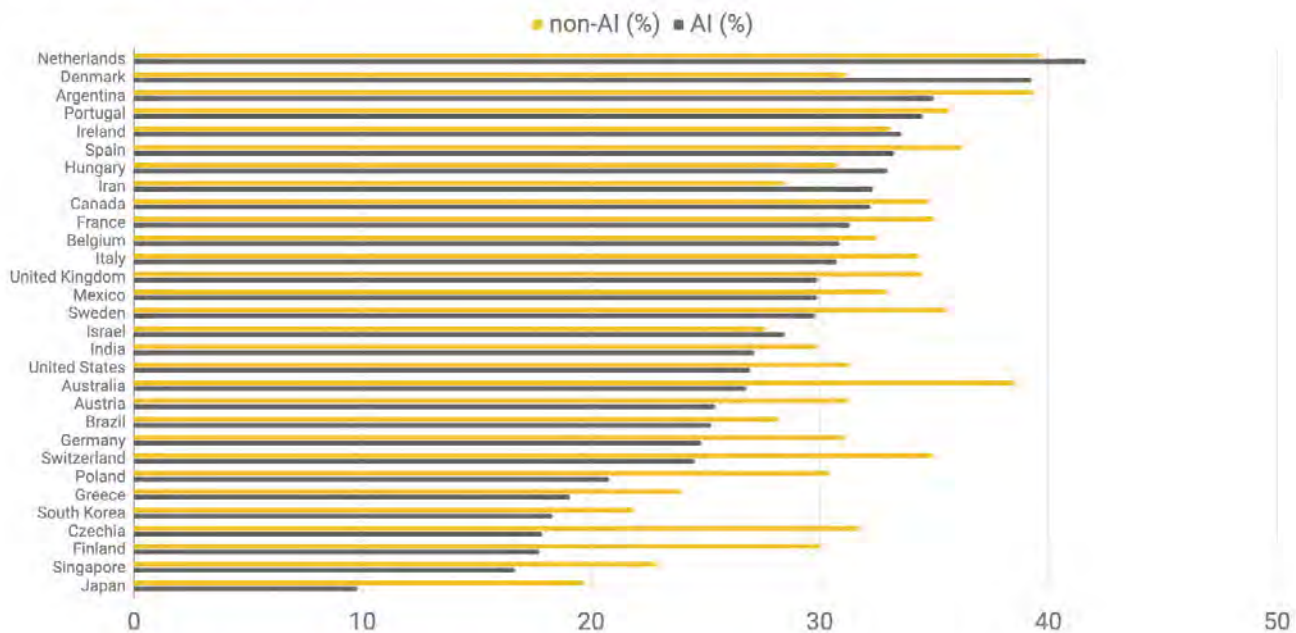


Fig. 1.16a.

"Our findings suggest that both geography 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."

Kostas Stathoulopoulos and Juan Mateos-Garcia, NESTA



Women in AI

Growth in female authorship of AI paper, 2000-18

Source: NESTA, 2019.

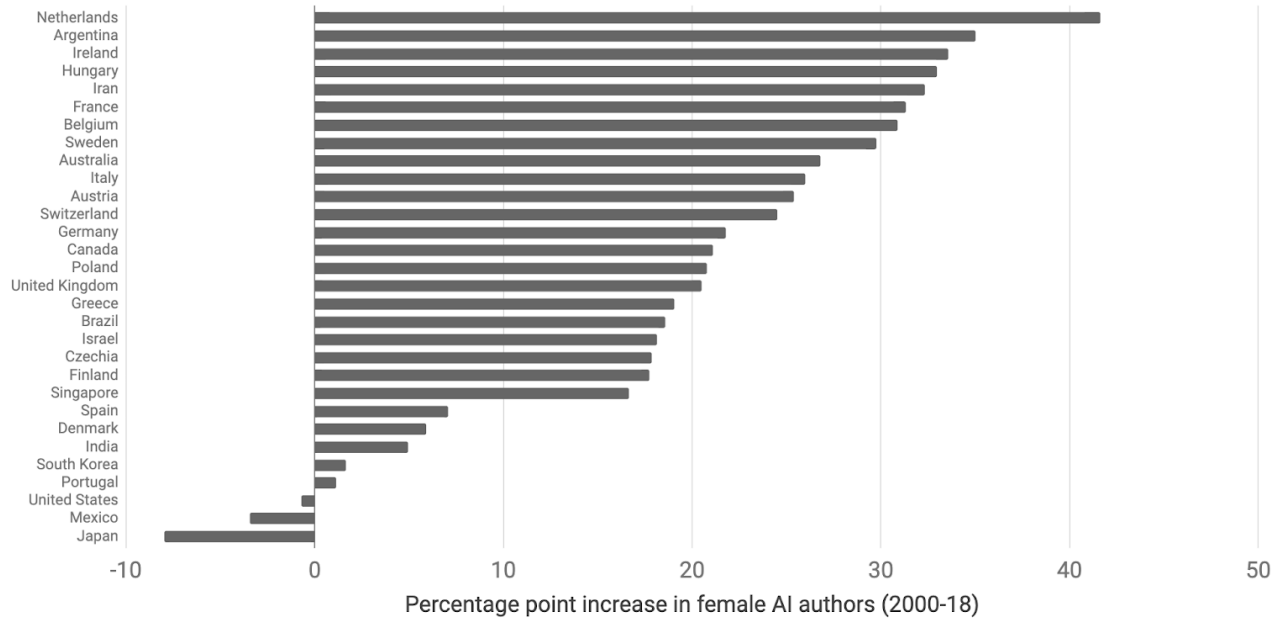


Fig 1.16b.

Many Western European countries as well as Argentina, Canada, and Iran show relatively high presence of women in AI research.



Measurement Questions

Some questions implied by the data in this section include:

- What is the best way to weight the relative importance of paper publications on preprint services like arXiv versus traditional journal publications?
- What tools are available to help us neatly attribute papers to a specific region or originating institution and/or funding source?
- Is it possible to measure and assess the gender of AI researchers without the addition of specific metadata to preprints and published papers?