Artificial Intelligence
Index Report 2022

CHAPTER 1:
Research & Development

CHAPTER 1:

Chapter Preview

ACCESS THE PUBLIC DATA

Overview

Research and development is an integral force that drives the rapid progress of artificial intelligence (AI). Every year, a wide range of academic, industry, government, and civil society experts and organizations contribute to AI R&D via a slew of papers, journal articles, and other AI-related publications, conferences on AI or on particular subtopics like image recognition or natural language processing, international collaboration across borders, and the development of open-source software libraries. These R&D efforts are diverse in focus and geographically dispersed.

Another key feature of AI R&D, making it somewhat distinct from other areas of STEM research, is its openness. Each year, thousands and thousands of AI publications are released in the open source, whether at conferences or on file-sharing websites. Researchers will openly share their findings at conferences; government agencies will fund AI research that ends up in the open source; and developers use open software libraries, freely available to the public, to produce state-of-the-art AI applications. This openness also contributes to the globally interdependent and interconnected nature of modern AI R&D.

This first chapter draws on multiple datasets to analyze key trends in the AI research and development space in 2021. It first looks at AI publications, including conference papers, journal articles, patents, and repositories. It then analyzes AI conference attendance. And finally, it examines AI open-source software libraries used in the R&D process.

CHAPTER HIGHLIGHTS

- Despite rising geopolitical tensions, the United States and China had the greatest number of cross-country collaborations in AI publications from 2010 to 2021, **increasing five times since 2010.** The collaboration between the two countries **produced 2.7 times more publications** than between the United Kingdom and China—the second highest on the list.

- In 2021, China continued to lead the world in the number of AI journal, conference, and repository publications—**63.2% higher than the United States** with all three publication types combined. In the meantime, the United States **held a dominant lead among major AI powers** in the number of AI conference and repository citations.

- From 2010 to 2021, **the collaboration between educational and nonprofit organizations produced the highest number of AI publications**, followed by the collaboration between private companies and educational institutions and between educational and government institutions.

- **The number of AI patents filed in 2021 is more than 30 times higher than in 2015**, showing a compound annual growth rate of 76.9%.

This section draws on data from the Center for Security and Emerging Technology (CSET) at Georgetown University. CSET maintains a merged corpus of scholarly literature that includes Digital Science's Dimensions, Clarivate's Web of Science, Microsoft Academic Graph, China National Knowledge Infrastructure, arXiv, and Papers with Code. In that corpus, CSET applied a classifier to identify English-language publications related to the development or application of AI and ML since 2010.[1]

1.1 PUBLICATIONS[2]

OVERVIEW

The figures below capture the total number of English-language AI publications globally from 2010 to 2021—by type, affiliation, cross-country collaboration, and cross-industry collaboration. The section also breaks down publication and citation data by region for AI journal articles, conference papers, repositories, and patents.

Total Number of AI Publications

Figure 1.1.1 shows the number of AI publications in the world. From 2010 to 2021, the total number of AI publications doubled, growing from 162,444 in 2010 to 334,497 in 2021.

NUMBER of AI PUBLICATIONS in the WORLD, 2010–21
Source: Center for Security and Emerging Technology, 2021 | Chart: 2022 AI Index Report

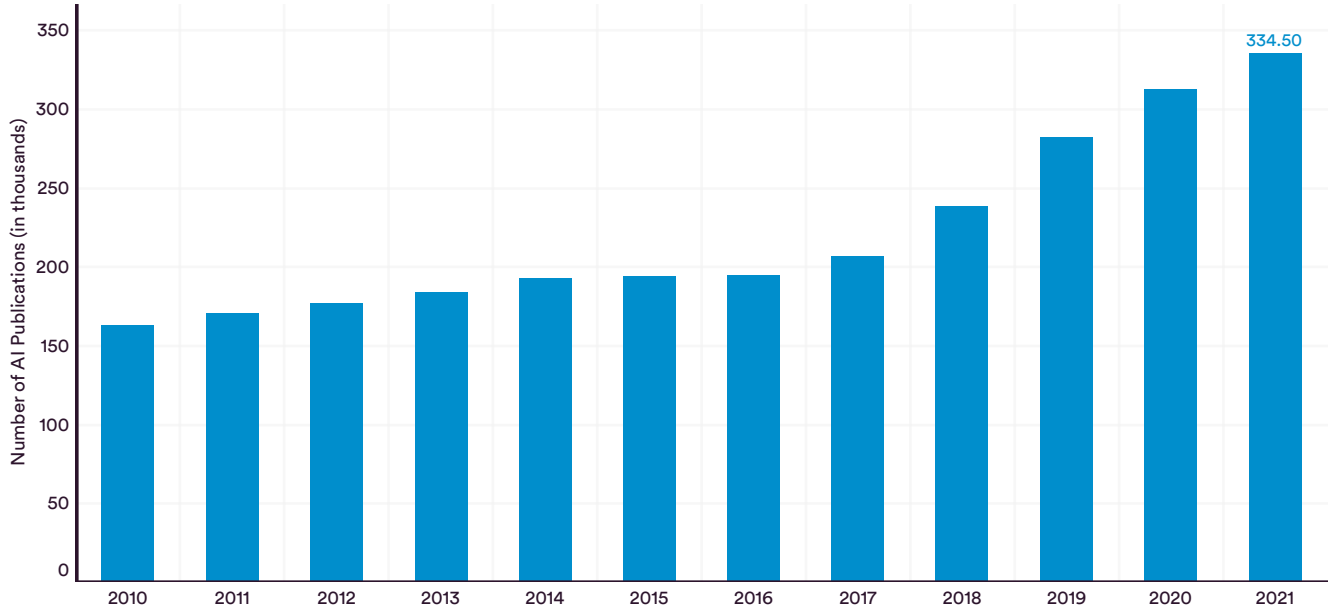

Figure 1.1.1

1 See the Appendix for more information on CSET's methodology. Due to the change in data provider and classification method, the publication trend/data may be different from past reports. For more on the challenge of defining AI and correctly capturing relevant bibliometric data, see the AI Index team's discussion in the paper "Measurement in AI Policy: Opportunities and Challenges."
2 The number of AI publications in 2021 in this section may be lower than the actual count due to the lag in the collection of publication metadata by aforementioned databases.

By Type of Publication

Figure 1.1.2 shows the variation overtime in the types of AI publications released globally. In 2021, 51.5% of all AI documents published were journal articles, 21.5% were conference papers, and 17.0% were from repositories. Books, book chapters, theses, and unknown document types comprised the remaining 10.1% of publications. While journal and repository publications have grown 2.5 and 30 times, respectively, in the past 12 years, the number of conference papers has declined since 2018.

NUMBER of AI PUBLICATIONS by TYPE, 2010–21

Source: Center for Security and Emerging Technology, 2021 | Chart: 2022 AI Index Report

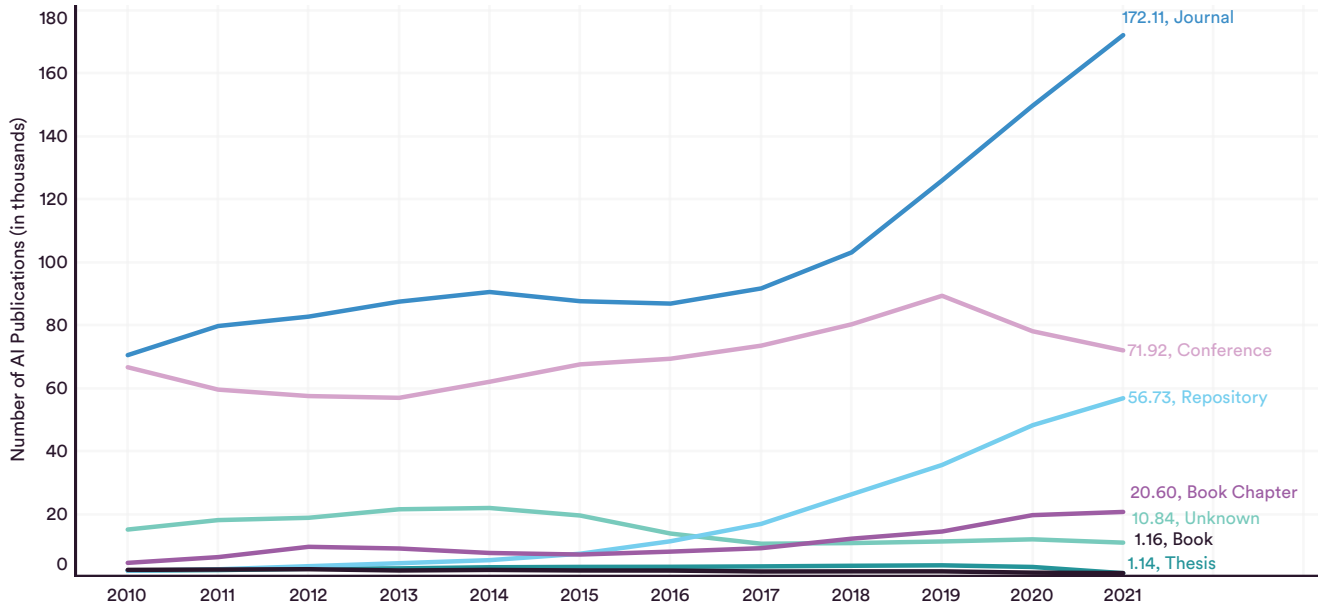

Figure 1.1.2

By Field of Study

Figure 1.1.3 shows that publications in pattern recognition and machine learning have more than doubled since 2015. Other areas strongly influenced by deep learning, such as computer vision, data mining and natural language processing, have shown smaller increases.

NUMBER of AI PUBLICATIONS by FIELD of STUDY (EXCLUDING OTHER AI), 2010–21

Source: Center for Security and Emerging Technology, 2021 | Chart: 2022 AI Index Report

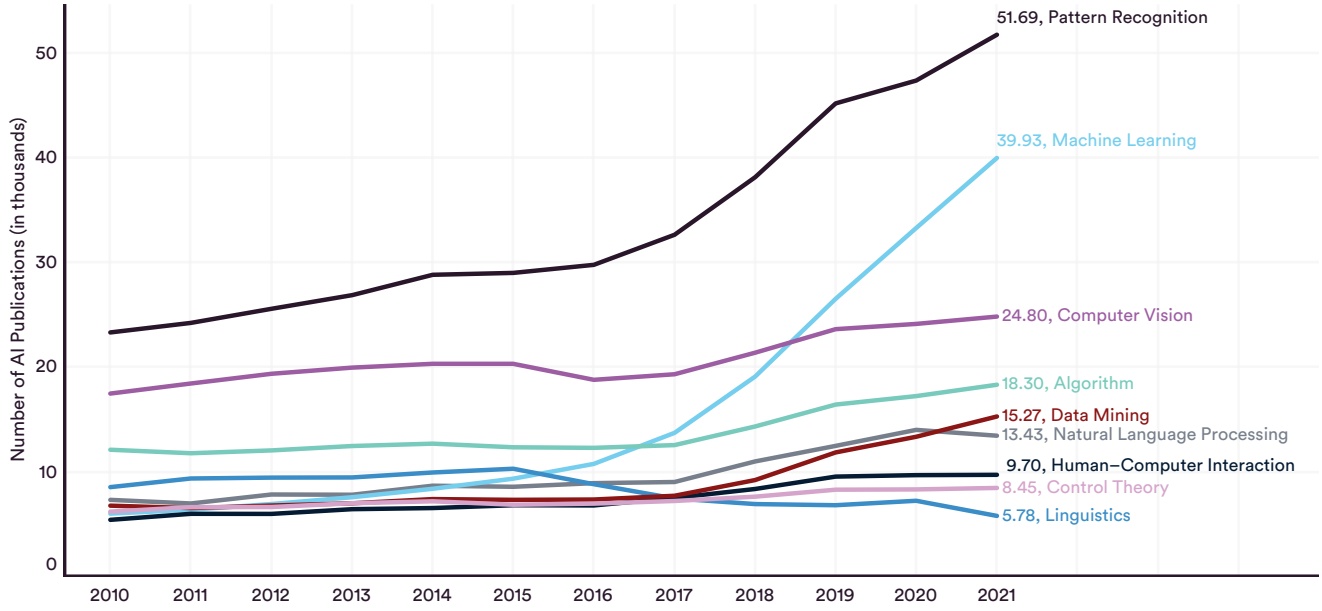

Figure 1.1.3

By Sector

This section shows the number of AI publications affiliated with industry, education, government, and nonprofit in the world (Figure 1.1.4a), China (Figure 1.1.4b), the United States (Figure 1.1.4c), and the European Union plus the United Kingdom (Figure 1.1.4d).[3] The education sector dominates in each of the regions. The level of company participation is highest in the United States, then in the European Union. China is the only area in which the share of education has been rising.

AI PUBLICATIONS (% of TOTAL) by SECTOR, 2010–21
Source: Center for Security and Emerging Technology, 2021 | Chart: 2022 AI Index Report

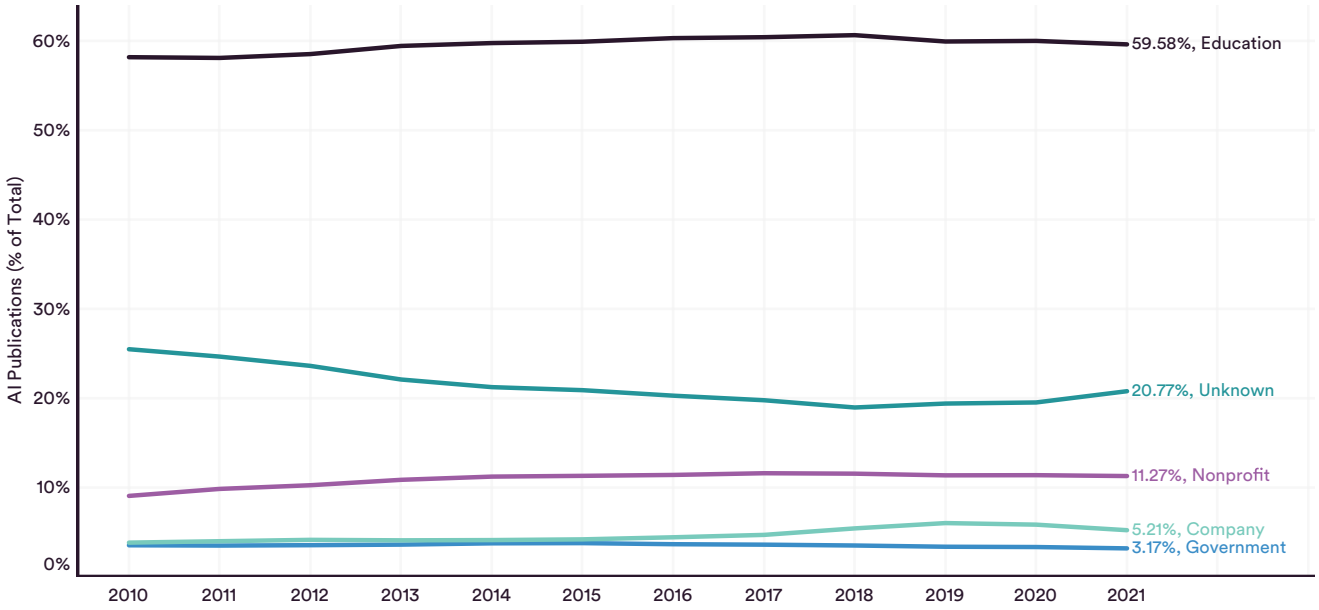

Figure 1.1.4a

AI PUBLICATIONS in UNITED STATES (% of TOTAL) by SECTOR, 2010–21
Source: Center for Security and Emerging Technology, 2021 | Chart: 2022 AI Index Report

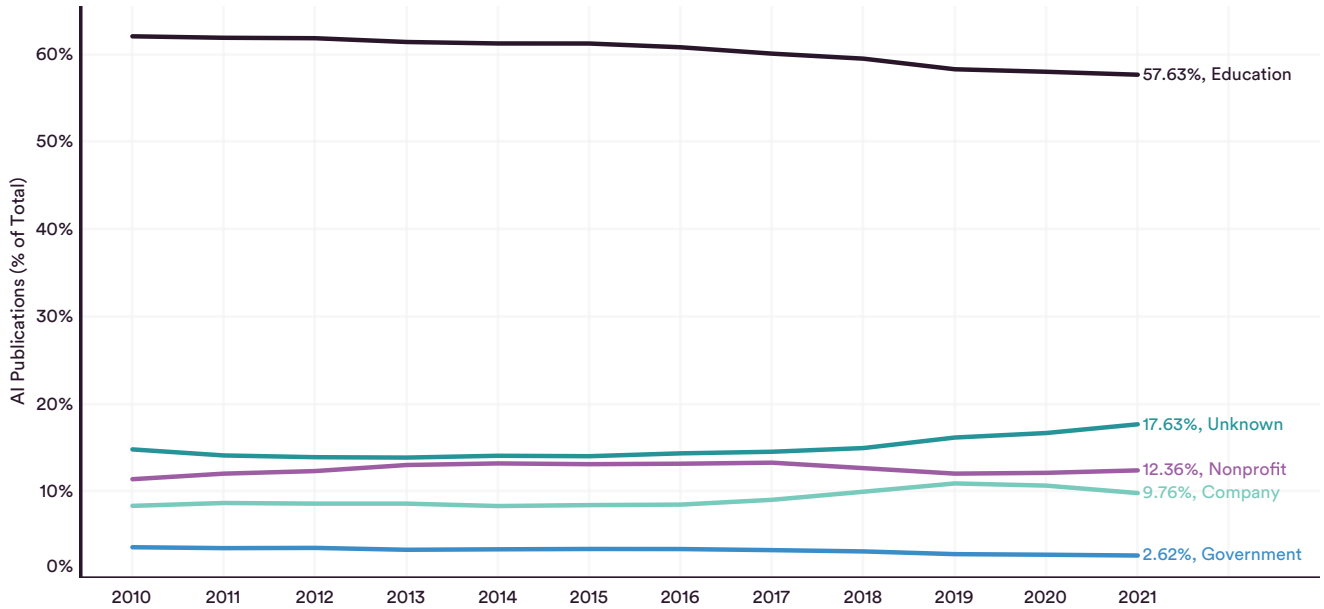

Figure 1.1.4b

3 The categorization is adapted based on the Global Research Identifier Database (GRID). See definitions of each category here. Healthcare, including hospitals and facilities, are included under nonprofit. Publications affiliated with state-sponsored universities are included in the education sector.

AI PUBLICATIONS in CHINA (% of TOTAL) by SECTOR, 2010–21

Source: Center for Security and Emerging Technology, 2021 | Chart: 2022 AI Index Report

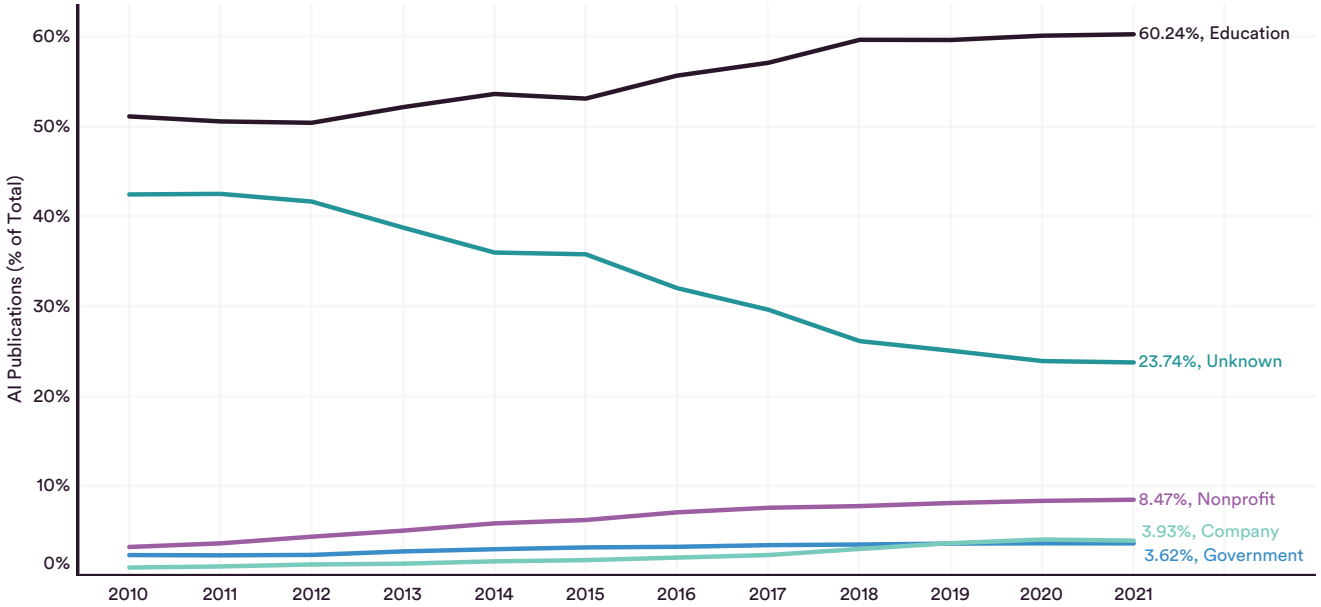

Figure 1.1.4c

AI PUBLICATIONS in EUROPEAN UNION and UNITED KINGDOM (% of TOTAL) by SECTOR, 2010–21

Source: Center for Security and Emerging Technology, 2021 | Chart: 2022 AI Index Report

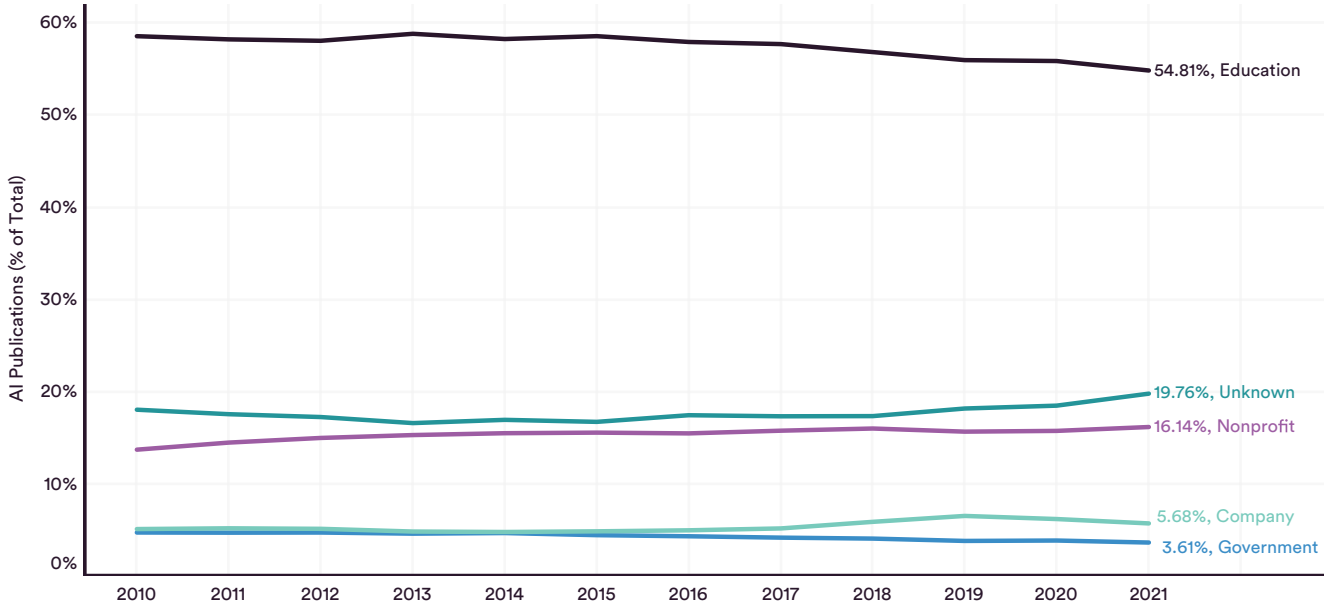

Figure 1.1.4d

Cross-Country Collaboration

Cross-border collaborations between academics, researchers, industry experts, and others are a key component of modern STEM development that accelerate the dissemination of new ideas and the growth of research teams. Figures 1.1.5a and 1.1.5b depict the top cross-country AI collaborations from 2010 to 2021. CSET counted cross-country collaborations as distinct pairs of countries across authors for each publication (e.g., four U.S. and four Chinese-affiliated authors on a single publication are counted as one U.S.-China collaboration; two publications between the same authors counts as two collaborations).

By far, the greatest number of collaborations in the past 12 years took place between the United States and China, increasing five times since 2010. The next largest set of collaborations is between the United Kingdom and both the United States and China, which have increased more than three times since 2010. In 2021, the number of collaborations between the United States and China was 2.7 times greater than between the United Kingdom and China.

By far, the greatest number of collaborations in the past 12 years took place between the United States and China, increasing five times since 2010.

UNITED STATES and CHINA COLLABORATIONS in AI PUBLICATIONS, 2010–21
Source: Center for Security and Emerging Technology, 2021 | Chart: 2022 AI Index Report

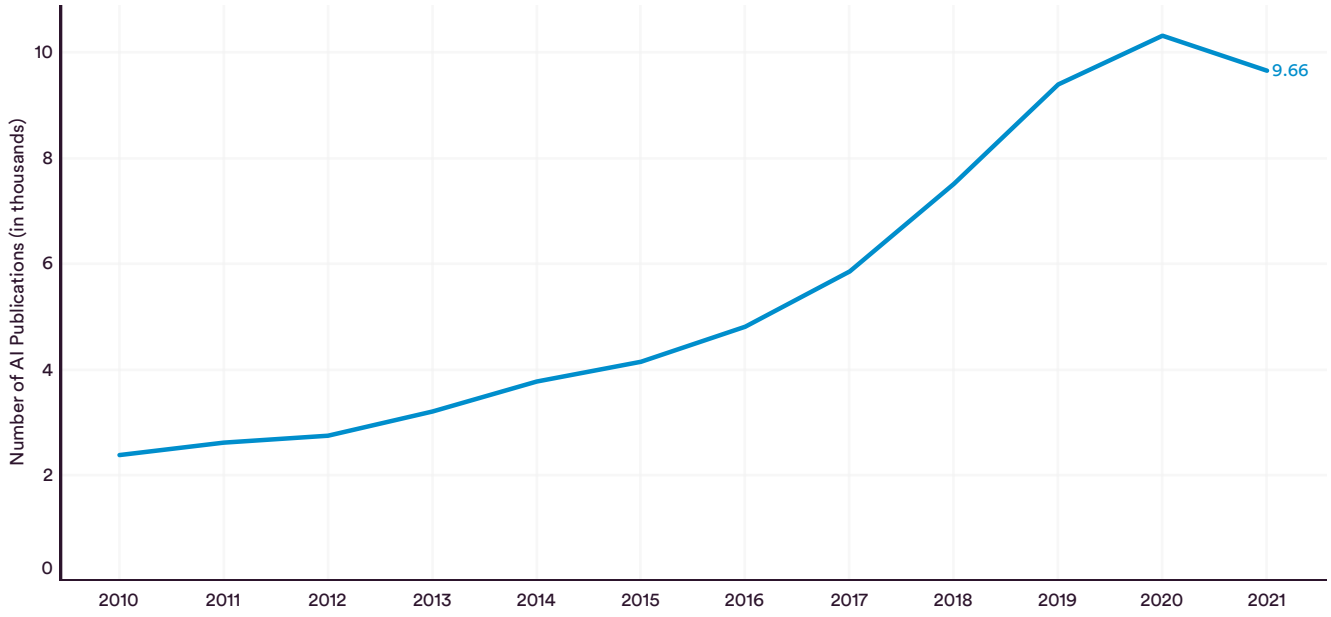

Figure 1.1.5a

CROSS-COUNTRY COLLABORATIONS in AI PUBLICATIONS (EXCLUDING U.S. and CHINA), 2010–21
Source: Center for Security and Emerging Technology, 2021 | Chart: 2022 AI Index Report

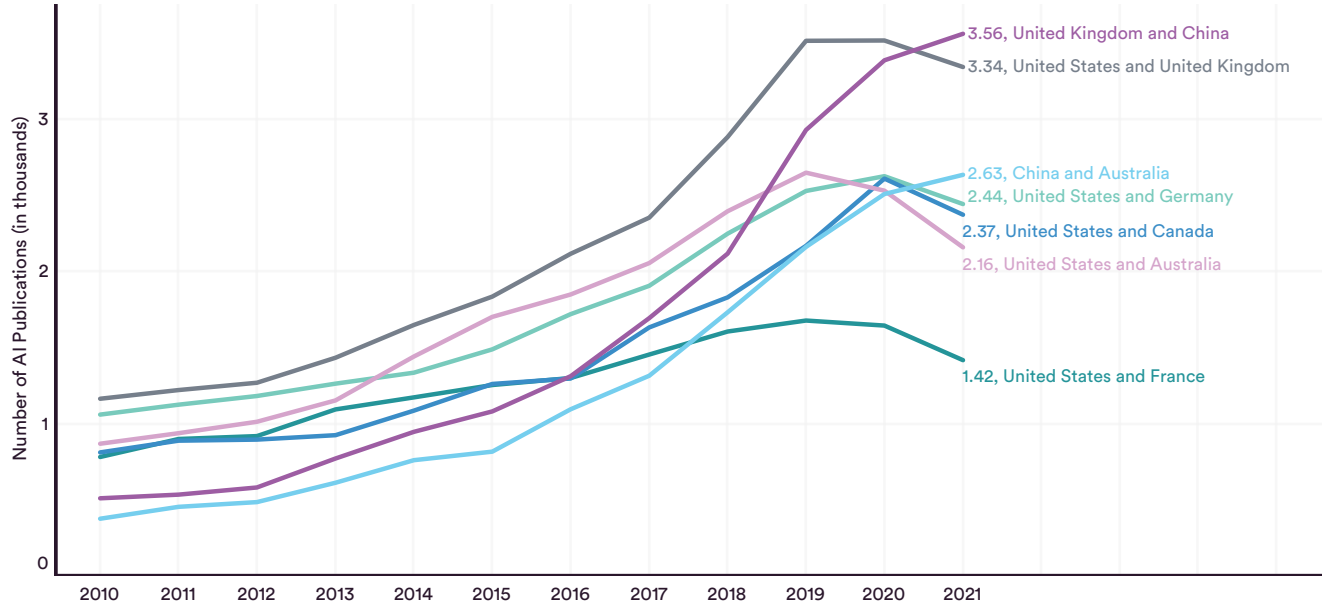

Figure 1.1.5b

Cross-Sector Collaboration

The increase in AI research outside of universities has brought collaborations between universities and other industries. Figure 1.1.6 shows that in 2021, educational institutions and nonprofits had the greatest number of collaborations (29,839), followed by companies and educational institutions (11,576), and governments and educational institutions (8,087). There were 2.5 times as many collaborations between educational institutions and nonprofits in 2021 as between educational institutions and companies.

CROSS-SECTOR COLLABORATIONS in AI PUBLICATIONS, 2010–21
Source: Center for Security and Emerging Technology, 2021 | Chart: 2022 AI Index Report

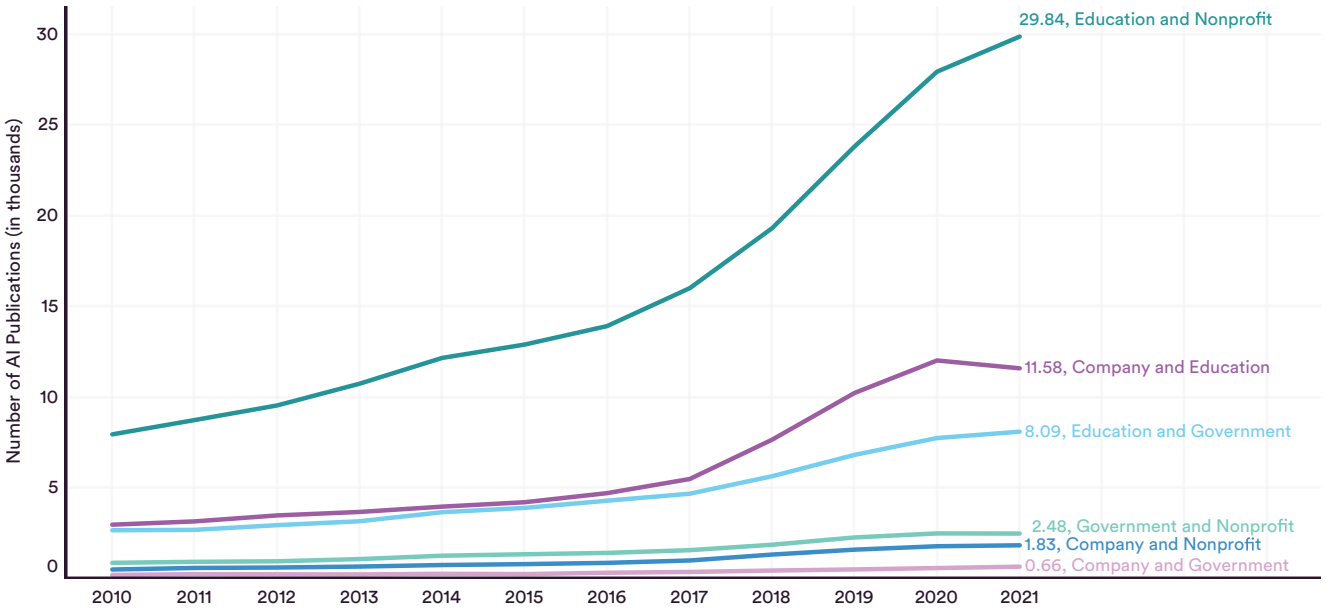

Figure 1.1.6

AI JOURNAL PUBLICATIONS

Overview

After growing only slightly from 2010 to 2015, the number of AI journal publications grew almost 2.5 times since 2015 (Figure 1.1.7). As a percentage of all journal publications, as captured in Figure 1.1.8, AI journal publications in 2021 were about 2.5% of all publications, compared to 1.5% in 2010.

NUMBER of AI JOURNAL PUBLICATIONS, 2010–21
Source: Center for Security and Emerging Technology, 2021 | Chart: 2022 AI Index Report

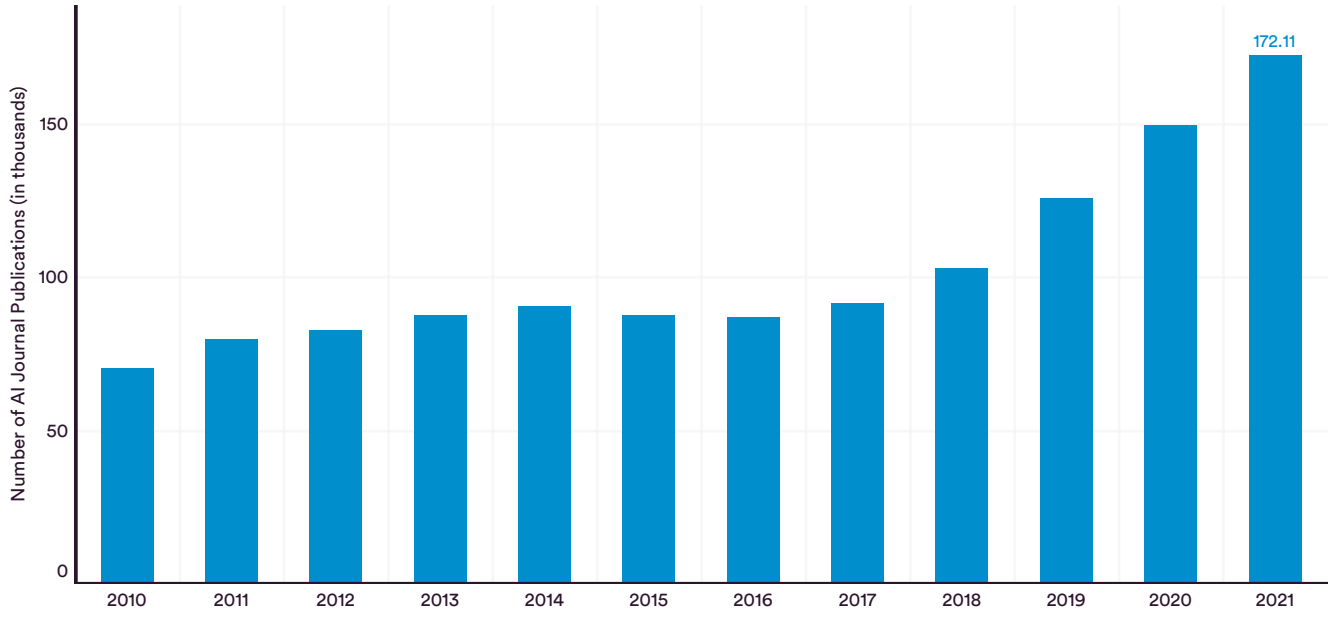

Figure 1.1.7

AI JOURNAL PUBLICATIONS (% of TOTAL JOURNAL PUBLICATIONS), 2010–21
Source: Center for Security and Emerging Technology, 2021 | Chart: 2022 AI Index Report

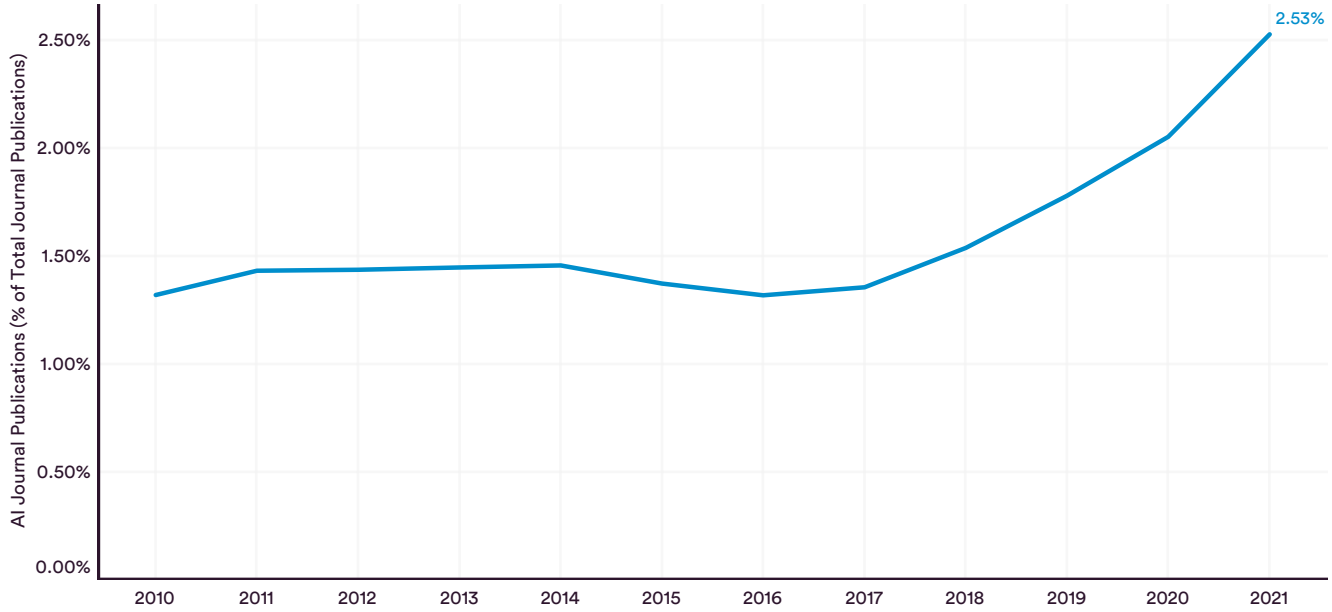

Figure 1.1.8

By Region[4]

Figure 1.1.9 shows the share of AI journal publications by region between 2010 and 2021. In 2021, East Asia and Pacific leads with 42.9%, followed by Europe and Central Asia (22.7%) and North America (15.6%). In addition, South Asia and the Middle East and North Africa saw the most significant growth as their number of AI journal publications grew around 12 and 7 times, respectively, in the last 12 years.

In addition, South Asia and the Middle East and North Africa saw the most significant growth as their number of AI journal publications grew around 12 and 7 times, respectively, in the last 12 years.

AI JOURNAL PUBLICATIONS (% of WORLD TOTAL) by REGION, 2010–21

Source: Center for Security and Emerging Technology, 2021 | Chart: 2022 AI Index Report

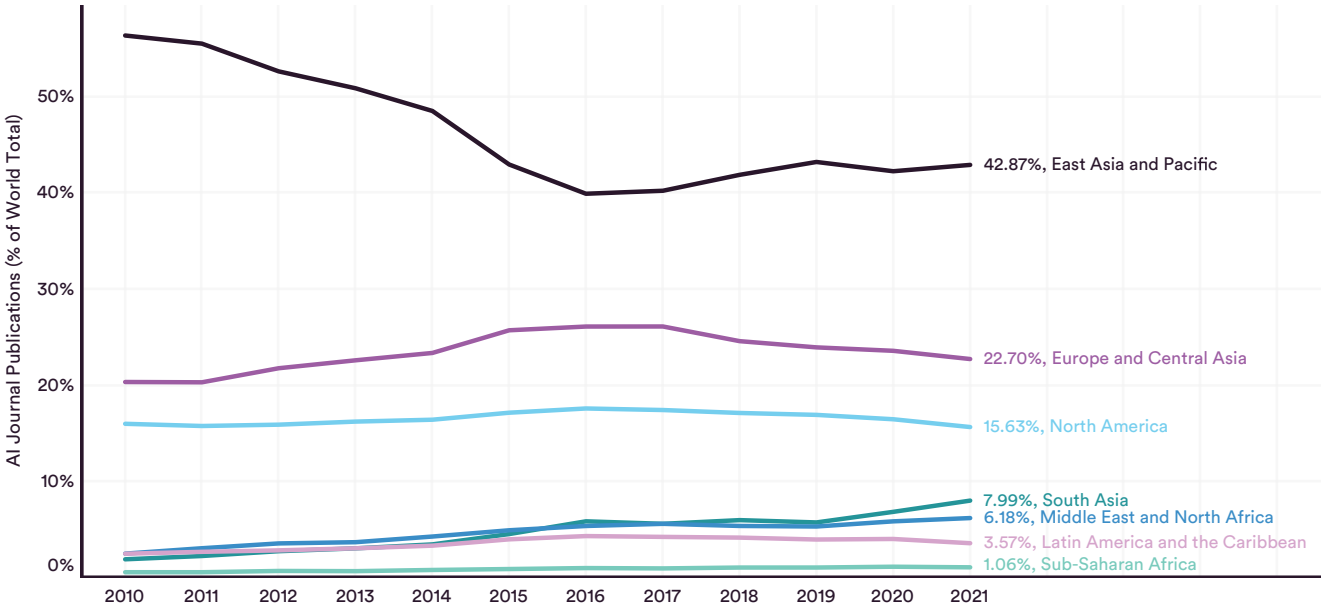

Figure 1.1.9

4 Regions in this chapter are classified according to the World Bank analytical grouping.

By Geographic Area[5]

Figure 1.1.10 breaks down the share of AI journal publications over the past 12 years by three major AI powers. China has remained the leader throughout, with 31.0% in 2021, followed by the European Union plus the United Kingdom at 19.1% and the United States at 13.7%

AI JOURNAL PUBLICATIONS (% of WORLD TOTAL) by GEOGRAPHIC AREA, 2010–21

Source: Center for Security and Emerging Technology, 2021 | Chart: 2022 AI Index Report

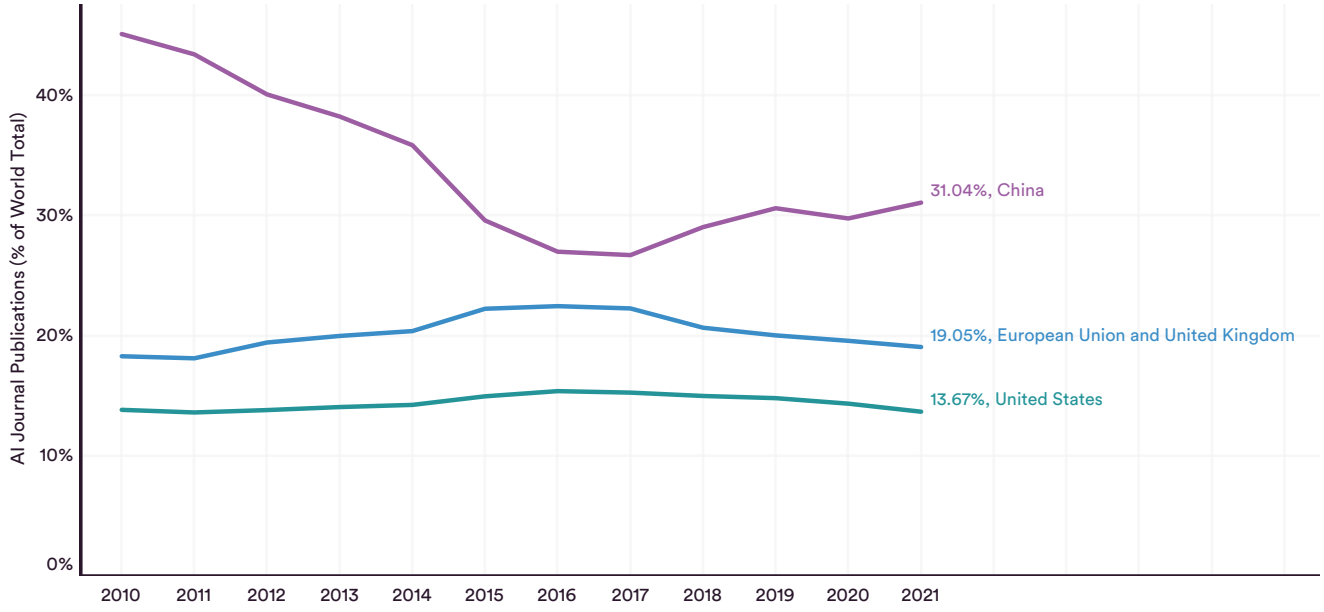

Figure 1.1.10

5 Geographic areas in this chapter combine the number of publications between the European Union and the United Kingdom to reflect the historically strong association between them with regards to research collaboration.

Citations

On the number of citations of AI journal publications, China's share has gradually increased while those of the European Union plus the United Kingdom and the United States have decreased. The three geographic areas combined accounted for more than 66% of the total citations in the world.

The three geographic areas combined accounted for more than 66% of the total citations in the world.

AI JOURNAL CITATIONS (% of WORLD TOTAL) by GEOGRAPHIC AREA, 2010–21

Source: Center for Security and Emerging Technology, 2021 | Chart: 2022 AI Index Report

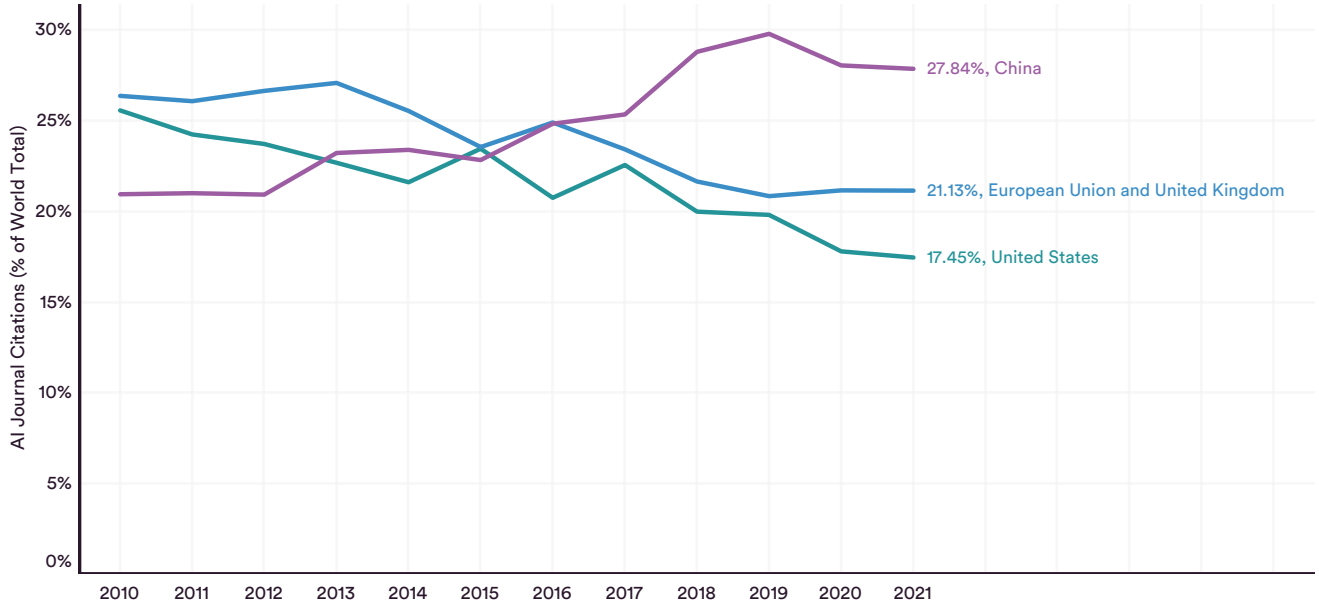

Figure 1.1.11

AI CONFERENCE PUBLICATIONS

Overview

The number of AI conference publications peaked in 2019, and fell about 19.4% below the peak in 2021 (Figure 1.1.12). Despite the decline in the total numbers, however, the share of AI conference publications among total conference publications in the world has increased by more than five percentage points since 2010 (Figure 1.1.13).

NUMBER of AI CONFERENCE PUBLICATIONS, 2010–21

Source: Center for Security and Emerging Technology, 2021 | Chart: 2022 AI Index Report

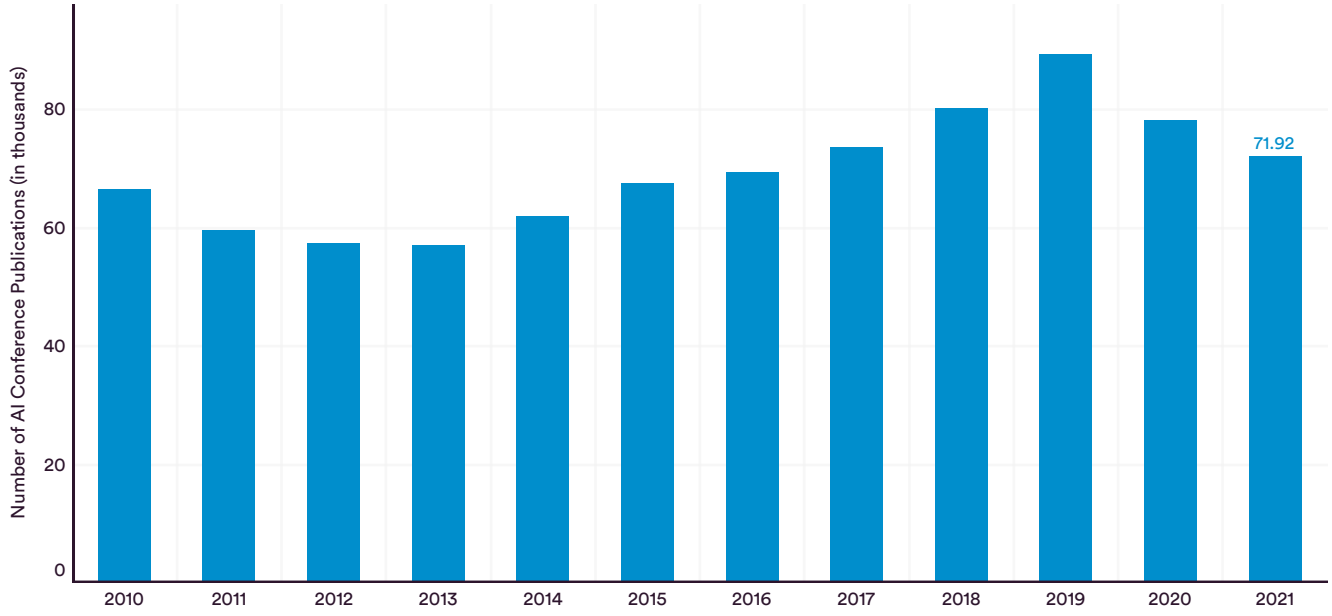

Figure 1.1.12

AI CONFERENCE PUBLICATIONS (% of TOTAL CONFERENCE PUBLICATIONS), 2010–21

Source: Center for Security and Emerging Technology, 2021 | Chart: 2022 AI Index Report

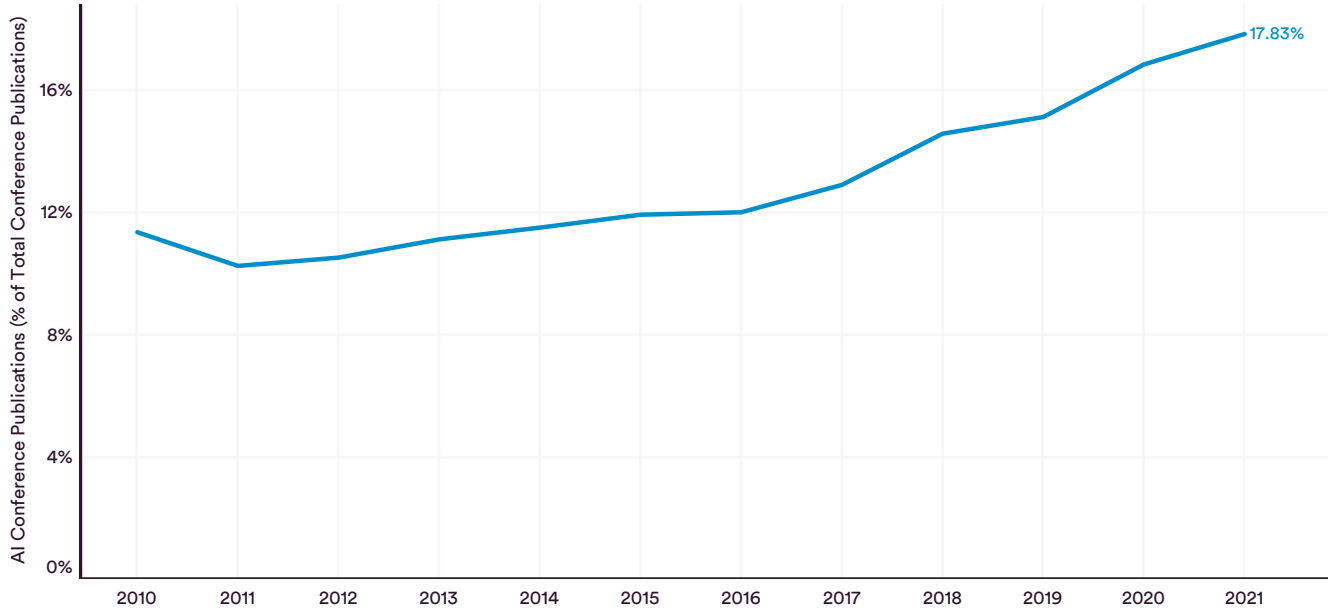

Figure 1.1.13

By Region

Figure 1.1.14 shows the number of AI conference publications by region. Similar to the trend in AI journal publication, East Asia and Pacific, Europe and Central Asia, and North America account for the world's highest numbers of AI conference publications. Specifically, the share represented by East Asia and Pacific continues to rise since taking the lead in 2014, accounting for 40.4% in 2021, followed by Europe and Central Asia (23.0%) and North America (19.0%). The percentage of AI conference publications in South Asia saw a noticeable rise in the past 12 years, from 4.0% in 2010 to 10.4% in 2021.

AI CONFERENCE PUBLICATIONS (% of WORLD TOTAL) by REGION, 2010–21
Source: Center for Security and Emerging Technology, 2021 | Chart: 2022 AI Index Report

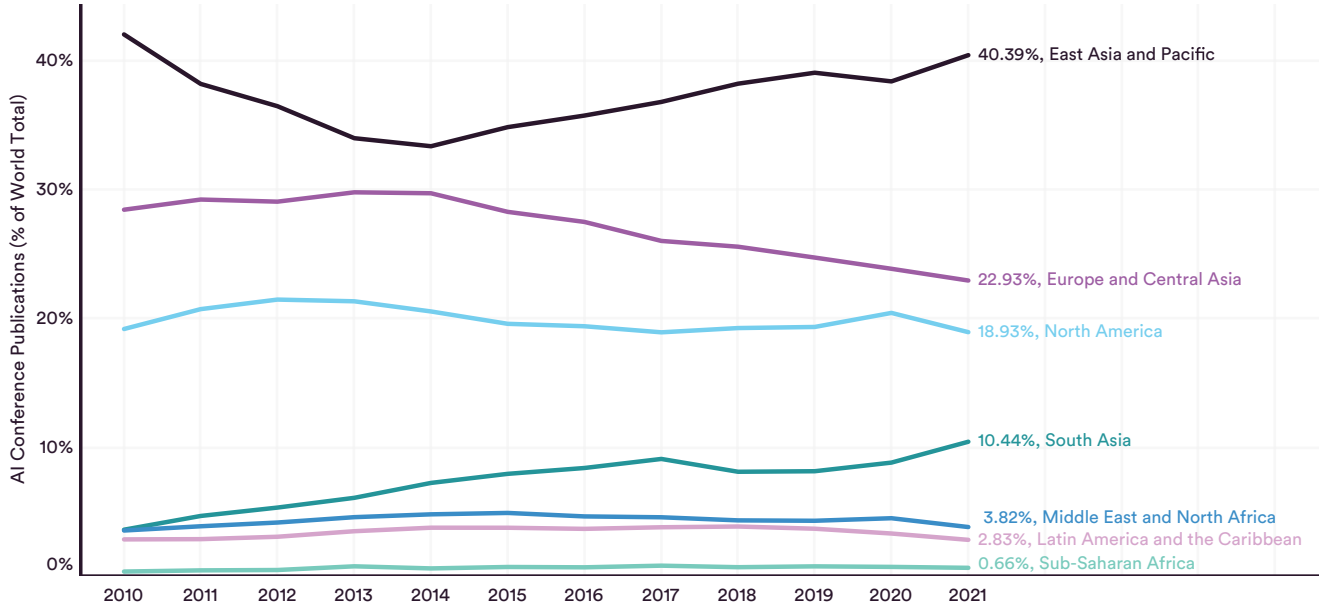

Figure 1.1.14

By Geographic Area

In 2021, China produced the greatest share of the world's AI conference publications at 27.6%, opening an even greater lead than in 2020, while the European Union plus the United Kingdom followed at 19.0% and the United States came in third at 16.9% (Figure 1.1.15).

AI CONFERENCE PUBLICATIONS (% of WORLD TOTAL) by GEOGRAPHIC AREA, 2010–21

Source: Center for Security and Emerging Technology, 2021 | Chart: 2022 AI Index Report

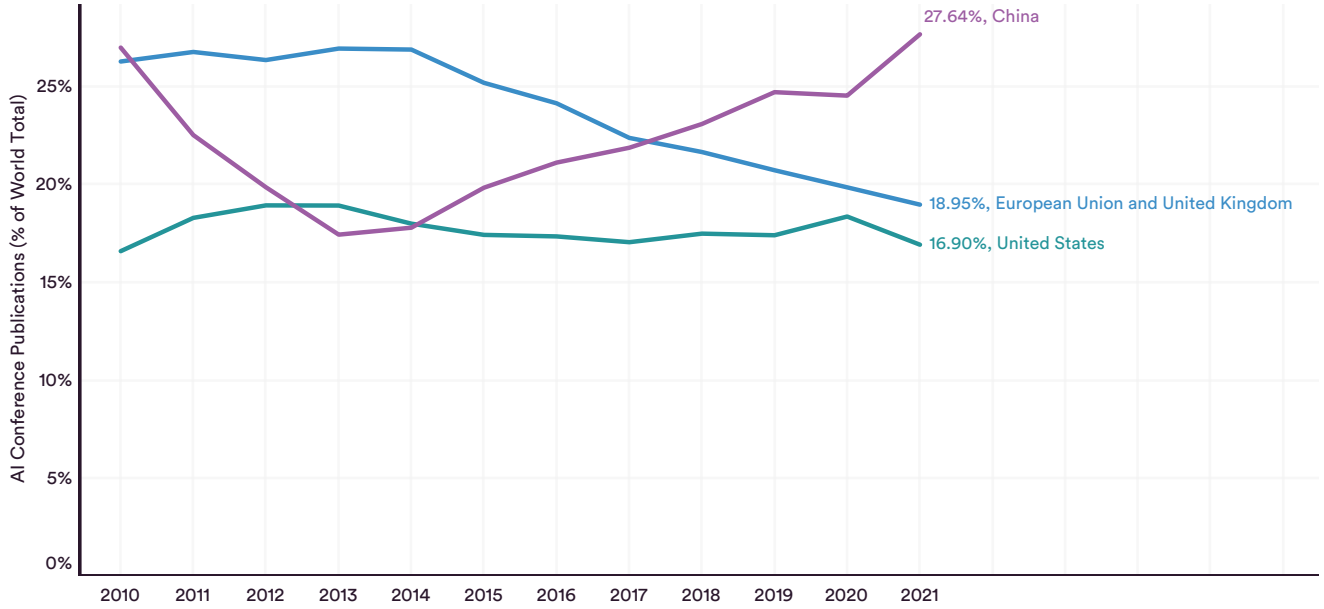

Figure 1.1.15

Citations

Despite China producing the most AI conference publications in 2021, Figure 1.1.16 shows that the United States led among the major powers with respect to the number of AI conference citations, with 29.5% in 2021, followed by the European Union plus the United Kingdom (23.3%) and China (15.3%).

AI CONFERENCE CITATIONS (% of WORLD TOTAL) by GEOGRAPHIC AREA, 2010–21

Source: Center for Security and Emerging Technology, 2021 | Chart: 2022 AI Index Report

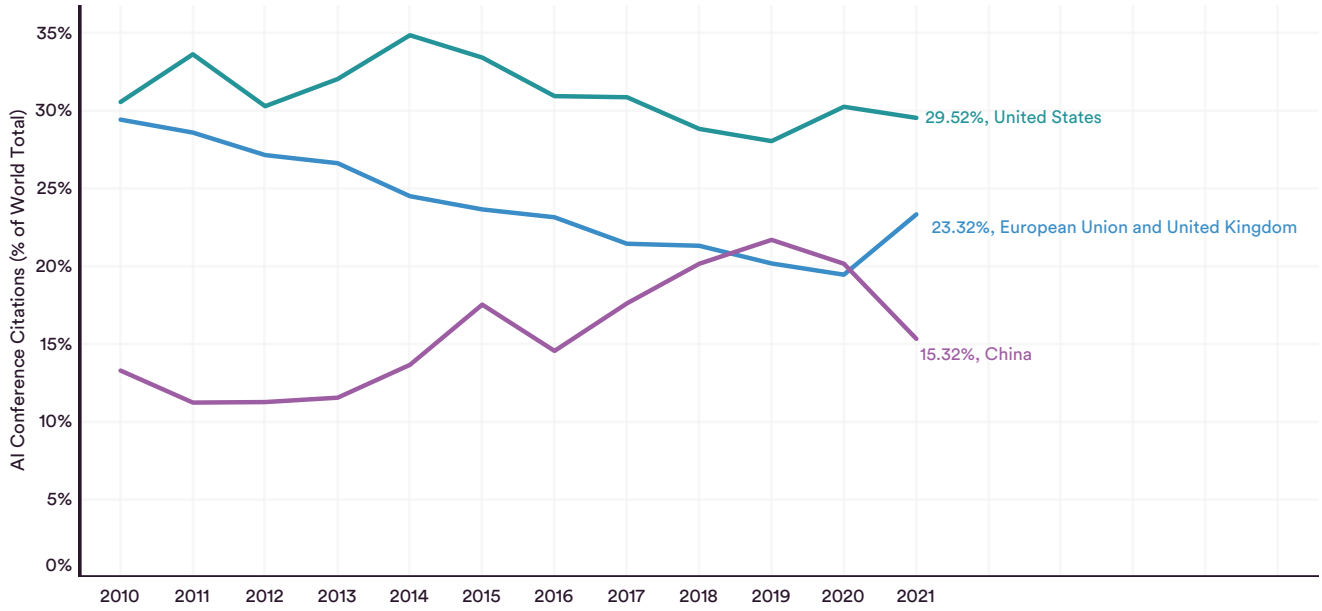

Figure 1.1.16

AI REPOSITORIES

Overview

Publishing pre-peer-reviewed papers on repositories of electronic preprints (such as arXiv and SSRN) has become a popular way among AI researchers to disseminate their work outside traditional avenues for publications. Those repositories allow researchers to share their findings before submitting them to journals and conferences, which greatly accelerates the cycle of information discovery. The number of AI repository publications grew almost 30 times in the past 12 years (Figure 1.1.17), now account for 15.3% of all repository publications (Figure 1.1.18).

NUMBER of AI REPOSITORY PUBLICATIONS, 2010–21
Source: Center for Security and Emerging Technology, 2021 | Chart: 2022 AI Index Report

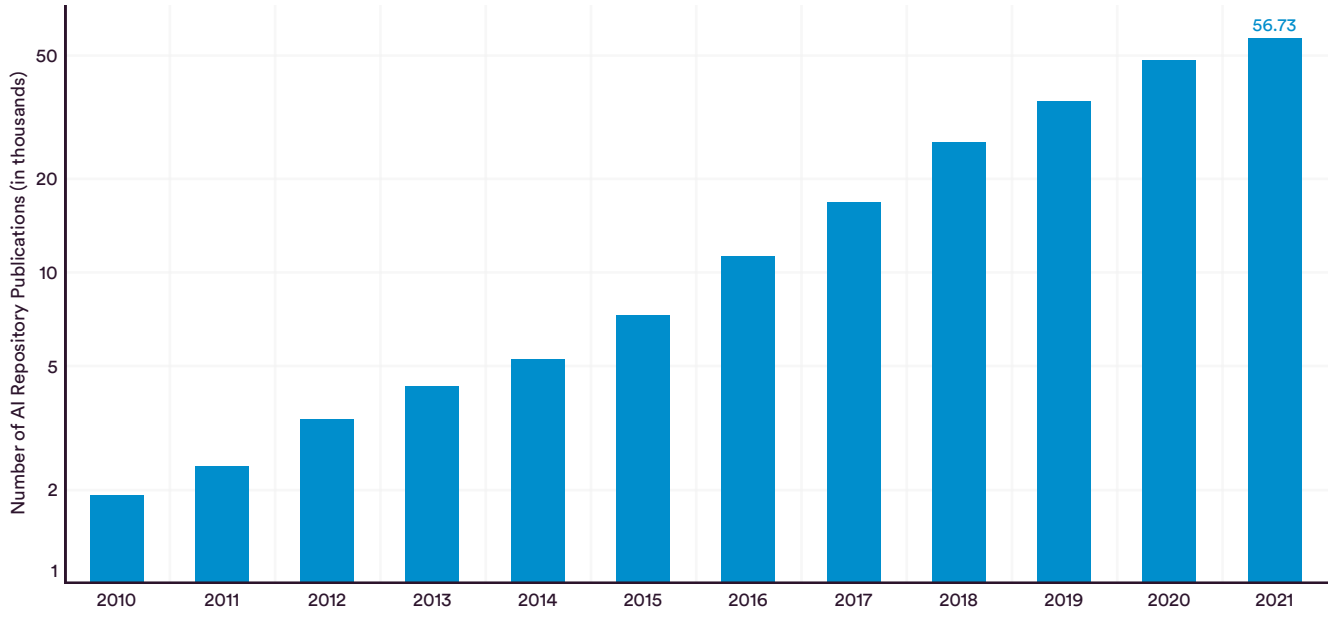

Figure 1.1.17

AI REPOSITORY PUBLICATIONS (% of TOTAL REPOSITORY PUBLICATIONS), 2010–21
Source: Center for Security and Emerging Technology, 2021 | Chart: 2022 AI Index Report

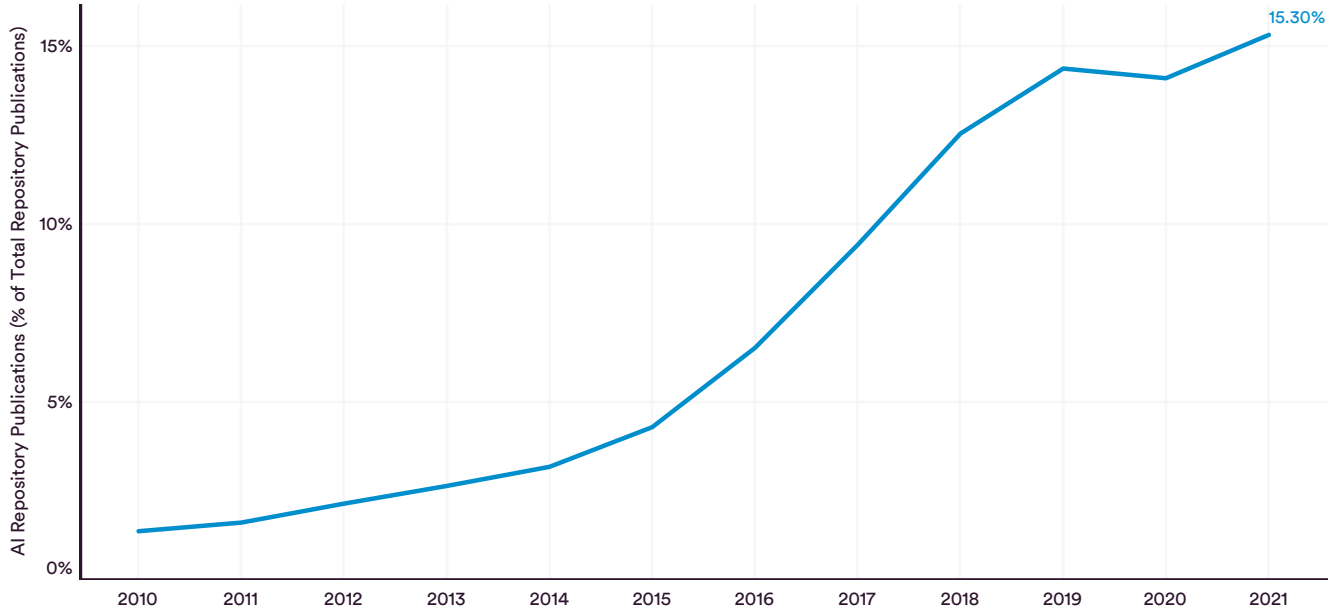

Figure 1.1.18

By Region

The analysis by region in Figure 1.1.19 shows that North America has maintained a steady lead in the share of AI repository publications in the world since 2014 while that of Europe and Central Asia has declined. Since 2013, the share of East Asia and Pacific has grown significantly.

AI REPOSITORY PUBLICATIONS (% of WORLD TOTAL) by REGION, 2010–21

Source: Center for Security and Emerging Technology, 2021 | Chart: 2022 AI Index Report

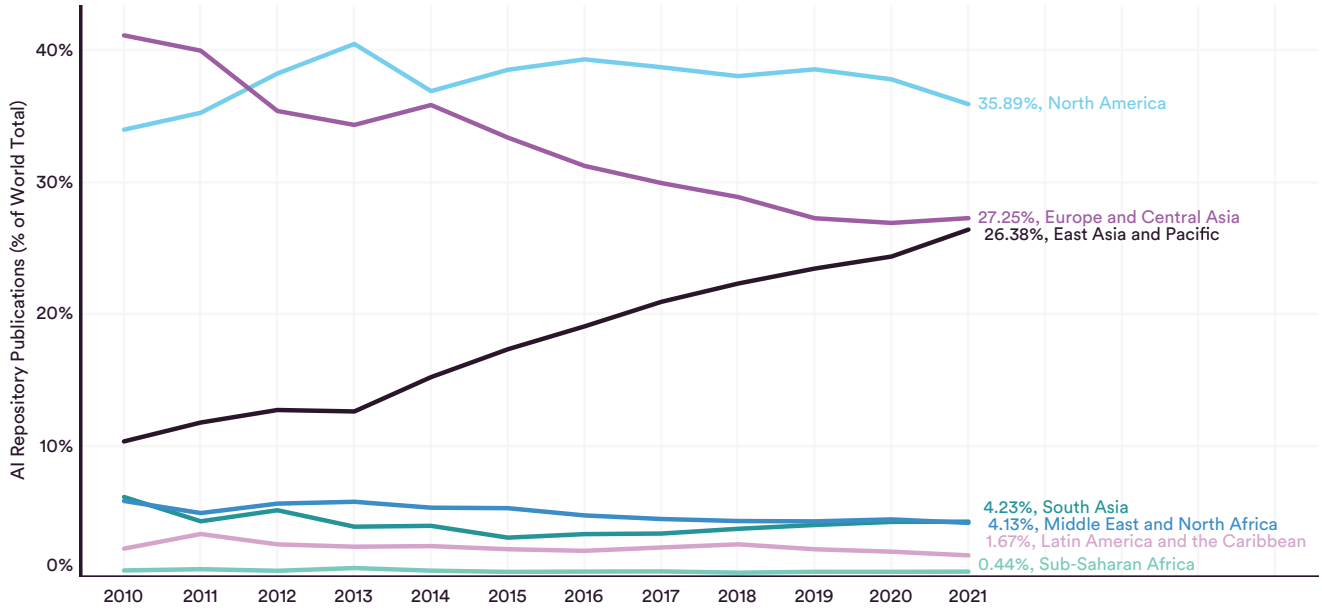

Figure 1.1.19

By Geographic Area

While the United States has held the lead in the percentage of AI repository publications in the world since 2011, China is catching up while the European Union plus the United Kingdom's share continues to drop (Figure 1.1.20). In 2021, the United States accounted for 32.5% of the world's AI repository publications—a higher percentage compared to journal and conference publications, followed by the European Union plus the United Kingdom (23.9%) and China (16.6%).

AI REPOSITORY PUBLICATIONS (% of WORLD TOTAL) by GEOGRAPHIC AREA, 2010–21

Source: Center for Security and Emerging Technology, 2021 | Chart: 2022 AI Index Report

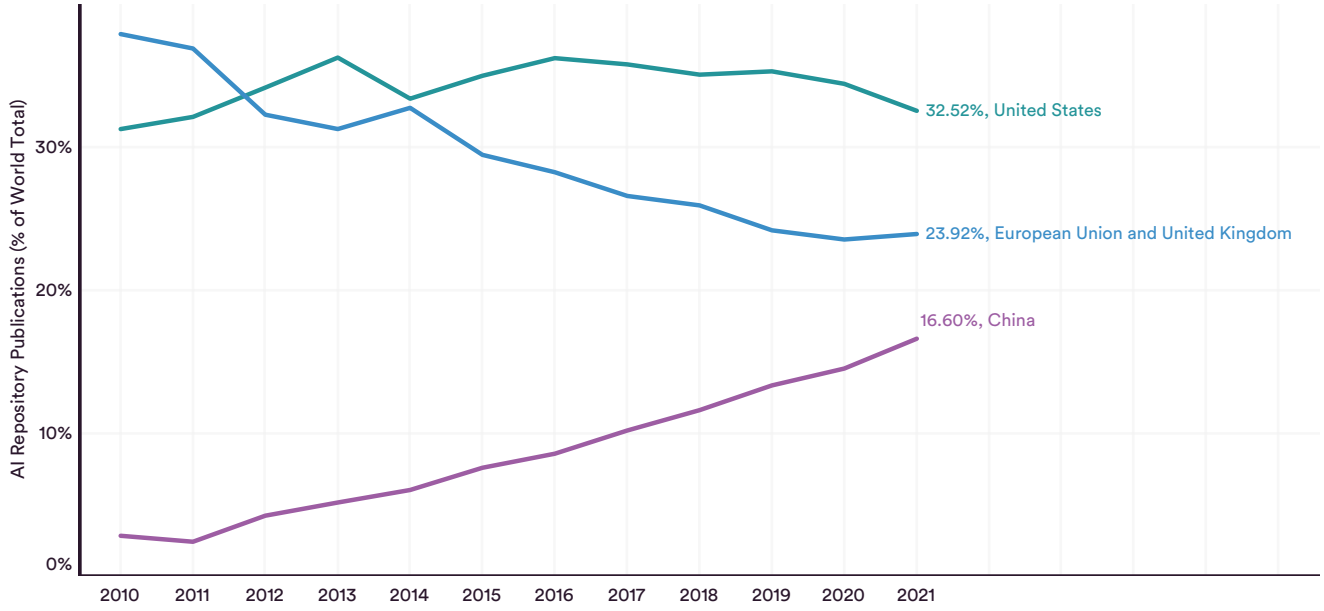

Figure 1.1.20

Citations

On the citations of AI repository publications, Figure 1.1.21 shows that the United States tops the list with 38.6% of overall citations in 2021, establishing a dominant lead over the European Union plus the United Kingdom (20.1%) and China (16.4%).

AI REPOSITORY CITATIONS (% of WORLD TOTAL) by GEOGRAPHIC AREA, 2010–21

Source: Center for Security and Emerging Technology, 2021 | Chart: 2022 AI Index Report

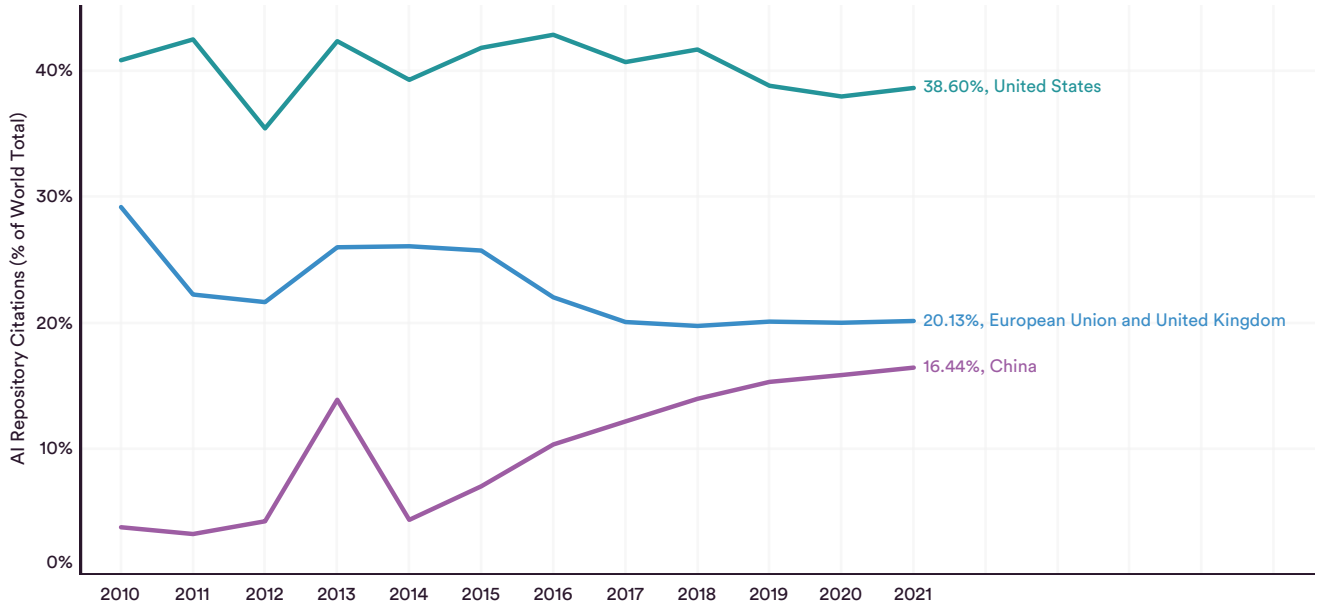

Figure 1.1.21

AI PATENTS

This section draws on data from CSET and 1790 Analytics on patents relevant to AI development and applications—indicated by Cooperative Patent Classification (CPC)/International Patent Classification (IPC) codes and keywords. Patents were grouped by country and year and then counted at the "patent family" level, before CSET extracted year values from the most recent publication date within a family.

Overview

Figure 1.1.22 captures the number of AI patents filed from 2010 to 2021. The number of patents filed in 2021 is more than 30 times higher than in 2015, showing a compound annual growth rate of 76.9%.

The number of patents filed in 2021 is more than 30 times higher than in 2015, showing a compound annual growth rate of 76.9%.

NUMBER of AI PATENT FILINGS, 2010–21
Source: Center for Security and Emerging Technology, 2021 | Chart: 2022 AI Index Report

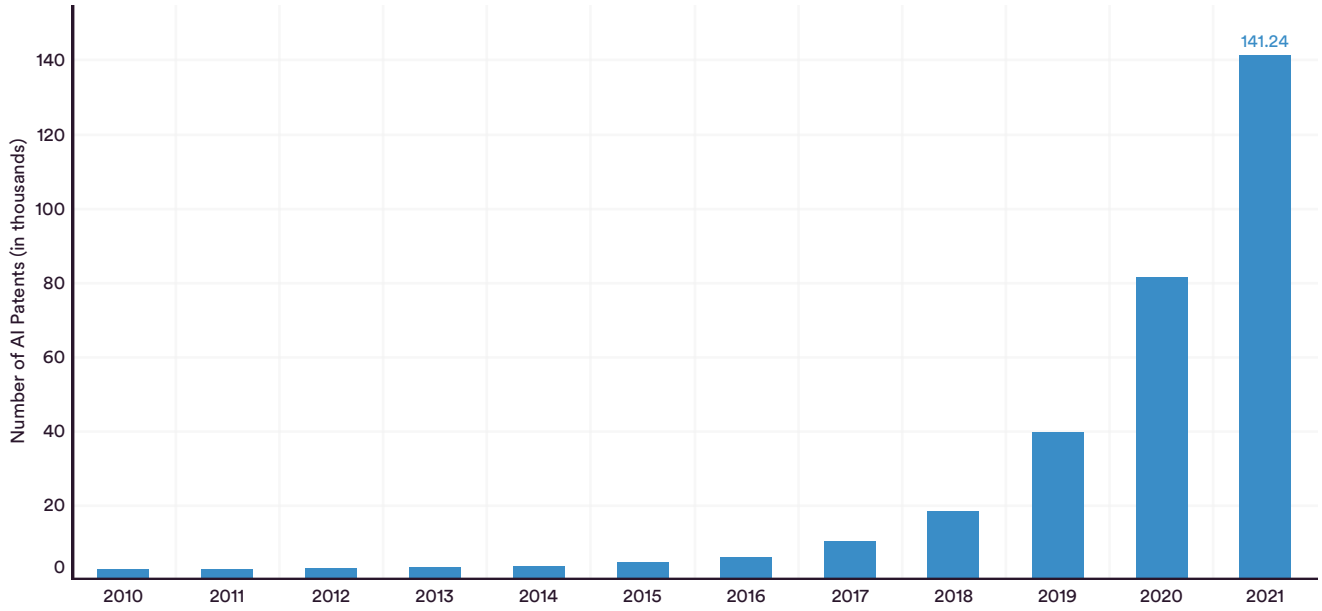

Figure 1.1.22

By Region and Application Status

Figure 1.1.23a breaks down AI patent filings by region. The share of East Asia and Pacific took off in 2014 and led the rest of the world in 2021 with 62.1% of all patent applications, followed by North America and Europe and Central Asia. In terms of granted patents in those regions, North America leads with 57.0%, followed by East Asia and Pacific (31.0%), and Europe and Central Asia (11.3%) (Figure 1.1.23b). The other regions combine to make up roughly 1% of world patents (Figure 1.1.23c).

AI PATENT FILINGS (% of WORLD TOTAL) by REGION, 2010–21

Source: Center for Security and Emerging Technology, 2021 | Chart: 2022 AI Index Report

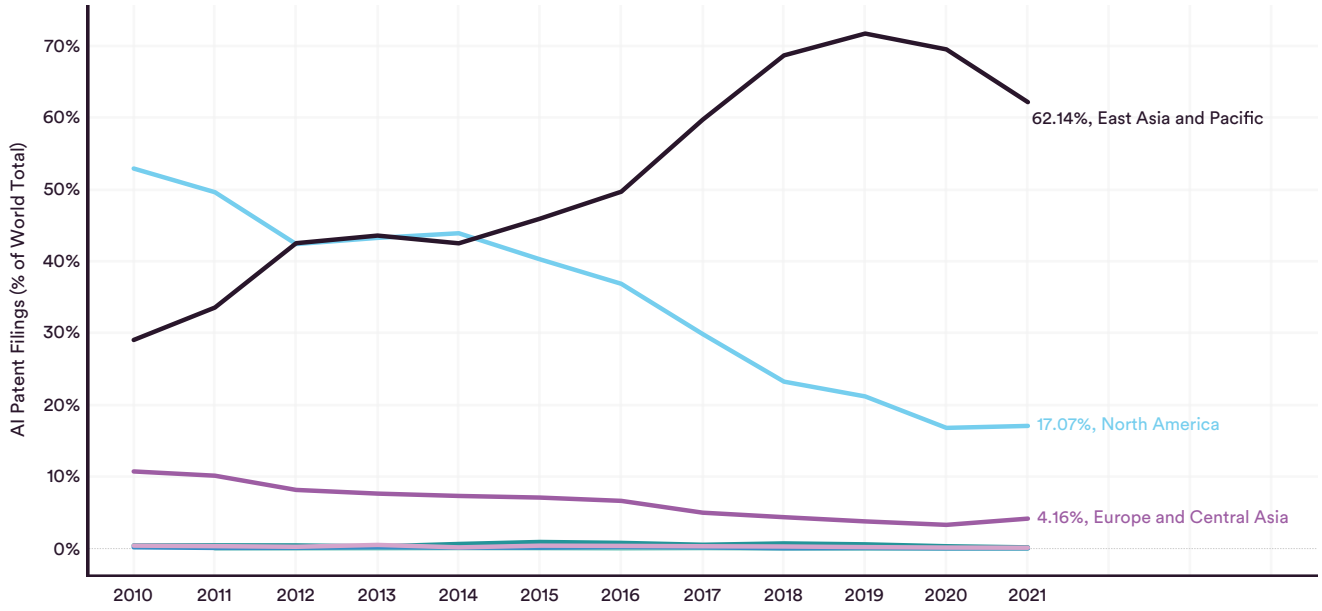

Figure 1.1.23a

GRANTED AI PATENTS (% of WORLD TOTAL) by REGION, 2010–21 (Part 1)

Source: Center for Security and Emerging Technology, 2021 | Chart: 2022 AI Index Report

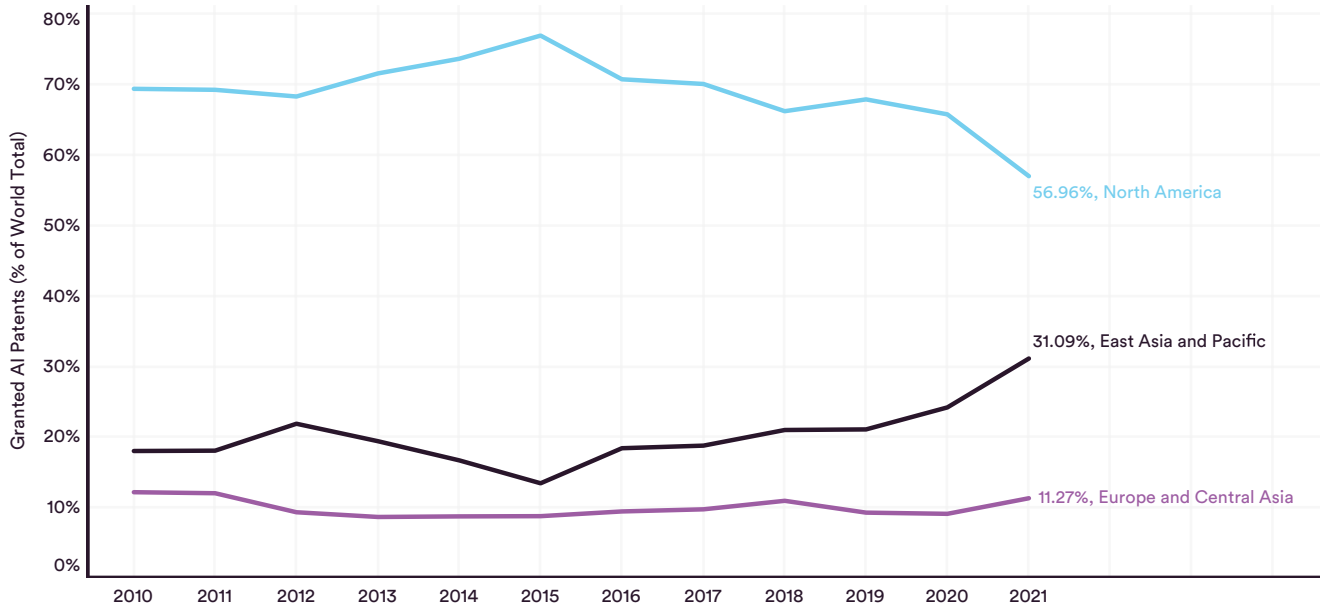

Figure 1.1.23b

GRANTED AI PATENTS (% of WORLD TOTAL) by REGION, 2010–21 (Part 2)

Source: Center for Security and Emerging Technology, 2021 | Chart: 2022 AI Index Report

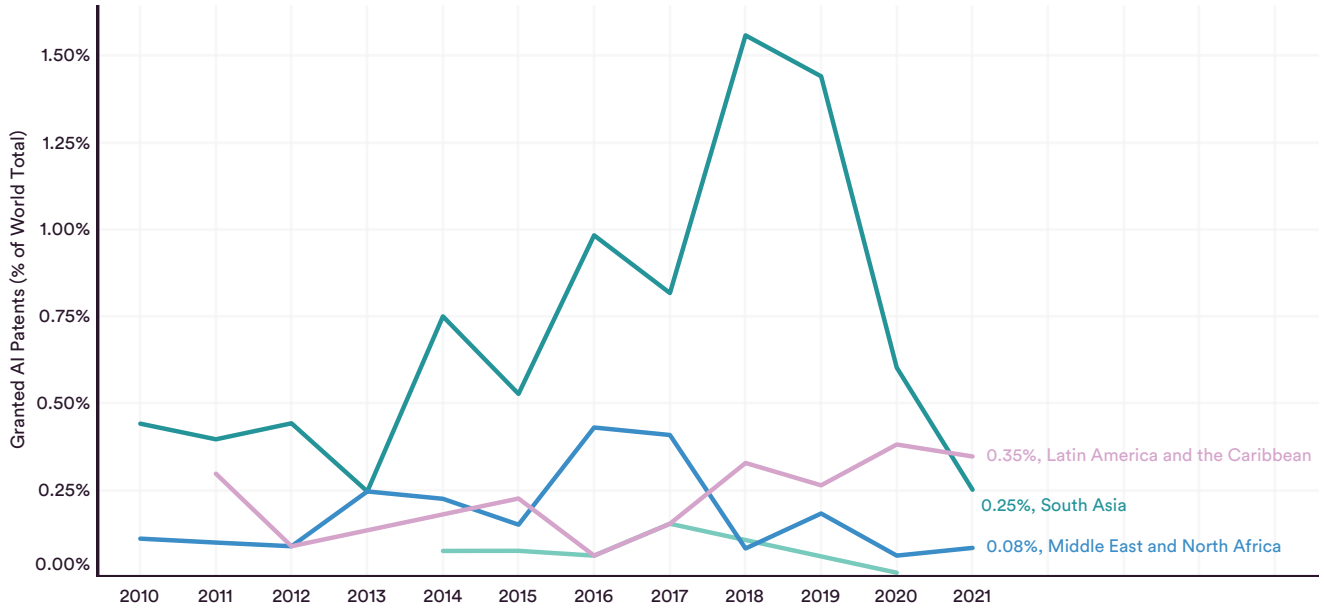

Figure 1.1.23c

By Geographic Area and Application Status

Trends revealed by the regional analysis can also be observed in AI patent data broken down by geographic area (Figure 1.1.24a and Figure 1.1.24b). China is now filing over half of the world's AI patents and being granted about 6%, about the same as the European Union plus the United Kingdom. The United States, which files almost all the patents in North America, does so at one-third the rate of China. Figure 1.1.24c shows that compared to the increasing numbers of AI patents applied and granted, China has far greater numbers of patent applications (87,343 in 2021) than those granted (1,407 in 2021).

AI PATENT FILINGS (% of WORLD TOTAL) by GEOGRAPHIC AREA, 2010–21

Source: Center for Security and Emerging Technology, 2021 | Chart: 2022 AI Index Report

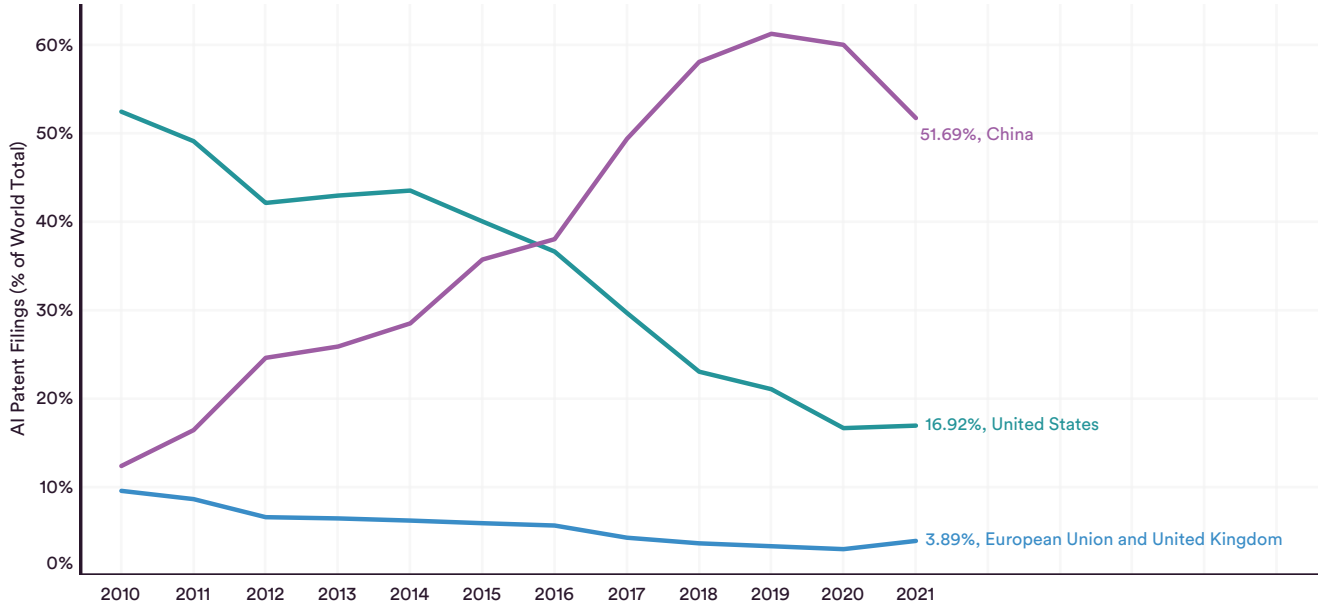

Figure 1.1.24a

GRANTED AI PATENTS (% of WORLD TOTAL) by GEOGRAPHIC AREA, 2010–21

Source: Center for Security and Emerging Technology, 2021 | Chart: 2022 AI Index Report

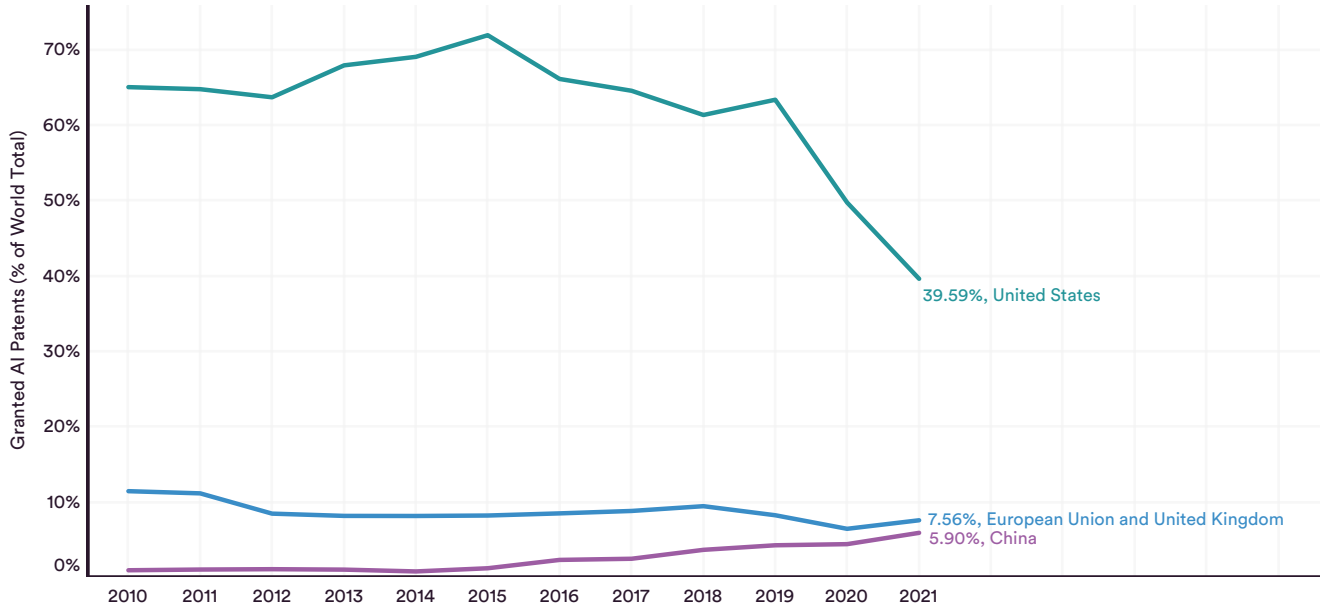

Figure 1.1.24b

AI PATENTS by APPLICATION STATUS by GEOGRAPHIC AREA, 2010–21

Source: Center for Security and Emerging Technology, 2021 | Chart: 2022 AI Index Report

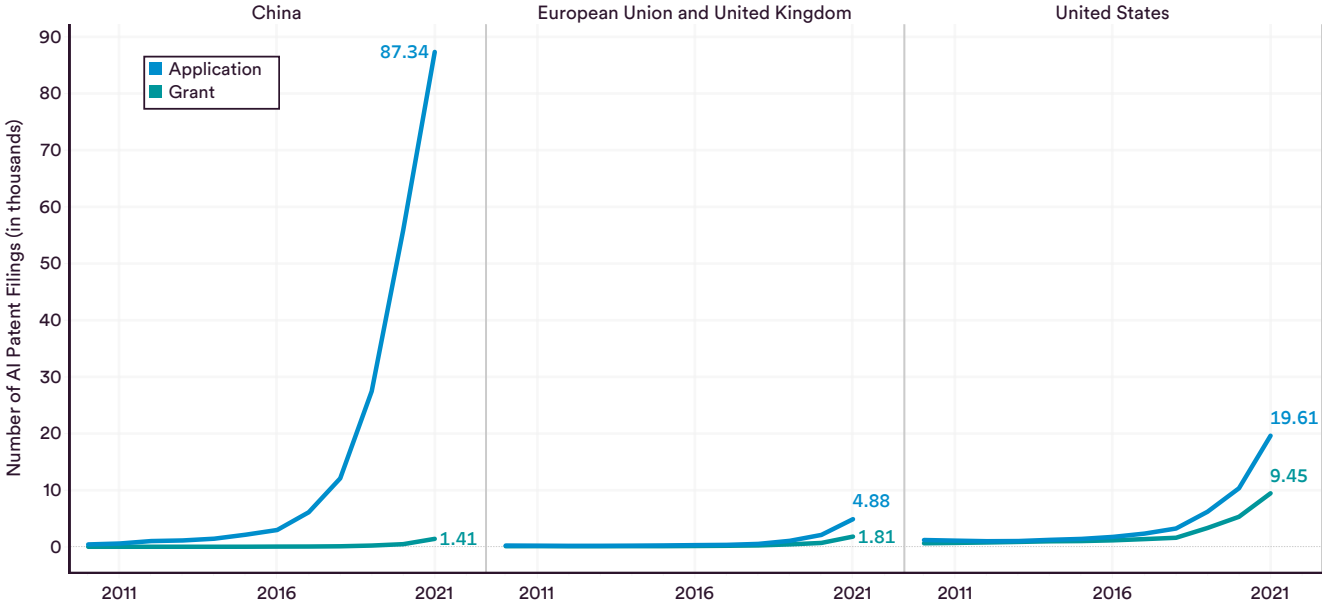

Figure 1.1.24c

AI conferences are key venues for researchers to publish and communicate their work, as well as to connect with peers and collaborators. Conference attendance is an indication of broader industrial and academic interest in a scientific field. In the past 20 years, AI conferences have grown not only in size but also in number and prestige. This section presents data on the trends in attendance at major AI conferences, covering more conferences (16) than previous Index reports.

1.2 CONFERENCES

CONFERENCE ATTENDANCE

Similar to 2020, most AI conferences were offered virtually in 2021. Only the International Conference on Robotics and Automation (ICRA) and the Conference on Empirical Methods in Natural Language Processing (EMNLP) were held using a hybrid format. Conference organizers reported measuring the exact attendance numbers at a virtual conference is difficult, as virtual conferences allow for higher attendance of researchers from all around the world.

Figure 1.2.1 shows that attendance at top AI conferences in 2021 was relatively consistent with 2020, with more than 88,000 participants worldwide. Figure 1.2.2 and Figure 1.2.3 show the attendance data for individual conferences, with 16 major AI conferences split into two categories: large AI conferences with over 2,500 attendees and small AI conferences with fewer than 2,500 attendees.[6]

NUMBER of ATTENDEES at SELECT AI CONFERENCES, 2010–21
Source: Conference Data, 2021 | Chart: 2022 AI Index Report

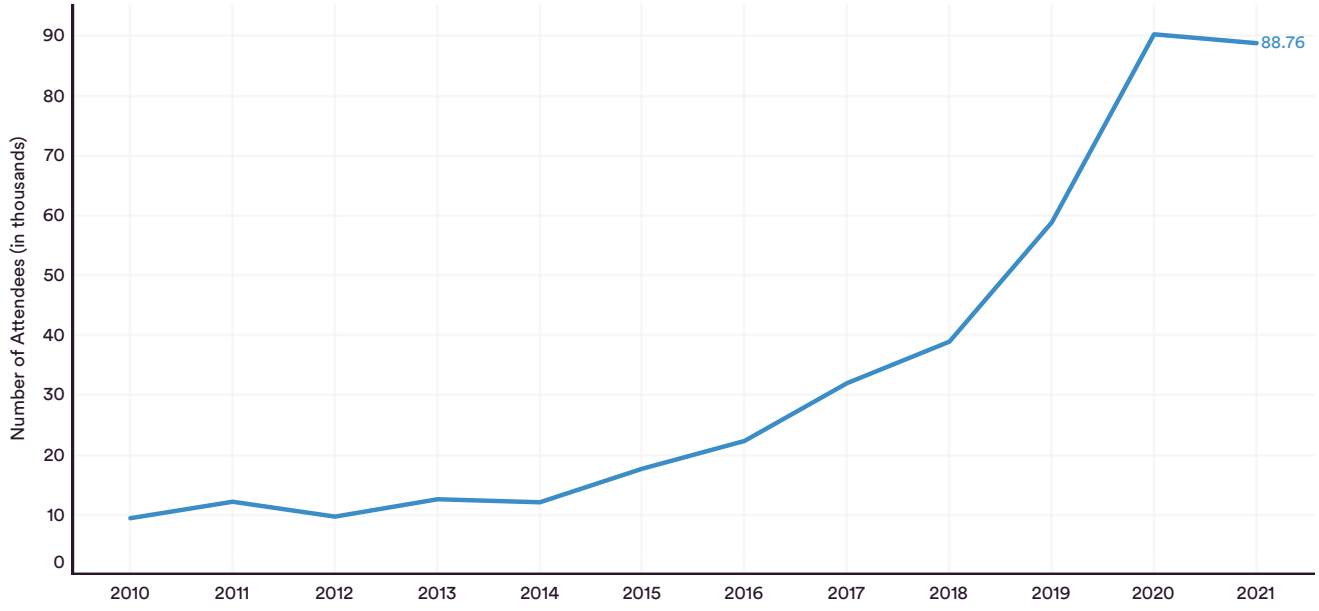

Figure 1.2.1

6 The International Conference on Machine Learning (ICML) used the number of session visitors as a proxy for the number of conference attendees, which explains the high attendance count in 2021. The International Conference on Intelligent Robots and Systems (IROS) extended the virtual conference to allow users to watch events for up to three months, which explains the high attendance count in 2020. For the AAMAS conference, the attendance in 2020 is based on the number of users on site reported by the platform that recorded the talks and managed the online conference, while the 2021 number is for total registrants.

ATTENDANCE at LARGE AI CONFERENCES, 2010–21

Source: Conference Data, 2021 | Chart: 2022 AI Index Report

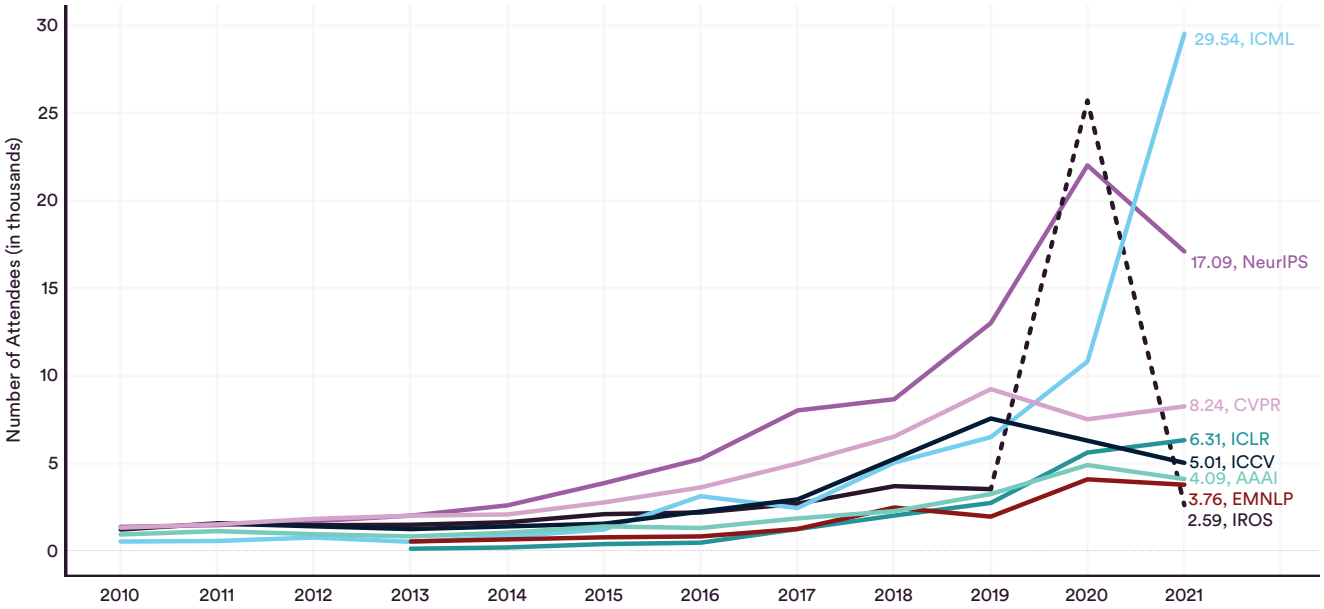

Figure 1.2.2

ATTENDANCE at SMALL AI CONFERENCES, 2010–21

Source: Conference Data, 2021 | Chart: 2022 AI Index Report

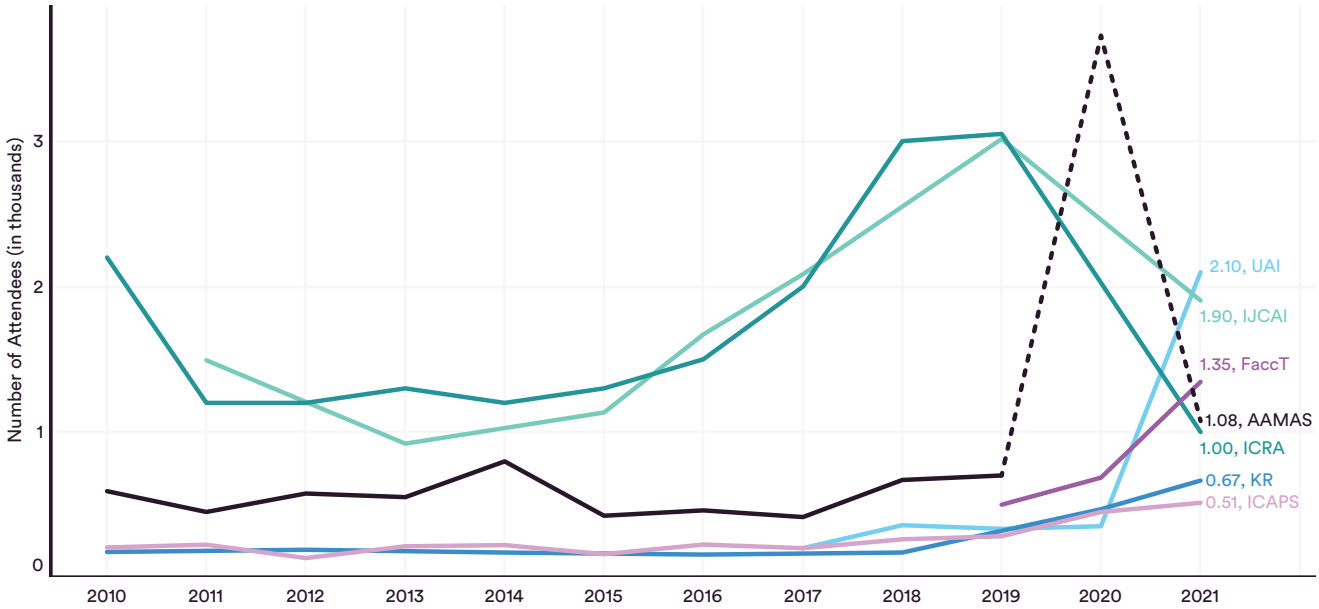

Figure 1.2.3

WOMEN IN MACHINE LEARNING (WIML) NEURIPS WORKSHOP

Founded in 2006, Women in Machine Learning is an organization dedicated to supporting and increasing the impact of women in machine learning. This section presents data from its annual technical workshop colocated with NeurIPS. Starting in 2020, WiML has also been hosting the Un-Workshop, which aims to advance research via collaboration and interaction among participants from diverse backgrounds at ICML.

Workshop Participants

The number of participants attending the WiML workshop has steadily increased since it was first introduced in 2006. For the 2021 edition, Figure 1.2.4 shows an estimate of 1,486 attendees over all workshop sessions, counted as the number of unique individuals who accessed the virtual workshop platform at neurips.cc. The 2021 WiML Workshop at NeurIPS happened as multiple sessions over three days, which was a change in format from 2020. As in 2020, the workshop was held virtually due to the pandemic.

ATTENDANCE at NEURIPS WOMEN in MACHINE LEARNING WORKSHOP, 2010–21

Source: Women in Machine Learning, 2021 | Chart: 2022 AI Index Report

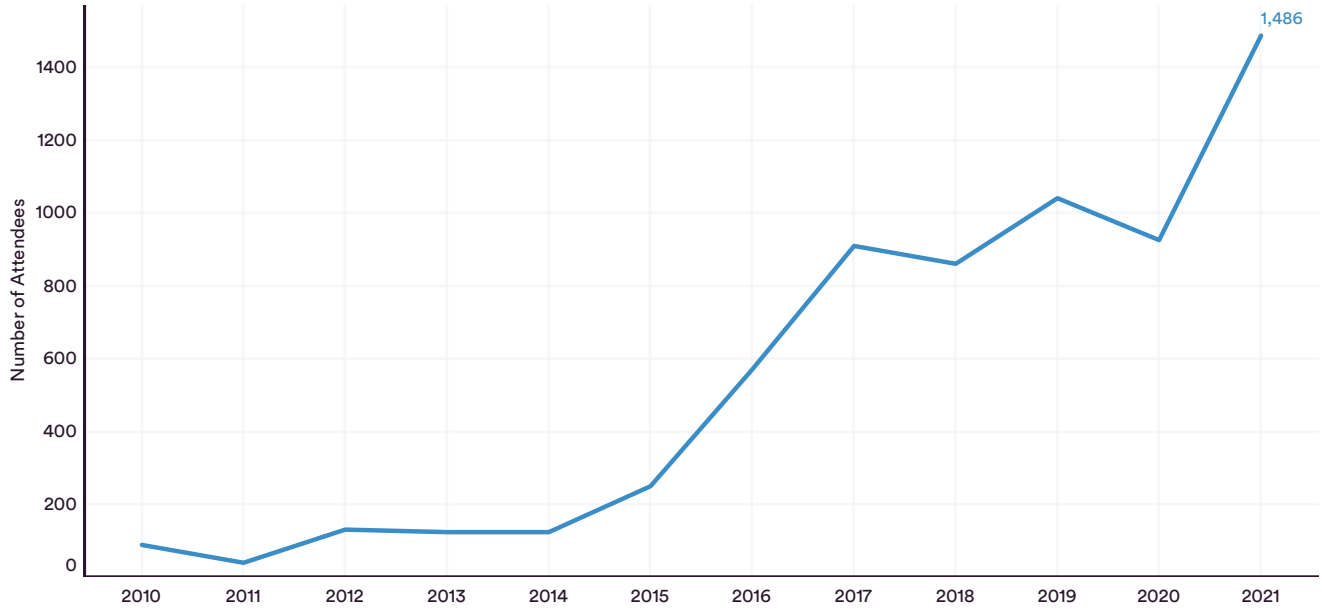

Figure 1.2.4

Demographics Breakdown

This section shows the continent of residence and professional position breakdowns of the 2021 workshop participants based on a survey filled by participants who consented to have such information aggregated. Among the survey respondents, more than half of the survey respondents were from North America, followed by Europe (19.9%), Asia (16.2%), and Africa (7.3%) (Figure 1.2.5). Figure 1.2.6 shows that Ph.D. students made up almost half of the survey participants, while the share of university faculty is around 1.2%. Researcher scientists/engineers, data scientists/engineers, and software engineers were among the most commonly held professional positions.

CONTINENT of RESIDENCE of PARTICIPANTS at NEURIPS WOMEN in MACHINE LEARNING WORKSHOP, 2021

Source: Women in Machine Learning, 2021 | Chart: 2022 AI Index Report

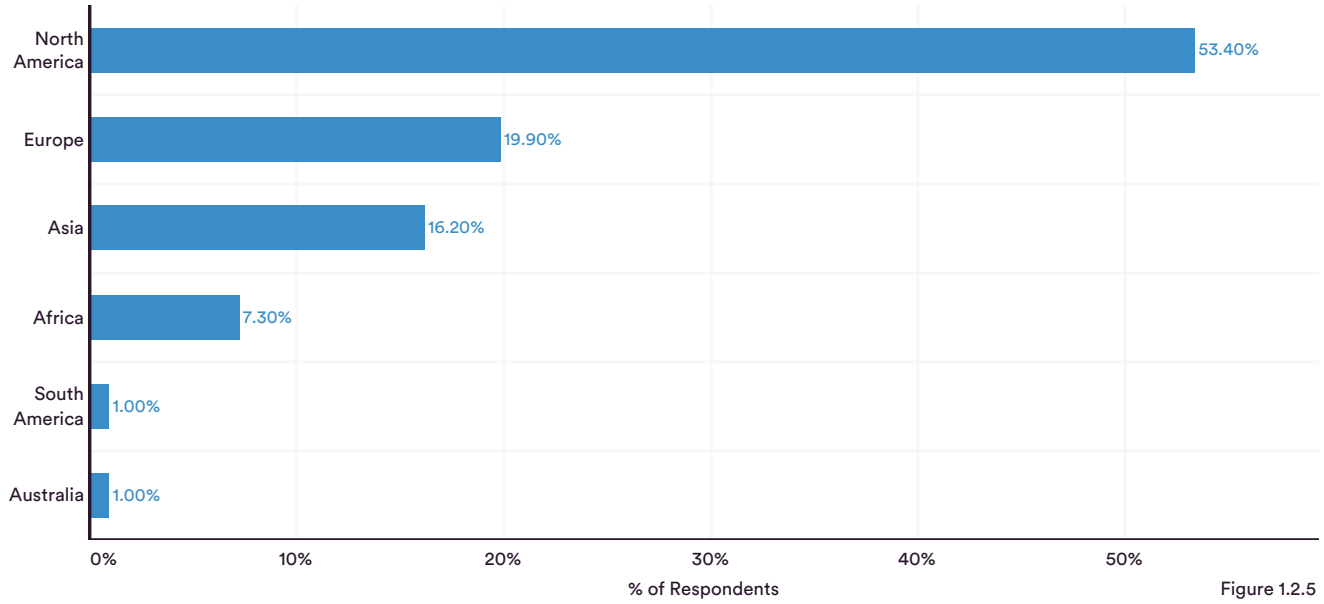

Figure 1.2.5

PROFESSIONAL POSITIONS of PARTICIPANTS at NEURIPS WOMEN in MACHINE LEARNING WORKSHOP, 2021

Source: Women in Machine Learning, 2021 | Chart: 2022 AI Index Report

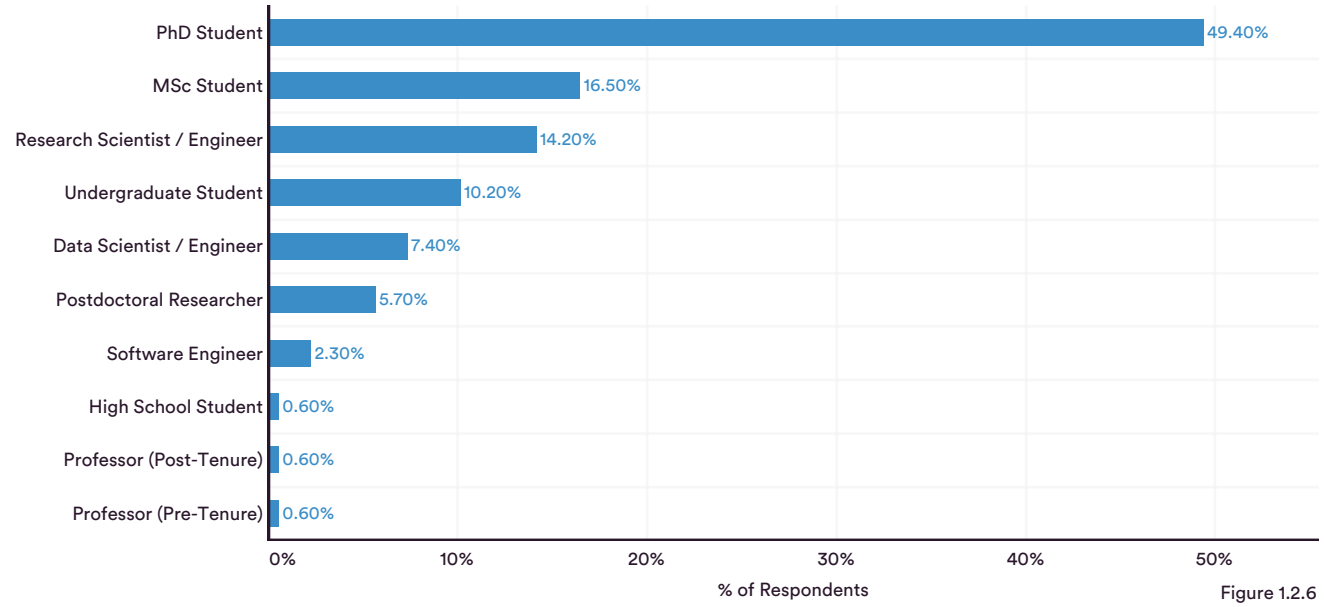

Figure 1.2.6

A software library is a collection of computer code that is used to create applications and products. Popular AI-specific software libraries—such as TensorFlow and PyTorch—help developers create their AI solutions quickly and efficiently. This section analyzes the popularity of software libraries through GitHub data.

1.3 AI OPEN-SOURCE SOFTWARE LIBRARIES

GITHUB STARS

Figures 1.3.1 and 1.3.2 reflect the number of users of GitHub open-source AI software libraries from 2015 to 2021. TensorFlow remained by far the most popular in 2021, with around 161,000 cumulative GitHub stars—a slight increase over 2020. TensorFlow was about three times as popular in 2021 as the next-most-starred GitHub open-source AI software library, OpenCV, which was followed by Keras, PyTorch, and Scikit-learn. Figure 1.3.2 shows library popularity for libraries with fewer than 40,000 GitHub stars—led by FaceSwap with around 40,000 stars, followed by 100-Days-Of-ML-Code, AiLearning, and BVLC/caffe.

NUMBER of GITHUB STARS by AI LIBRARY (OVER 40K STARS), 2014–21
Source: GitHub, 2021 | Chart: 2022 AI Index Report

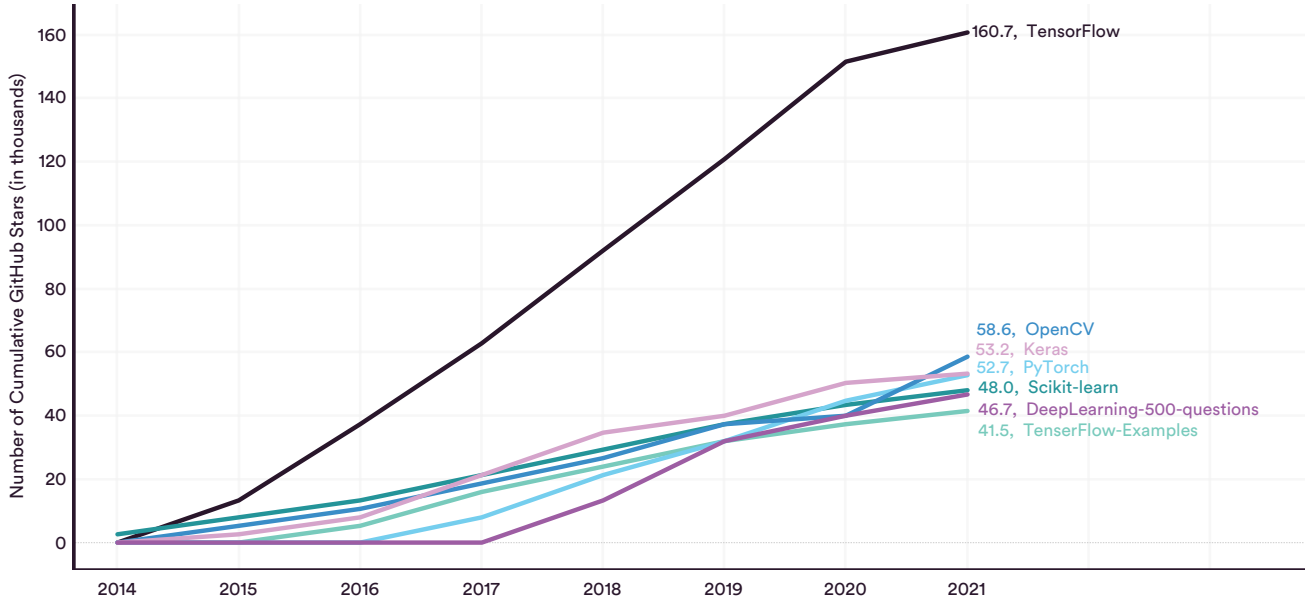

Figure 1.3.1

NUMBER of GITHUB STARS by AI LIBRARY (UNDER 40K STARS), 2014–21

Source: GitHub, 2021 | Chart: 2022 AI Index Report

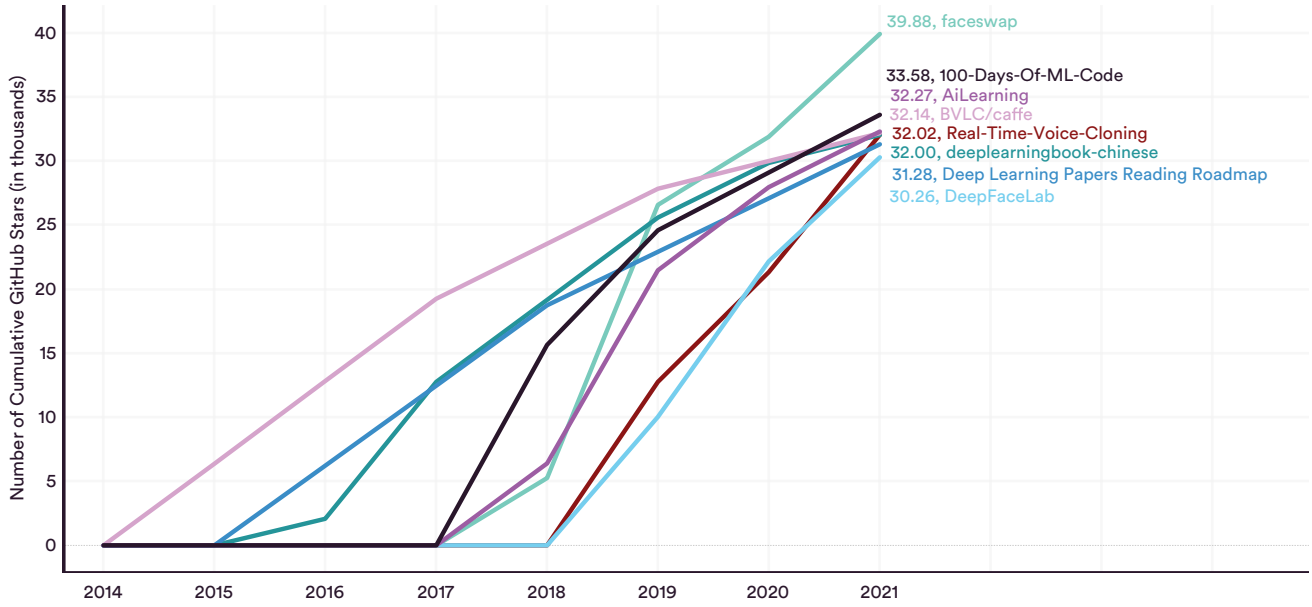

Figure 1.3.2

APPENDIX

CENTER FOR SECURITY AND EMERGING TECHNOLOGY, GEORGETOWN UNIVERSITY

Prepared by Sara Abdulla and James Dunham

The Center for Security and Emerging Technology (CSET) is a policy research organization within Georgetown University's Walsh School of Foreign Service that produces data-driven research at the intersection of security and technology, providing nonpartisan analysis to the policy community.

Publications from CSET Merged Corpus of Scholarly Literature

Source

CSET's merged corpus of scholarly literature combines distinct publications from Digital Science's Dimensions, Clarivate's Web of Science, Microsoft Academic Graph, China National Knowledge Infrastructure, arXiv, and Papers With Code.[1]

Methodology

To create the merged corpus, CSET deduplicated across the listed sources using publication metadata, and then combined the metadata for linked publications. To identify AI publications, CSET used an English-language subset of this corpus: publications since 2010 that appear AI-relevant.[2] CSET researchers developed a classifier for identifying AI-related publications by leveraging the arXiv repository, where authors and editors tag papers by subject.

To provide a publication's field of study, CSET matches each publication in the analytic corpus with predictions from Microsoft Academic Graph (MAG)'s field-of-study model, which yields hierarchical labels describing the published research field(s) of study and corresponding scores.[3] CSET researchers identified the most common fields of study in their corpus of AI-relevant publications since 2010 and recorded publications in all other fields as "Other AI." English-language AI-relevant publications were then tallied by their top-scoring field and publication year.

CSET also provided year-by-year citations for AI-relevant work associated with each country. A publication is associated with a country if it has at least one author whose organizational affiliation(s) is located in that country. Citation counts aren't available for all publications; those without counts weren't included in the citation analysis. Over 70% of English-language AI papers published between 2010 and 2020 have citation data available.

CSET counted cross-country collaborations as distinct pairs of countries across authors for each publication. Collaborations are only counted once: For example, if a publication has two authors from the United States and two authors from China, it is counted as a single United States-China collaboration.

Additionally, publication counts by year and by publication type (e.g., academic journal articles, conference papers) were provided where available. These publication types were disaggregated by affiliation country as described above.

CSET also provided publication affiliation sector(s) where, as in the country attribution analysis, sectors were associated with publications through authors' affiliations. Not all affiliations were characterized in terms of sectors; CSET researchers relied primarily on GRID from Digital Science for this purpose, and not all organizations can be found in or linked to GRID.[4] Where the affiliation

1 All CNKI content is furnished for CSET by East View Information Services, Minneapolis, MN, USA.
2 For more information, see James Dunham, Jennifer Melot, and Dewey Murdick, "Identifying the Development and Application of Artificial Intelligence in Scientific Text," arXiv [cs.DL], May 28, 2020, https://arxiv.org/abs/2002.07143.
3 These scores are based on cosine similarities between field-of-study and paper embeddings. See Zhihong Shen, Hao Ma, and Kuansan Wang, "A Web-Scale System for Scientific Knowledge Exploration," arXiv [cs.CL], May 30, 2018, https://arxiv.org/abs/1805.12216.
4 See https://www.grid.ac/ for more information about the GRID dataset from Digital Science.

sector is available, papers were counted toward these sectors, by year. Cross-sector collaborations on academic publications were calculated using the same method as in the cross-country collaborations analysis.

Patents from CSET's AI Patents Dataset

Source

CSET's AI patents dataset was developed by CSET and 1790 Analytics. It includes patents relevant to the development and application of AI, as indicated by CPC/IPC codes and keywords.

Methodology

In this analysis, patents were grouped by year and country, and then counted at the "patent family" level.[5] CSET extracted year values from the most recent publication date within a family. This method has the advantage of capturing updates within a patent family (such as amendments).

The country of origin for a patent is derived from the first country in which a patent was filed.[6]

GITHUB STARS

Source

GitHub: star-history (available at star history website) was used to retrieve the data.

Methodology

The visual in the report shows the number of stars for various GitHub repositories over time. The repositories include the following:
apachecn/ailearning, apache/incubator-mxnet, Avik-Jain/100-Days-Of-ML-Code, aymericdamien/TensorFlow-Examples, BVLC/cafe, cafe2/cafe2, CorentinJ/Real-Time-Voice-Cloning, deepfakes / faceswap, dmlc/mxnet, exacity/deeplearningbook-chinese, fchollet/keras, floodsung/Deep-Learning-Papers-Reading-Roadmap, iperov/DeepFaceLab, Microsoft/CNTK, opencv/opencv, pytorch/pytorch, scikit-learn/scikit-learn, scutan90/DeepLearning-500-questions, tensorflow/tensorflow, Theano/Theano, Torch/Torch7.

Nuance

The GitHub Archive currently does not provide a way to count when users remove a star from a repository. Therefore, the reported data slightly overestimates the number of stars. A comparison with the actual number of stars for the repositories on GitHub reveals that the numbers are fairly close and that the trends remain unchanged.

5 Patents are analyzed at the "patent family" level rather than "patent document" level because patent families are a collective of patent documents all associated with a single invention and/or innovation by the same inventors/assignees. Thus, counting at the "patent family" level mitigates artificial number inflation when there are multiple patent documents in a patent family or if a patent is filed in multiple jurisdictions.
6 For more details on CSET's approach and experimentation for assigning country values, see footnote 26 in "Patents and Artificial Intelligence: A Primer," by Dewey Murdick and Patrick Thomas. (Center for Security and Emerging Technology, September 2020), https://doi.org/10.51593/20200038.