artificial intelligence index

2018 annual report

Steering Committee

Yoav Shoham *(Chair)*
Stanford University

Raymond Perrault
SRI International

Erik Brynjolfsson
MIT

Jack Clark
OpenAI

James Manyika
McKinsey Global Institute

Juan Carlos Niebles
Stanford University

Terah Lyons
Partnership On AI

John Etchemendy
Stanford University

Barbara Grosz
Harvard University

Project Manager

Zoe Bauer

Welcome to the AI Index 2018 Report

Our Mission is to ground the conversation about AI in data.

The AI Index is an effort to track, collate, distill, and visualize data relating to artificial intelligence. It aspires to be a comprehensive resource of data and analysis for policymakers, researchers, executives, journalists, and the general public to develop intuitions about the complex field of AI.

AI INDEX 2018

Table of contents

Introduction to the AI Index 2018 Annual Report

We are pleased to introduce the AI Index 2018 Annual Report. This year's report accomplishes two objectives. First, it refreshes last year's metrics. Second, it provides global context whenever possible. The former is critical to the Index's mission — grounding the AI conversation means tracking volumetric and technical progress on an ongoing basis. But the latter is also essential. There is no AI story without global perspective. The 2017 report was heavily skewed towards North American activities. This reflected a limited number of global partnerships established by the project, not an intrinsic bias. This year, we begin to close the global gap. We recognize that there is a long journey ahead — one that involves further collaboration and outside participation — to make this report truly comprehensive.

Still, we can assert that *AI is global*. 83 percent of 2017 AI papers on Scopus originate outside the U.S. 28 percent of these papers originate in Europe — the largest percentage of any region. University course enrollment in artificial intelligence (AI) and machine learning (ML) is increasing all over the world, most notably at Tsinghua in China, whose combined AI + ML 2017 course enrollment was 16x larger than it was in 2010. And there is progress beyond just the United States, China, and Europe. South Korea and Japan were the 2nd and 3rd largest producers of AI patents in 2014, after the U.S. Additionally, South Africa hosted the second Deep Learning Indaba conference, one of the world's largest ML teaching events, which drew over 500 participants from 20+ African countries.

AI's diversity is not just geographic. Today, over 50% of the Partnership on AI's members are nonprofits — including the ACLU, Oxford's Future of Humanity Institute, and the United Nations Development Programme. Also, there is heightened awareness of gender and racial diversity's importance to progress in AI. For example, we see increased participation in organizations like AI4ALL and Women in Machine Learning (WiML), which encourage involvement by underrepresented groups.

AI Index Report Overview

The report has four sections:
1. Data: Volume of Activity and Technical Performance
2. Other measures: Recent Government Initiatives, Derivative measures, and Human-Level Performance
3. Discussion: What's Missing?
4. Appendix

DATA

The ***Volume of Activity*** metrics capture engagement in AI activities by academics, corporations, entrepreneurs, and the general public. Volumetric data ranges from the number of undergraduates studying AI, to the percent of female applicants for AI jobs, to the growth in venture capital funding of AI startups.

The ***Technical Performance*** metrics capture changes in AI performance over time. For example, we measure the quality of question answering and the speed at which computers can be trained to detect objects.

The 2018 AI Index adds additional country-level granularity to many of last year's metrics, such as robot installations and AI conference attendance. Additionally, we have added several new metrics and areas of study, such as patents, robot operating system downloads, the GLUE metric, and the COCO leaderboard. Overall, we see a continuation of last year's main takeaway: AI activity is increasing nearly everywhere and technological performance is improving across the board. Still, there were certain takeaways this year that were particularly interesting. These include the considerable improvement in natural language and the limited gender diversity in the classroom.

OTHER MEASURES

Like last year, the ***Derivative Measures*** section investigates relationships between trends. We also show an exploratory measure, the AI Vibrancy Index, which combines trends across academia and industry to quantify the liveliness of AI as a field.

We introduce a new qualitative metric this year: ***Recent Government Initiatives***. This is a simplified overview of recent government investments in artificial intelligence. We include initiatives from the U.S., China, and Europe. The AI Index looks forward to including more government data and analysis in future reports by collaborating with additional organizations.

The ***Human-Level Performance Milestones*** section of the report builds on our timeline of instances where AI shows human and superhuman abilities. We include four new achievements from 2018.

AI Index Report Overview (continued)

Finally, to start a conversation in the AI community, the **What's Missing?** section presents suggestions from a few experts in the field, who offer ideas about how the AI Index could be made more comprehensive and representative.

APPENDIX

The **Appendix** supplies readers with a fully transparent description of sources, methodologies, and nuances. Our appendix also houses underlying data for nearly every graph in the report. We hope that each member of the AI community interacts with the data most relevant to their work and interests.

SYMBOLS

We earmark pages with the globe symbol below when discussing AI's universality. This includes country comparisons, deep dives into regions outside of the U.S., and data on diversity in the AI community.

VOLUME OF ACTIVITY

Published Papers: Papers by topic

The graph below shows growth in annual publishing rates of academic papers, relative to their rates in 1996. The graph compares the growth of papers across *All fields*, *Computer Science (CS)*, and *Artificial Intelligence (AI)*.

The growth of annually published papers in AI continues to outpace that of annually published papers in CS, suggesting that growth in AI publishing is driven by more than a heightened interest in computer science.

See Appendix 1 for data and methodology.

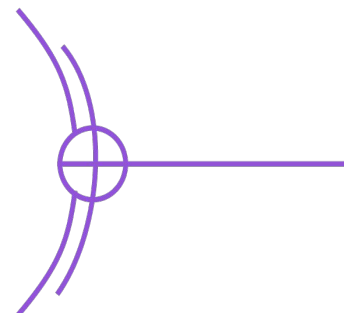Growth of annually published papers by topic (1996–2017)
Source: Scopus
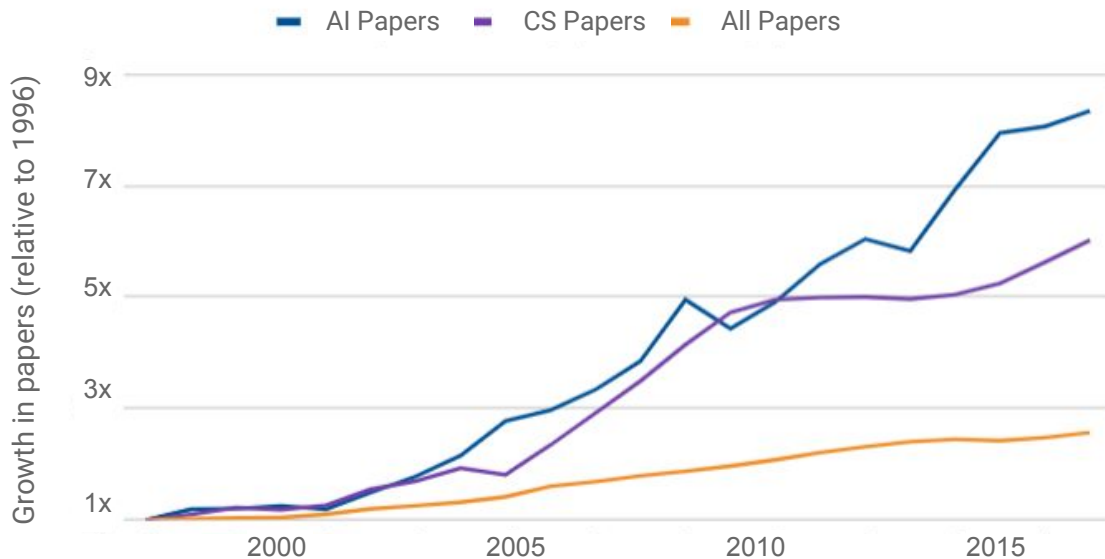
Note: This visual uses the Scopus query search term "Artificial Intelligence," not the Elsevier keyword approach. See more details in the appendix.

AI outpaces CS
AI papers on Scopus have increased 8x since 1996.
CS papers increased 6x during the same timeframe.

Published Papers: AI papers by region

The graph below shows the number of AI papers published annually by region.

Europe has consistently been the largest publisher of AI papers — 28% of AI papers on Scopus in 2017 originated in Europe. Meanwhile, the number of papers published in China increased 150% between 2007 and 2017. This is despite the spike and drop in Chinese papers around 2008.

See Appendix 2 for data and methodology.

Annually published AI papers on Scopus by region (1998—2017)
Source: Elsevier
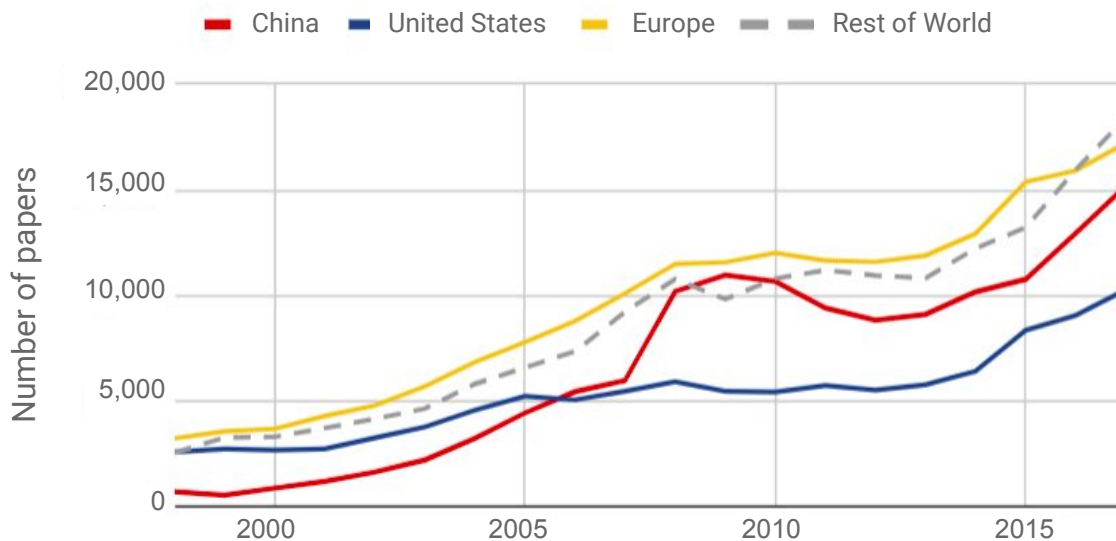
Note: We speculate that the increase in AI papers in China around 2008 is a result of The National Medium- and Long-Term Program for Science and Technology Development (2006— 2020), and other government programs that provide funding and a range of incentive policies for AI research. Similarly, FP7 (2007—2013) and other science and technology research programs in Europe may have contributed to the small uptick in papers around 2008—2010.

Europe is the largest publisher of AI papers

In 2017, 28% of AI papers on Scopus were affiliated with European authors, followed by China (25%) and the U.S. (17%).
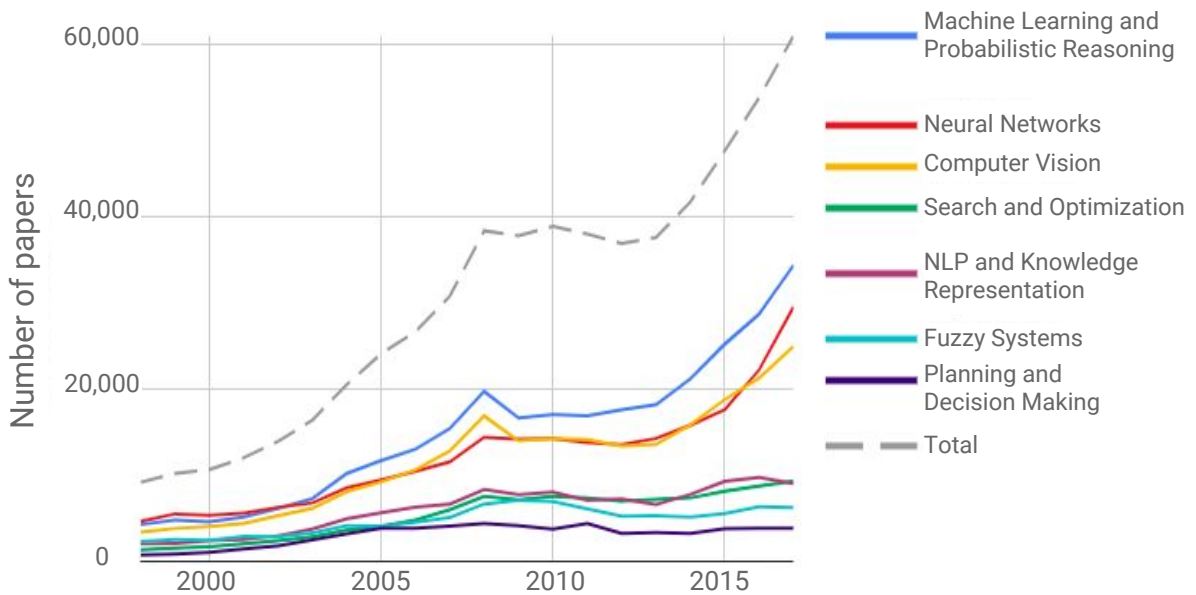
Published Papers: AI papers by subcategory

The graph below shows the number of AI papers on Scopus, by subcategory. Categories are not mutually exclusive.

56 percent of papers fell into the *Machine Learning and Probabilistic Reasoning* category in 2017, compared to 28% in 2010. For most categories below, papers were published at a faster rate during the period 2014—2017 than in the period 2010—2014. Most notably, *Neural Networks* had a compound annual growth rate (CAGR) of 3% from 2010—2014, followed by a CAGR of 37% from 2014—2017.

See Appendix 2 for data and methodology.

Number of AI papers on Scopus by subcategory (1998—2017)
Source: Elsevier

The number of Scopus papers on Neural Networks had a CAGR of 37% from 2014 to 2017

Published Papers: AI papers on arXiv

The graph below shows the number of AI papers on arXiv, by each paper's primary subcategory. The right axis refers the sum of all AI papers on arXiv (indicated by the grey dashed line).

The number of AI papers on arXiv is increasing overall and in a number of subcategories. This points to AI authors' tendency to disseminate their research, regardless of whether it is peer reviewed or has been accepted into AI conferences. This also points to the field's competitive nature.

Computer Vision (CV) and Pattern Recognition has been the largest AI subcategory on arXiv since 2014; prior to 2014, growth in this category closely tracked *Artificial Intelligence* and *Machine Learning*. In addition to showing a growing interest in *Computer Vision* (and its general applied applications), this also indicates the growth in other AI application areas, such as *Computation and Language* and *Robotics*.

See Appendix 3 for data and methodology.

Number of AI papers on arXiv by subcategory (2010—2017)
Source: arXiv
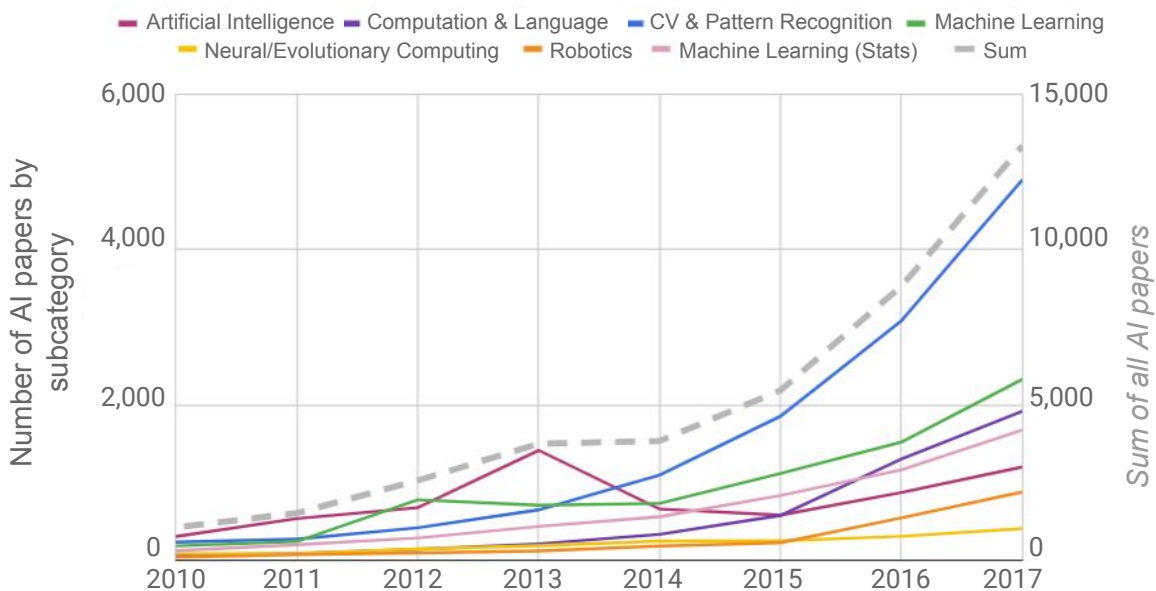
"…Aside from the increase in publications, it's important to note the adoption of arXiv by these communities for disseminating results. We've seen many times how establishing some critical mass then catalyzes ever higher levels of participation within a community…"
Paul Ginsparg, Cornell
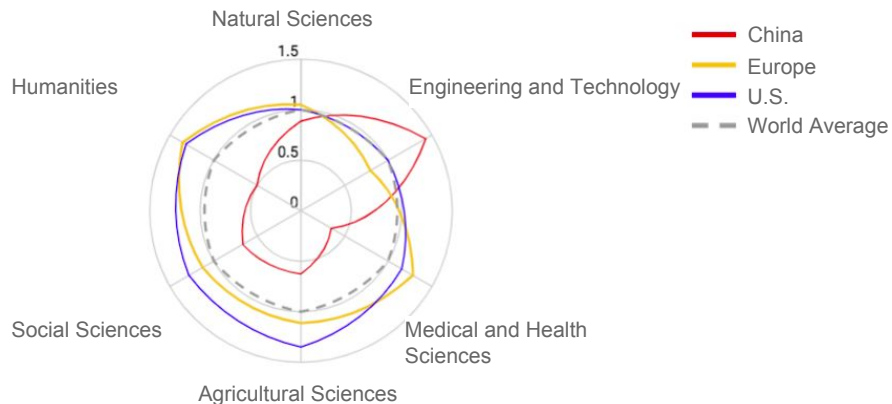
Published Papers: Activity focus by region

The graphs below show the Relative Activity Index (RAI) of the U.S., Europe, and China. RAI approximates a region's specialization by comparing it to global research activity in AI. RAI is defined as the share of a country's publication output in AI relative to the global share of publications in AI. A value of 1.0 indicates that a country's research activity in AI corresponds exactly with the global activity in AI. A value higher than 1.0 implies a greater emphasis, while a value lower than 1.0 suggests a lesser focus.

AI papers in China are more focused on *Engineering and Technology* and *Agricultural Sciences*, while AI papers in the U.S. and Europe tend to focus on *Humanities* and *Medical and Health Sciences*. Compared to data from 2000, data from 2017 show increased specialization across the three regions and a shift toward agriculture in China. This is inline with our expectations, as China is the largest food producer and tends to focus its research on applied AI.

See [Appendix 2](#) for data and methodology.

Relative activity focus by region and AI research sector (2000)
Source: Elsevier

Relative activity focus, by region and AI research sector (2017)
Source: Elsevier
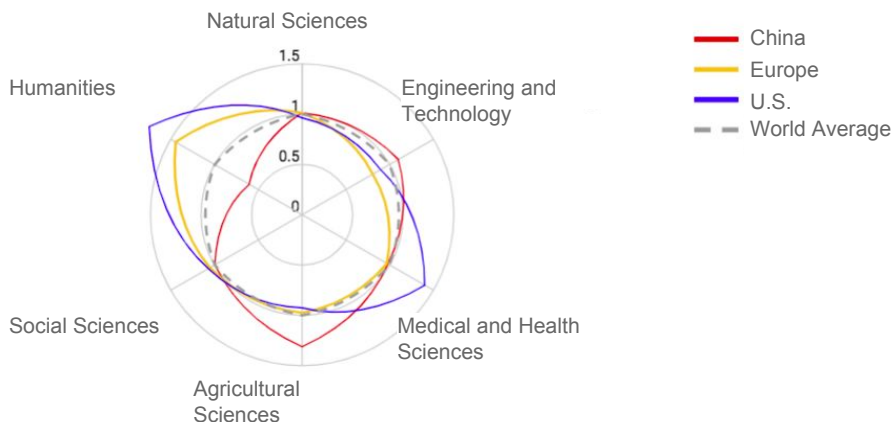
Published Papers: Sector focus

The graphs below and on the following page show the number of Scopus papers affiliated with government, corporate, and medical organizations. Then page 16 directly compares the three regions: we show one graph for the growth of papers in the corporate sector, and one for the growth of papers in the government sector.

In 2017, the Chinese government produced nearly 4x more AI papers than Chinese corporations. China has also experienced a 400% increase in government-affiliated AI papers since 2007, while corporate AI papers only increased by 73% in the same period.

In the U.S., a relatively large proportion of total AI papers are corporate. In 2017, the proportion of corporate AI papers in the U.S. was 6.6x greater than the proportion of corporate AI papers in China, and 4.1x greater than that in Europe.

Note that in all three regions academic papers (not shown) outweigh government, corporate, and medical papers. See comment under each visual.

See Appendix 2 for data and methodology.

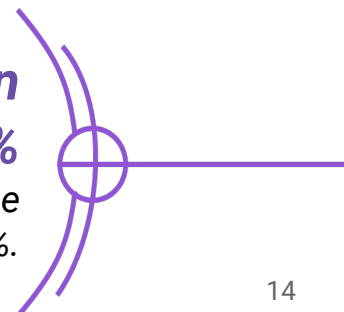Papers by sector affiliation — China (1998—2017)
Source: Elsevier
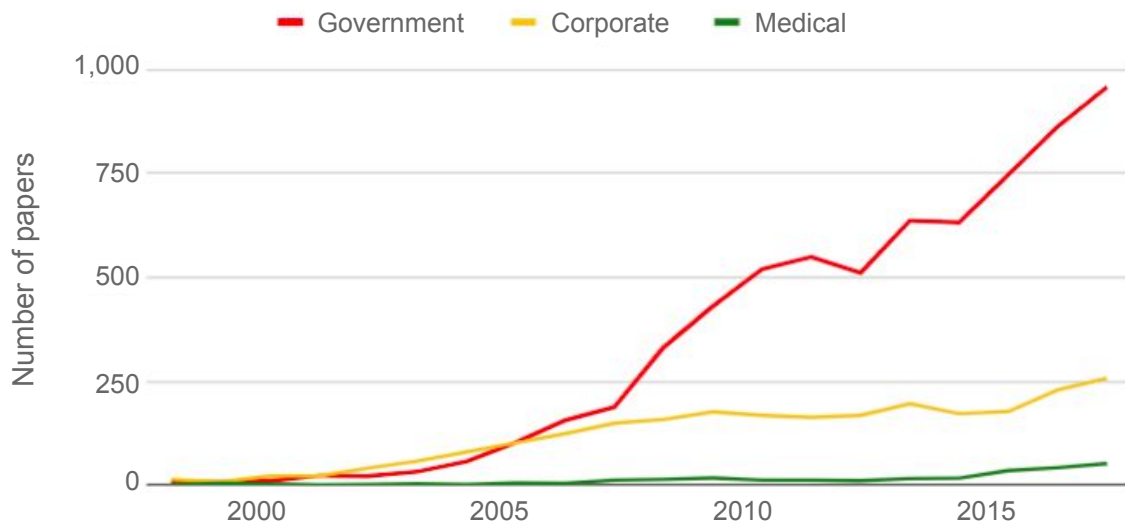
Note: For reference, in 2017 the number of AI papers on Scopus affiliated with the academic sector in China (not shown) was ~14,000 or ~92% of total AI papers in China.

Government-affiliated AI papers in China have increased 400%

From 2007—2017, government AI papers in China increase 400% while corporate AI papers in China increased by 73%.

Published Papers: Sector focus (continued)

Papers by sector affiliation — U.S. (1998—2017)
Source: Elsevier
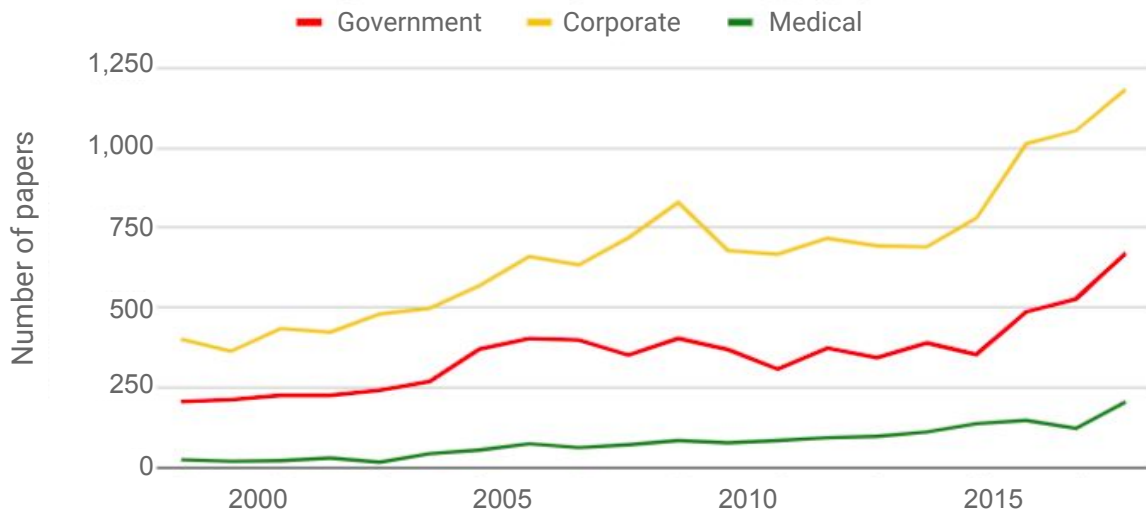
Note: For reference, in 2017 the number of AI papers on Scopus affiliated with the academic sector in the U.S. (not shown) was ~9,000 or ~81% of total AI papers in the U.S.

Papers by sector affiliation — Europe (1998—2017)
Source: Elsevier
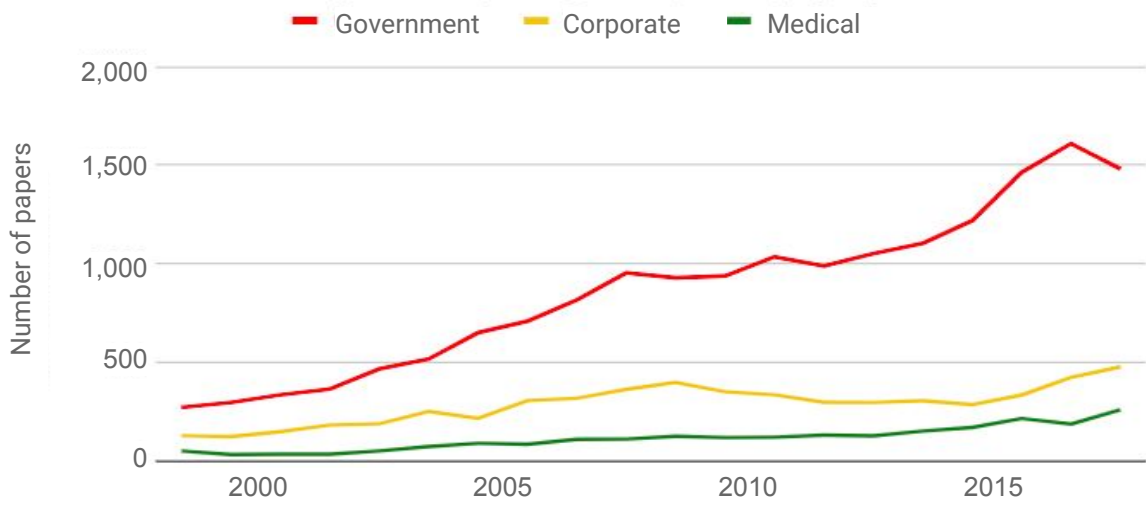
Note: For reference, in 2017 the number of AI papers on Scopus affiliated with the academic sector in Europe (not shown) was ~16,000 or ~88% of total AI papers in Europe.

The proportion of corporate papers in the U.S. is 6.6x greater than that in China

In 2017, the proportion of corporate AI papers in the U.S. was 6.6x greater than the proportion of corporate AI papers in China.

Published Papers: Sector focus (continued)

Growth in corporate-affiliated AI papers (2009—2017)
Source: Elsevier
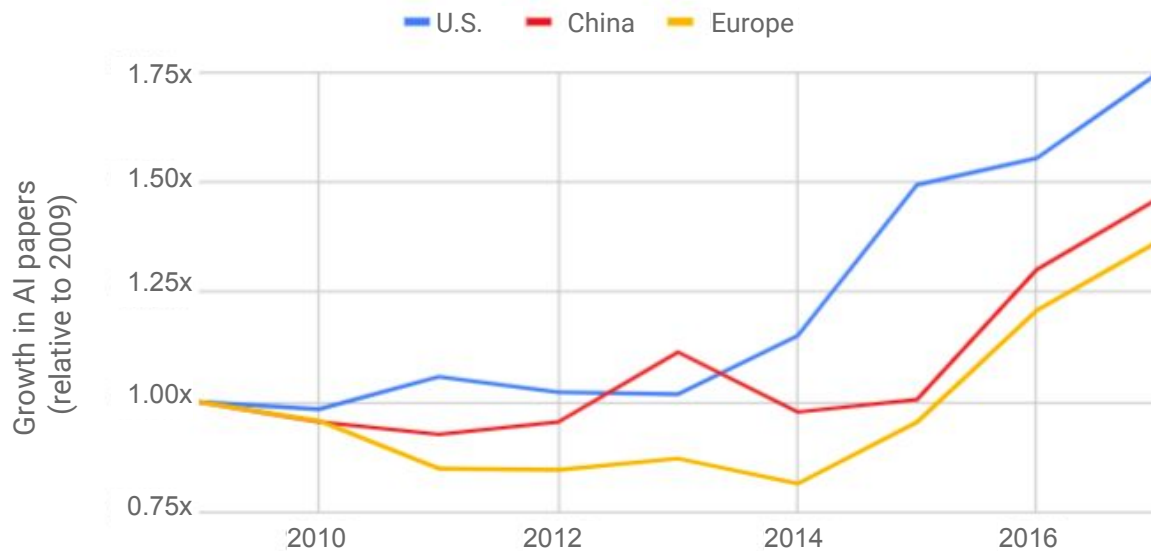
Growth in government-affiliated AI papers (2009—2017)
Source: Elsevier
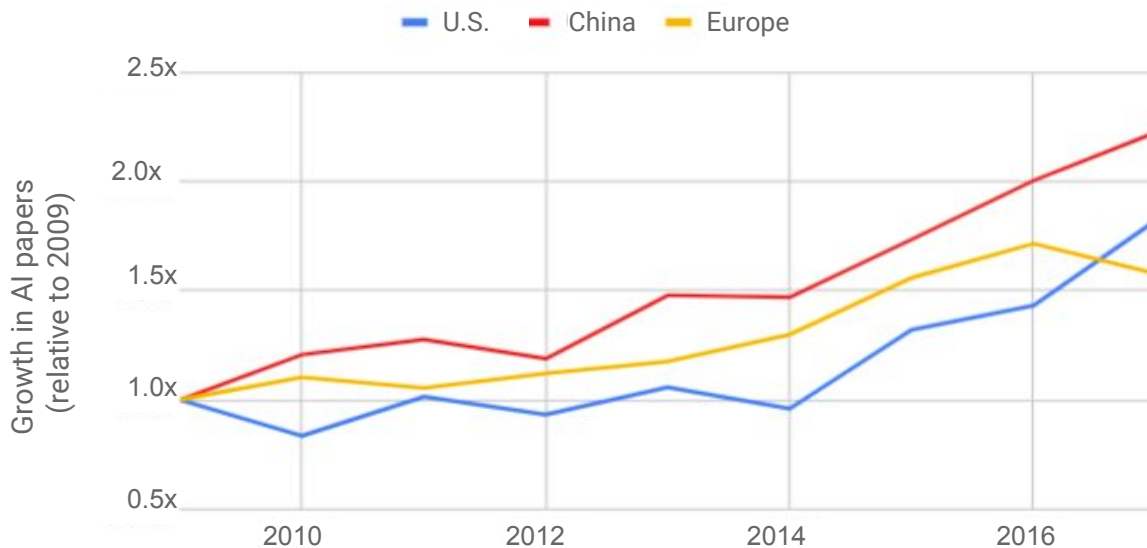
The number of Chinese government-affiliated AI papers has more than doubled since 2009

Meanwhile, the U.S. shows the greatest increase in corporate-affiliated AI papers. There were 1.7x as many corporate AI papers in 2017 as there were in 2009.

Published Papers: Citation impact by region

The graph below 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 authors in that region divided by the average number of citations by all AI authors. In this visual, the FWCI is *re-based,* meaning the citation impacts are shown relative to the world average. A re-based FWCI of 1 indicates that the publications have been cited on par with the world average. A re-based FWCI of 0.85 indicates that the papers are 15% less cited than the world average.

While Europe has the largest number of annually published AI papers, Europe's re-based FWCI has remained relatively flat and on par with the world average. In contrast, China has increased its re-based FWCI considerably. AI authors in China were cited 44% more in 2016 than they were in 2000. Still, the U.S. outperforms other regions in total citations. U.S. authors are cited 83% more than the global average.

See Appendix 2 for data and methodology.

Field-Weighted Citation Impact of AI authors by region (1998—2016)
Source: Elsevier
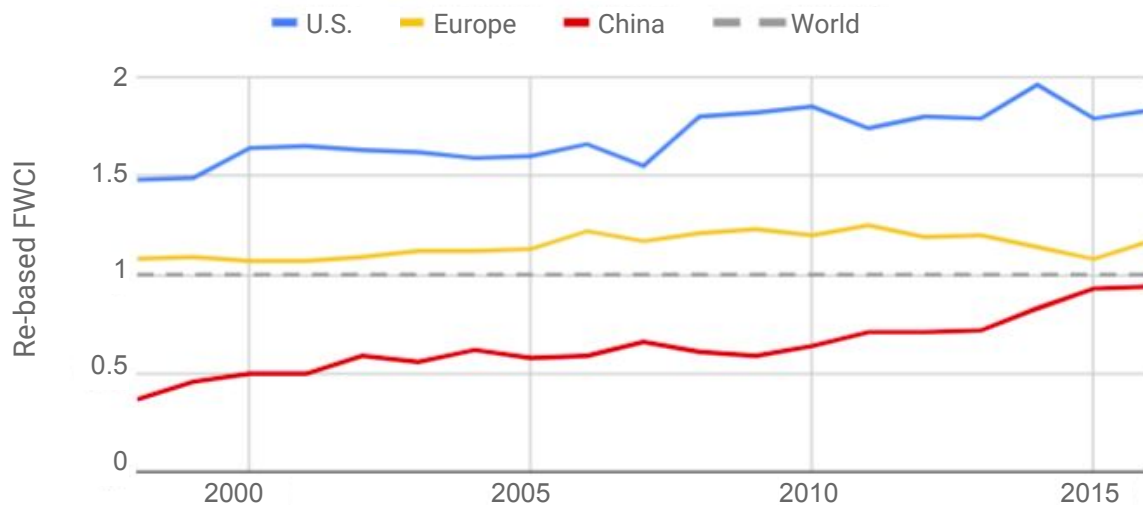
U.S. AI authors are cited 83% more than the global average.

Published Papers: Author mobility by region

The graphs on the [following page](#) show the effect of international mobility on the publication rate and the citation impact of AI authors.

We look at four mobility classes: *Sedentary, Transitory, Migratory Inflow and Migratory Outflow*. *Sedentary* authors are active researchers who have not published outside of their home region. *Transitory* authors have published in a region outside their home for two years or less. *Migratory* authors contribute papers to a region other than their own for two or more years. Whether an author is considered *migratory inflow* or *migratory outflow* depends on the perspective of the graph.

The x-axis shows the relative *publication rate* — the average number of publications for authors in each class divided by the average number of publications for that region overall. The y-axis shows the *field-weighted citation impact (FWCI)* — the average number of citations received by authors in each mobility class divided by the average number of citations for the region overall.

Only AI authors were considered in this analysis. An author is an "AI author" if at least 30% of his or her papers cover artificial intelligence. An author's home region is the region where they published their first paper.

…

Across the U.S., China, and Europe, the publication rate is lowest for sedentary authors. Additionally, in the three regions, migratory authors (a combination of migratory inflow and outflow) have the highest FWCI. So, authors that move tend to have a greater number of citations and tend to publish more frequently.

Of the three regions, China has the largest proportion of sedentary AI authors (76%), followed by Europe (52%), and then the U.S. (38%). Though a larger proportion of Chinese authors are sedentary, non-sedentary authors in China tend to have higher publication rates than non-sedentary authors in the other two regions. In other words, while geographically mobile Chinese authors are relatively few, they tend to be more productive than the average mobile author elsewhere.

See [Appendix 2](#) for data and methodology.

Mobile authors have a greater number of citations and publish more frequently

Published Papers: Author mobility by region
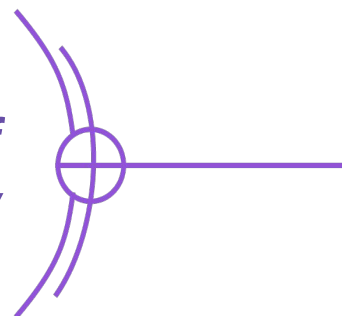
See description on the [previous page](#).

Publication rate and citation impact by mobility class — U.S. (1998—2017)
Source: Elsevier

Publication rate and citation impact by mobility class — China (1998—2017)
Source: Elsevier

Publication rate and citation impact by mobility class — Europe (1998—2017)
Source: Elsevier
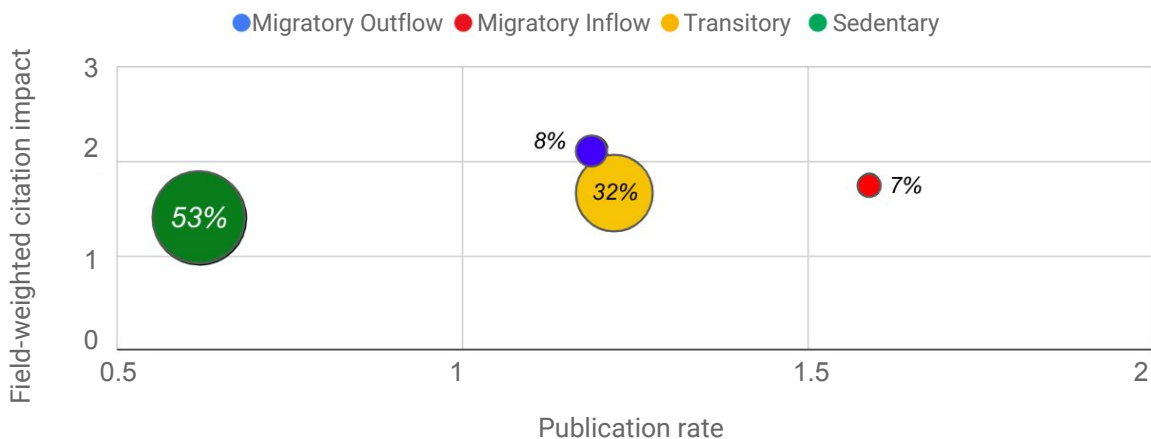
Published Papers: AAAI papers by country

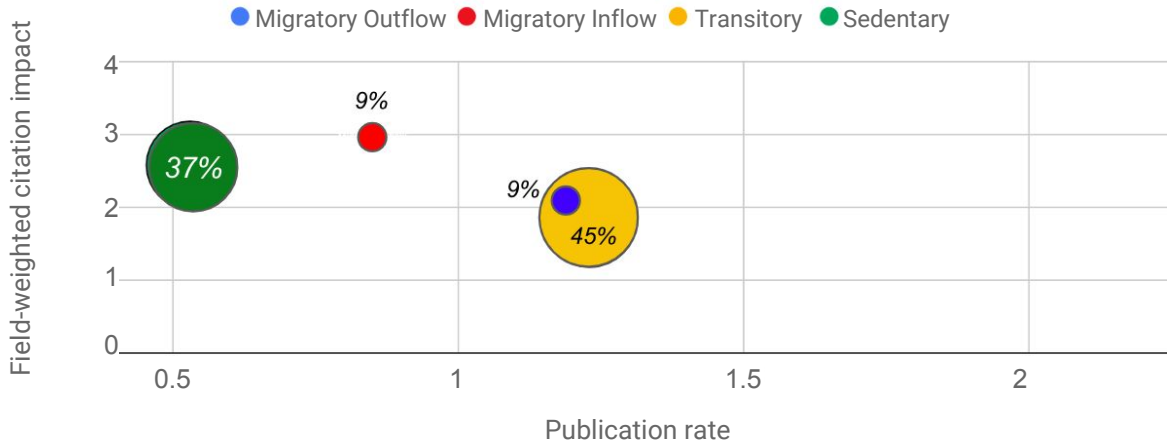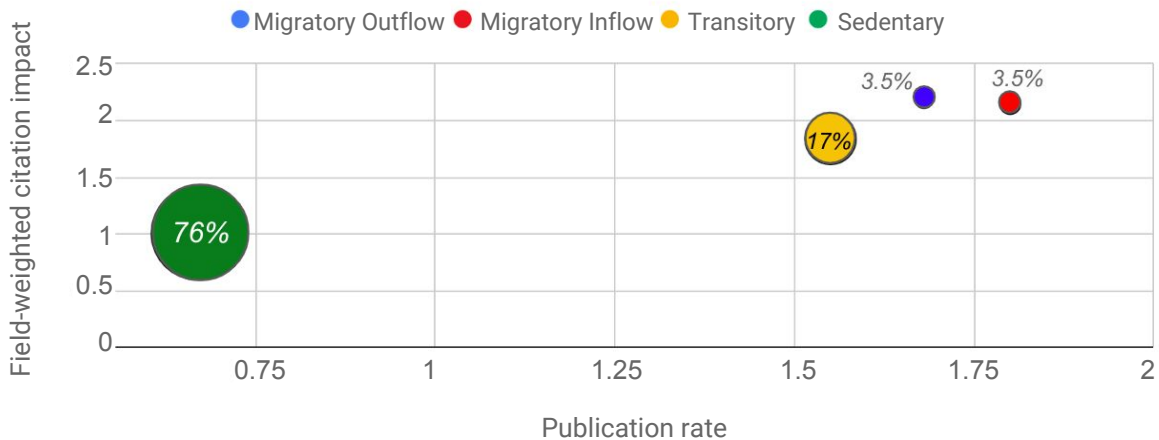The graph below shows the number of submitted and accepted papers for the 2018 Association for the Advancement of Artificial Intelligence (AAAI) conference, by country. The 2018 AAAI conference was held in February 2018 in New Orleans, Louisiana.

About 70% of papers submitted to AAAI in 2018 were affiliated with the U.S. or China. While China had the largest number of submitted papers, the U.S. and China had nearly the same number of accepted papers, at 268 and 265, respectively. As a result, U.S.-affiliated papers received a 29% acceptance rate while China-affiliated papers received a 21% acceptance rate. German and Italian papers received the highest acceptance rates (41%), though they had fewer submissions.

See Appendix 4 for data and methodology. For attendance by conference over time see page 26.

Number of accepted and submitted papers — 2018 AAAI conference
Source: AAAI

Legend: Submitted, Accepted

Data values (Submitted / Accepted):
- China: 1,242 / 265
- U.S.: 934 / 268
- U.K.: 128 / 32
- Australia: 124 / 32
- Singapore: 109 / 40
- Japan: 99 / 30
- Hong Kong: 92 / 32
- India: 76 / 17
- Canada: 69 / 16
- Germany: 68 / 28
- South Korea: 60 / 14
- Israel: 43 / 10
- Italy: 34 / 14

Y-axis: Number of papers (0, 250, 500, 750, 1,000, 1,250)
X-axis: Affiliated country
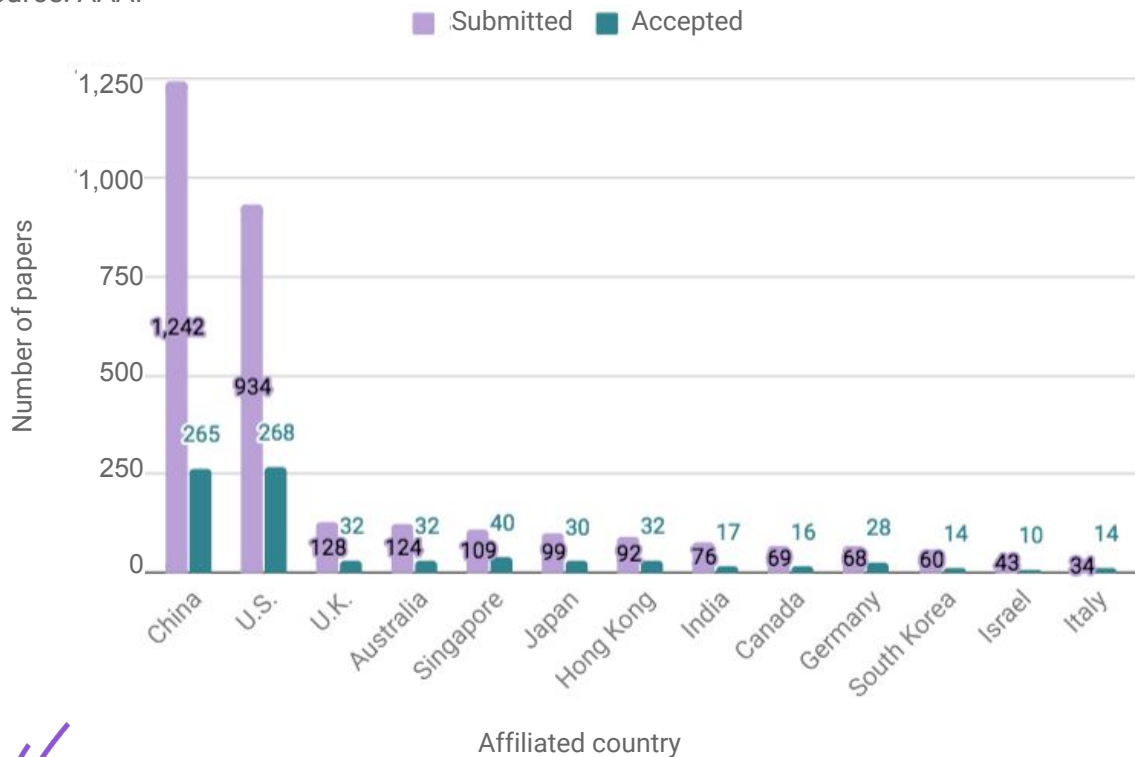
70% of AAAI papers are from the U.S. or China

At the 2018 AAAI conference, 70% of submitted papers and 67% of accepted papers were affiliated with the U.S. or China.

Course Enrollment: U.S. AI courses

The graphs below show the percentage of undergraduate students enrolled in introductory AI and ML courses. This is a calculated metric to account for school size. School selection criteria, actual enrollment numbers, and full university names can be found in the appendix.

While introductory AI courses tend to have a slightly larger proportion of undergraduate students than introductory ML courses (an average of 5.2% in AI vs 4.4% in ML), the number of undergraduate students in introductory ML courses are growing at a faster rate. See next page for growth comparison. This depicts the growing importance of machine learning as a subfield of AI.

See Appendix 5 for data and methodology.

Percent of undergraduates enrolled in Intro to AI (2010−2017)
Source: University provided data
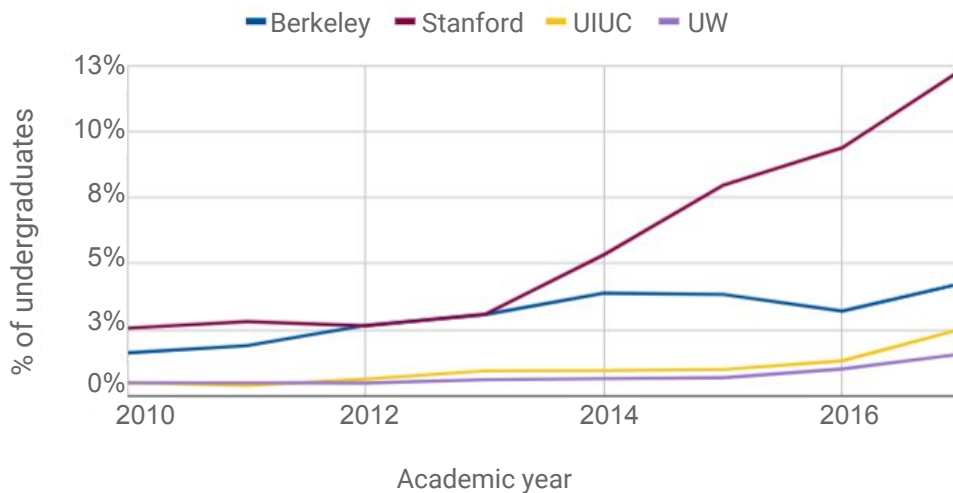
Percent of undergraduates enrolled in Intro to ML (2010−2017)
Source: University provided data
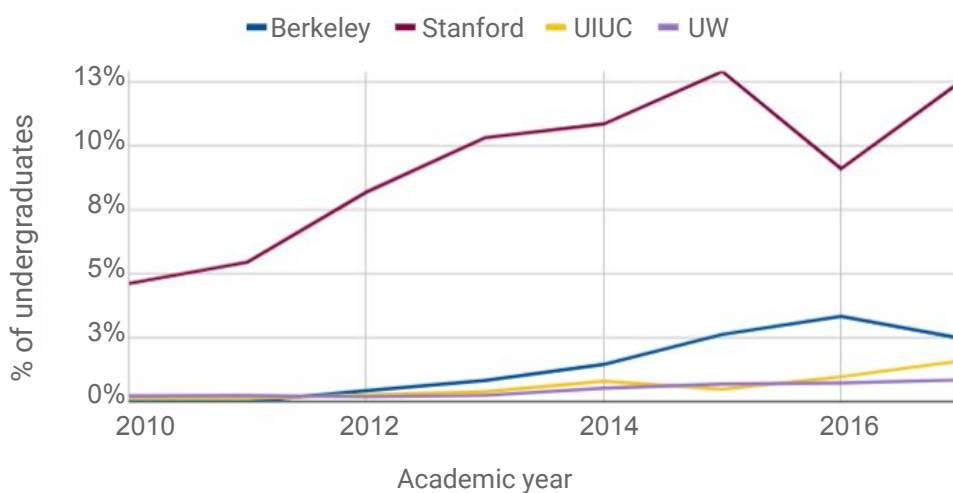
A note on gender diversity:
Stanford '17 Intro to AI: 74.45% male (869 students total); Berkeley '17 Intro to AI: 73.37% male (973 students total)
Stanford '17 Intro to ML: 75.91% male (876 students total); Berkeley '17 Intro to ML: 78.67% male (720 students total)

Course Enrollment: U.S. AI courses (continued)

The graphs below show the growth of AI and ML course enrollment at several leading computer science universities in the U.S. School selection criteria, actual enrollment numbers, and full university names can be found in the appendix.

2017 introductory AI enrollment was 3.4x larger than it was in 2012, while 2017 introductory ML course enrollment was 5x larger than it was in 2012. UC Berkeley's 2017 introductory ML course has 6.8x as many students as it had in 2012. That course grew more than any other introductory ML course we studied.

See Appendix 5 for data and methodology.

Growth in introductory AI course enrollment (2012−2017)
Source: University provided data
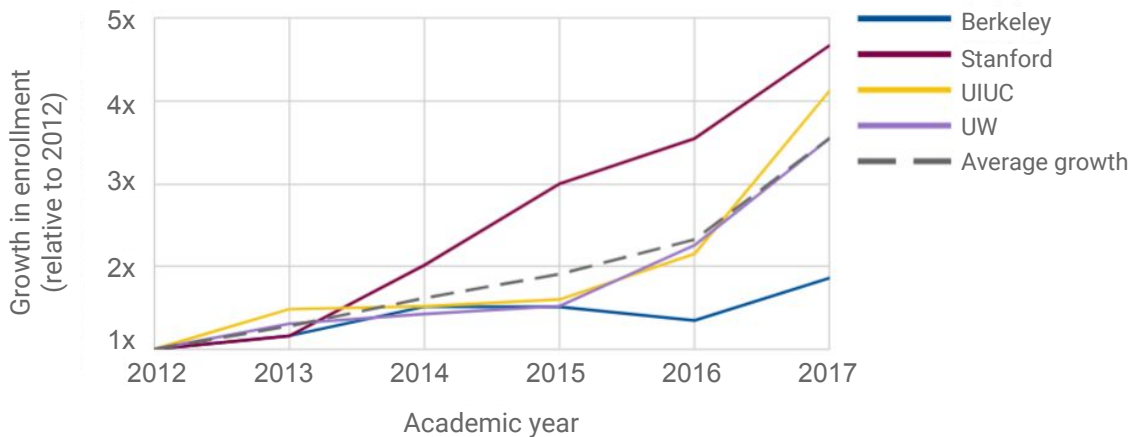
Growth in introductory ML course enrollment (2012−2017)
Source: University provided data
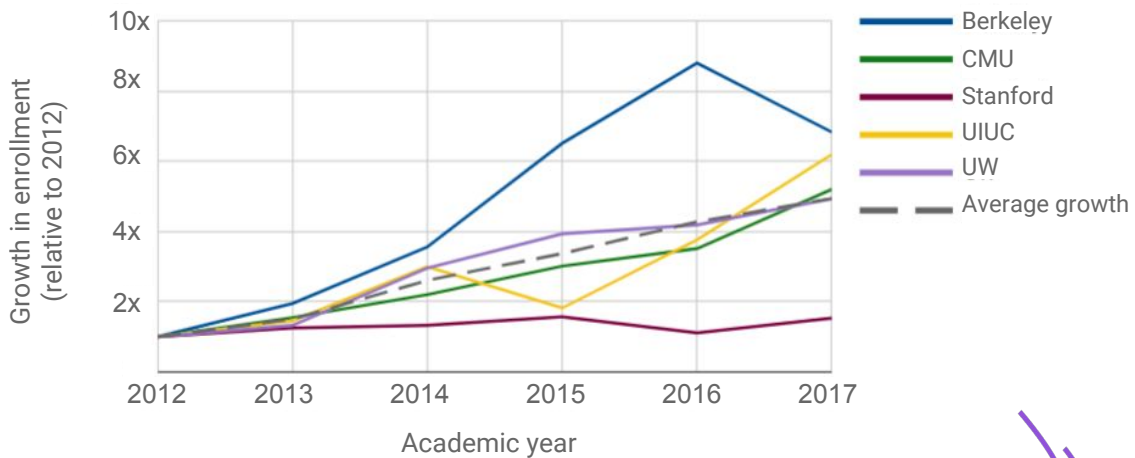
2017 introductory ML course enrollment was 5x greater than it was in 2012
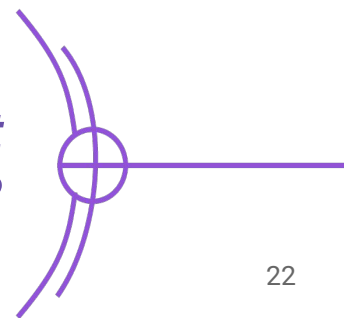
Course Enrollment: International courses

The graphs below and on the next page show AI and ML course enrollment at several leading computer science universities outside of the U.S. The first graph shows relative growth for international schools that provided data for academic years 2010 − 2017. The second graph shows relative growth for international schools that provided data for academic years 2016 − 2018. School selection criteria, actual enrollment numbers, and full university names can be found in the appendix.

Combined AI + ML course enrollment at Tsinghua University was 16x greater in 2017 than it was in 2010. Across the schools studied, we found that growth in AI course enrollment was relatively school dependent, and was not particularly influenced by geography. The AI Index looks forward to refining this hypothesis in future reports.

See Appendix 6 for data and methodology.

Growth of AI+ML course enrollment — Non-U.S. (2010−2017)
Source: University provided data
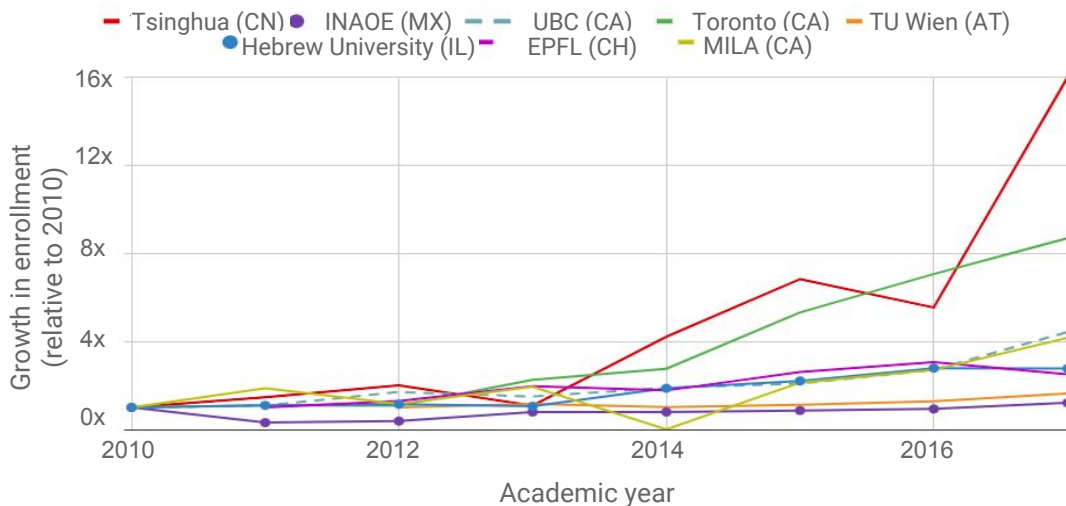
Note: This visual shows universities where historical years were available. All growth rates are relative to 2010 enrollment, with the exception of EPFL (2011) and TU Wien (2012).

Course Enrollment: Intl. courses (continued)

See description on .

Growth of AI course enrollment — Non-U.S. (2016—2018)
Source: University provided data
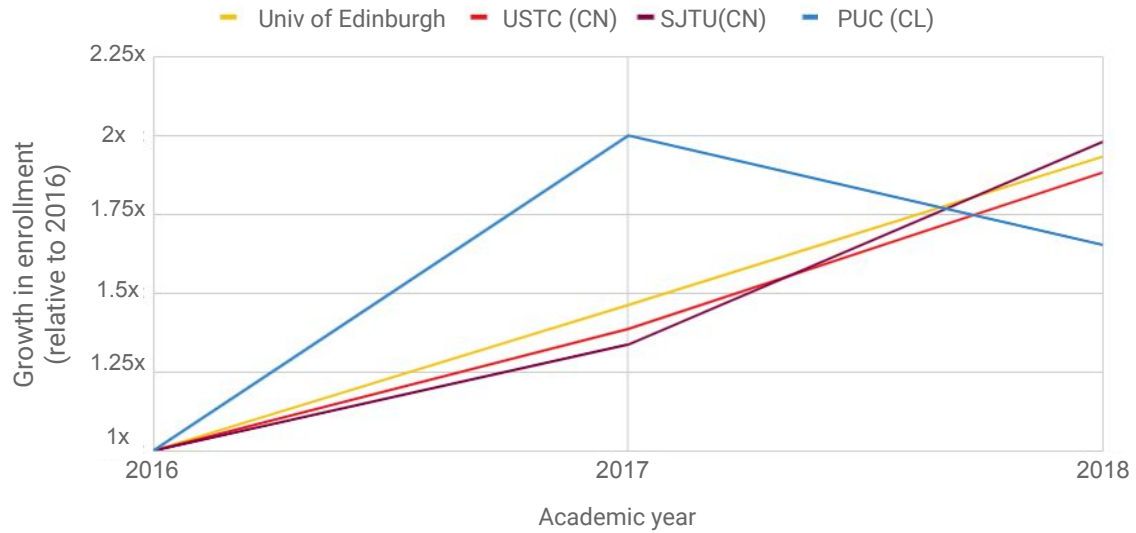
Note: This visual shows universities where historical years were limited. All growth rates are relative to 2016 enrollment.

Faculty Diversity

The graph below shows the gender break down of AI professors at several leading computer science universities. Data was collected using faculty rosters on September 21, 2018. We selected schools with easily accessible AI faculty rosters.

A significant barrier to improving diversity is the lack of access to data on diversity statistics in industry and in academia. We encourage institutions to be more transparent about diversity statistics in an effort to improve diversity in the field.

Across the schools studied, we found that on average 80% of AI professors are male. We noticed little variation between schools, regardless of geography. Still, due to the limited number of schools studied, these findings are a small view of a much larger picture.

See Appendix 7 for data and methodology.

Gender break down of AI professors — Select schools (September, 2018)
Source: University faculty rosters
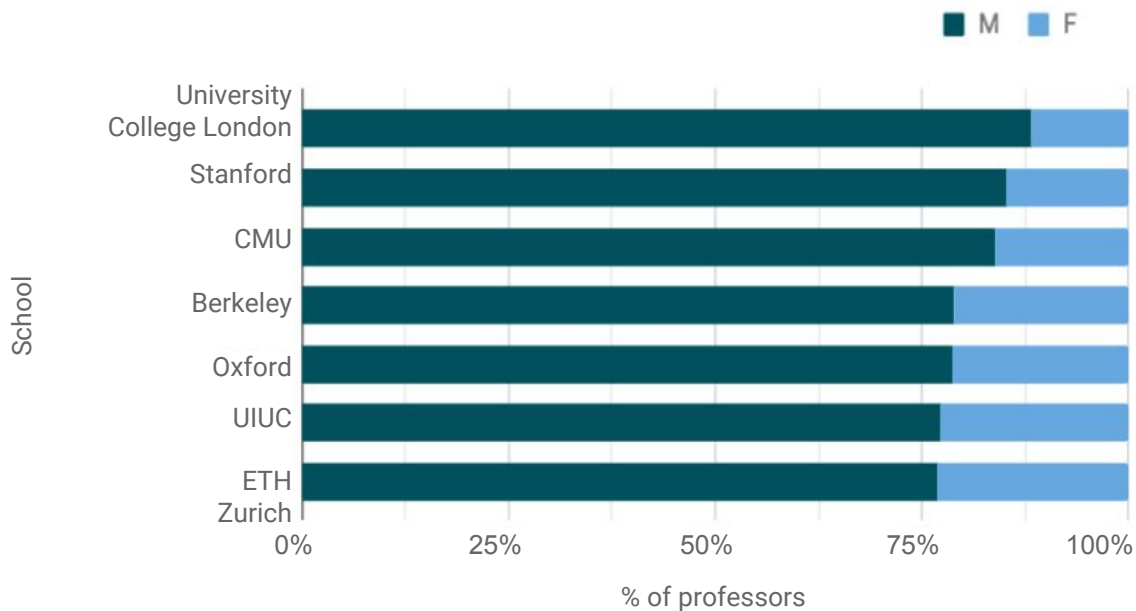
80% of AI professors are male

On average, 80% of professors from UC Berkeley, Stanford, UIUC, CMU, UC London, Oxford, and ETH Zurich are male

Participation: Large AI conferences

The graphs below show attendance at large AI conferences, and growth of large conference attendance relative to 2012. Large AI conferences are those with over two thousand attendees in 2017.

NeurIPS (originally NIPS), CVPR, and ICML, the most attended AI conferences, have experienced the most growth in attendance since 2012. NeurIPS and ICML are growing at the fastest rate — 4.8x and 6.8x their 2012 attendance, respectively. This shows continued interest in ML as a subfield of AI. Meanwhile, conferences focusing on symbolic reasoning continue to show little relative growth.

See Appendix 8 for data and methodology.

Attendance at large conferences (1984—2018)
Source: Conference provided data
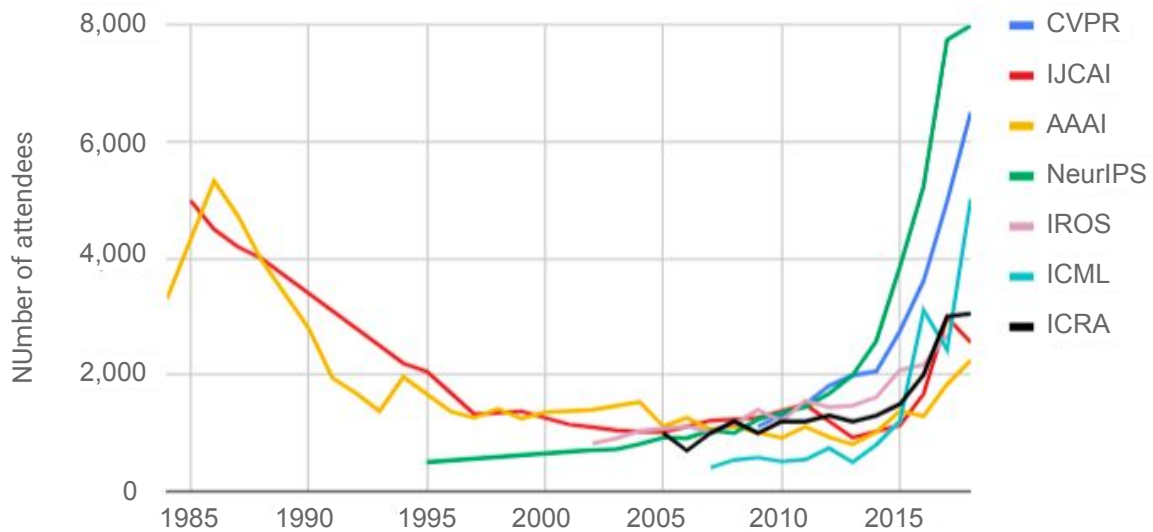
Growth of large conference attendance (2012—2018)
Source: Conference provided data
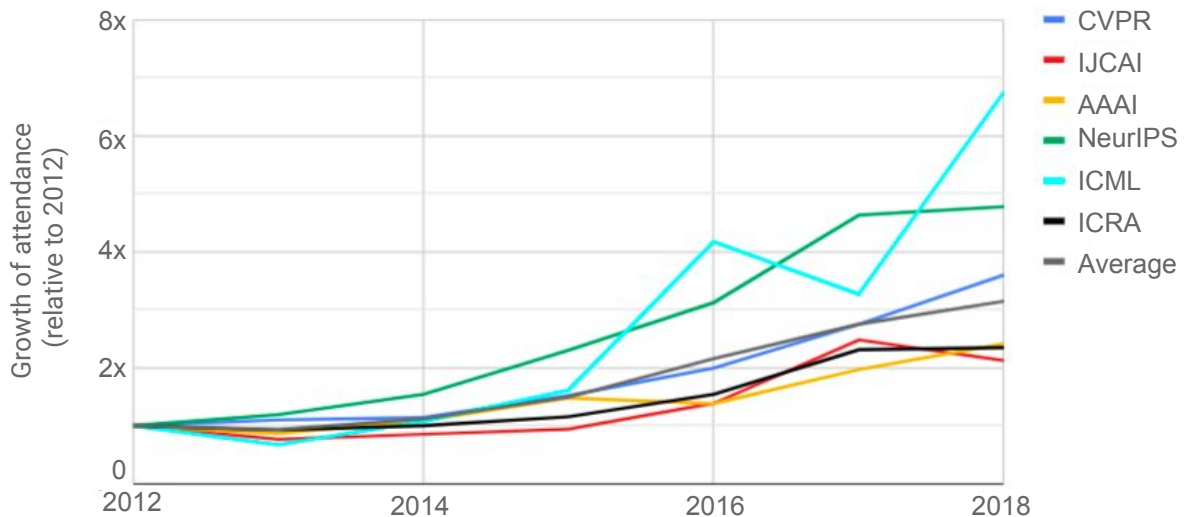
Participation: Small AI conferences

The graphs below show attendance at small AI conferences, and growth of small AI conference attendance relative to 2012. Small AI conferences are those with under two thousand attendees in 2017.

ICLR's 2018 attendance is 20x greater than it was in 2012. This increase is likely a result of a greater focus on deep and reinforcement learning within AI today.

See Appendix 8 for data and methodology.

Attendance at small conferences (1995−2018)
Source: Conference provided data
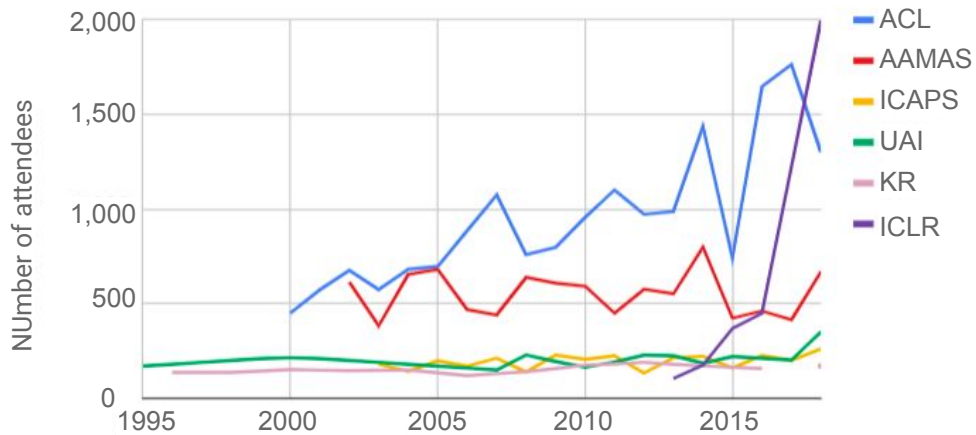
Note: 2018 was the first year that KR held a workshop. For consistency, workshop attendees are not included in KR's attendance count in the visual above.

Growth of small conference attendance (2012−2018)
Source: Conference provided data
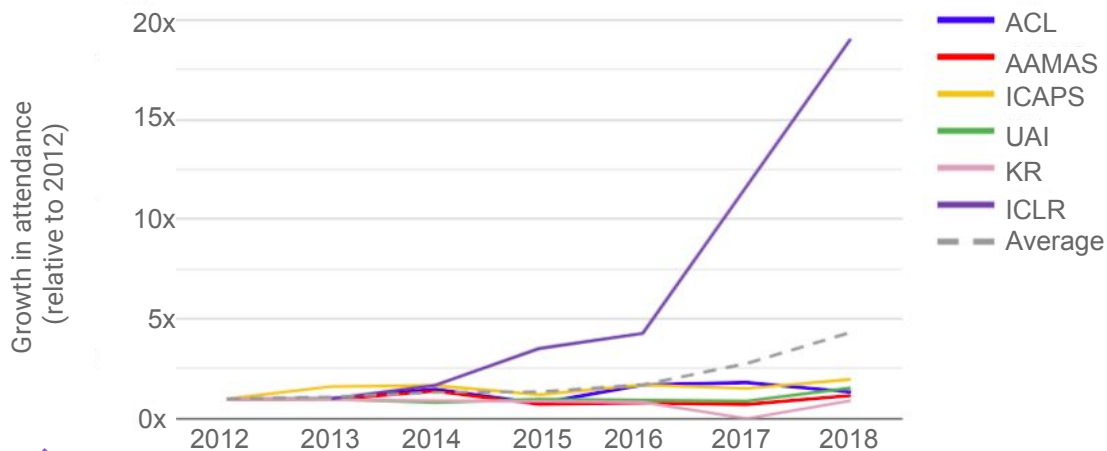
ICLR's 2018 conference attendance is 20x larger than it was in 2012

Participation: Diversity organizations

The graphs below show the number of registrations for the annual workshop hosted by Women in Machine Learning (WiML), an organization dedicated to supporting women in machine learning, and the number of alumni of AI4All, an AI education initiative designed to increase diversity and inclusion in AI.

Both the WiML workshop and AI4All have seen increased program enrollment over the past several years. The WiML workshop has 600% more participants than it had in 2014 and AI4ALL has 900% more alumni than it had in 2015. These increases show a continued effort to include women and underrepresented groups in the AI field.

See Appendix 8 for data and methodology.

AI4ALL alumni and programs (2015–2018)
Source: AI4ALL
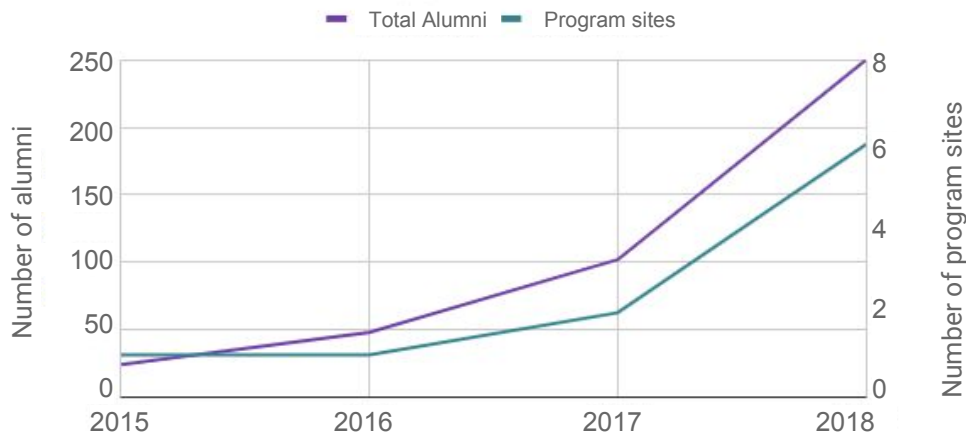
WiML workshop registration (2006–2018)
Source: WiML
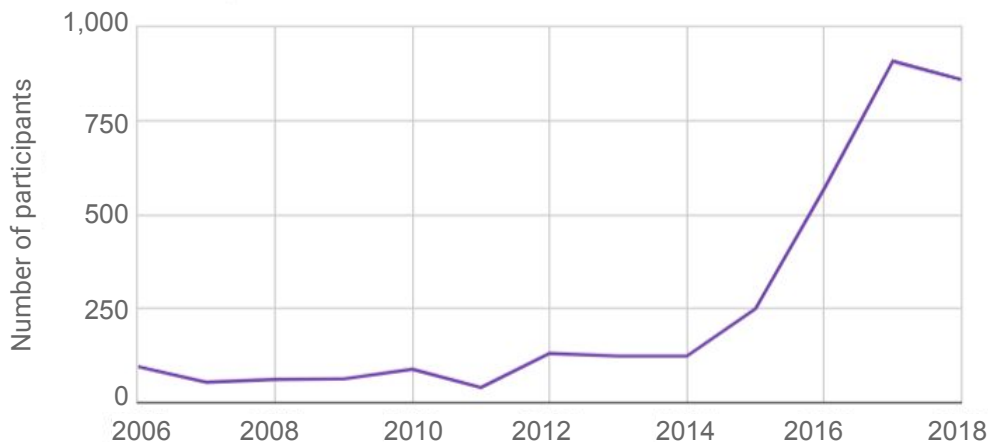
Note: WiML workshop registration was slightly inflated in 2017 due to a 2-day workshop, rather than the 1-day format of past years.

"…This data reflects the growth we've seen in the WiML community in recent years and, we think, correlates to increased attention to intentionally diversifying the communities within AI. But we have a long way to go, since attendance for WiML events still numbers in the hundreds, while overall attendance at commonly co-located conferences like NeurIPS is in the multiple thousands…"
WiML Board

Robot software downloads

The graph belows shows the number of robot operating system (ROS) binary packages downloaded from ROS.org over time. ROS is a widely used open source software stack for robotics. ROS is used by many commercial manufacturers and academic researchers. The left-axis shows the average number of monthly downloads in total, while the right-axis shows average monthly downloads from unique IP addresses only.

Since 2014, total downloads and unique downloads have increased by 352% and 567%, respectively. This represents an increased interest in both robotics and the use of robot systems. Because the number of unique downloads is growing at a faster rate than the total number of downloads, we can infer that there are more ROS users, not just that ROS is more frequently used.

See Appendix 9 for data and methodology.

ROS package downloads (2011—2018)
Source: ROS metrics report
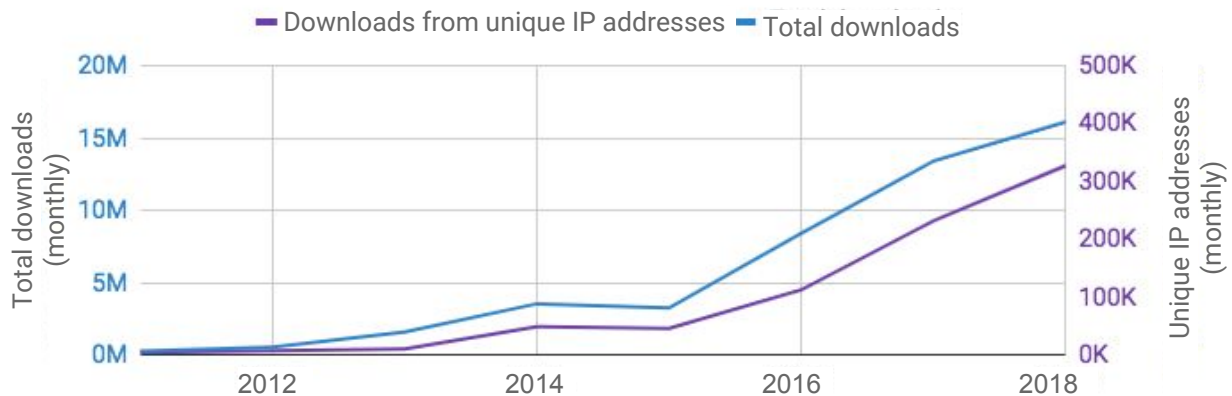
Since 2014, unique ROS package downloads from ROS.org have increased by 567%

Robot software downloads

The graph below shows the five largest regions by number of ROS.org pageviews since 2012.

The U.S. and Europe have the highest volume of ROS pageviews. China is not far behind after experiencing the most growth of any large region. The number Chinese views in 2017 was 18x larger than the number of Chinese views in 2012. ROS.org notes that growth in China is organic, and not a result of increased marketing or other resources in China.

See Appendix 9 for data and methodology.

ROS.org page views (2012—2017)
Source: ROS metrics report
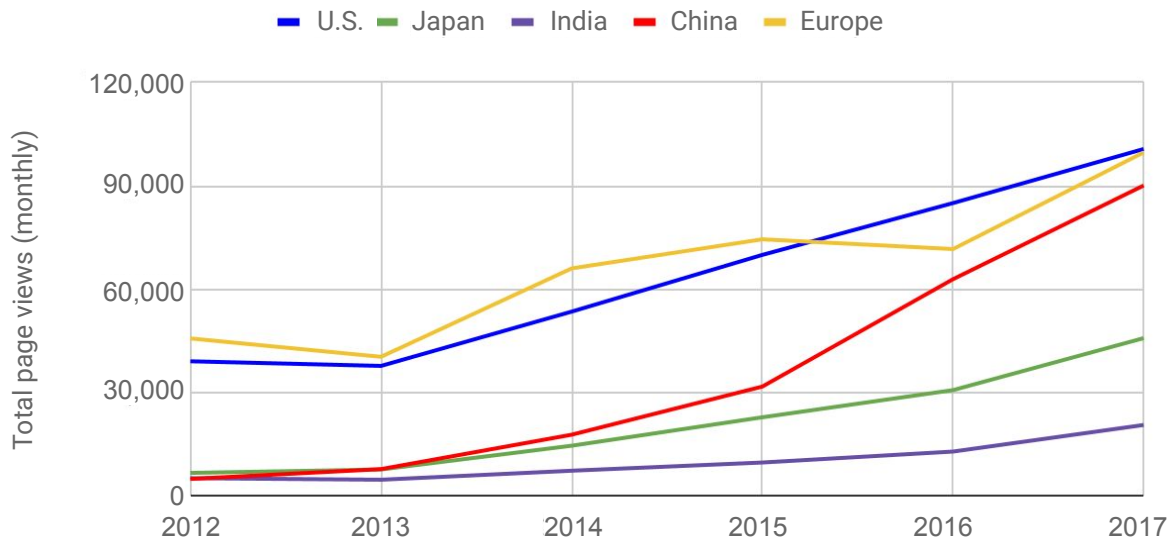
ROS.org pageviews in China were 18x greater in 2017 than in 2012
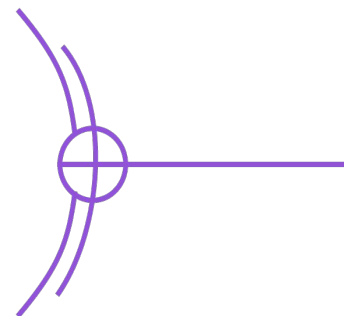
Startups / Investments: AI startups

The graph below shows the number of active venture-backed U.S. private startups in a given year. The blue line (left-axis) shows AI startups <u>only</u>, while the grey line (right-axis) shows all venture-backed startups, including AI startups. The graph plots the total number of startups in January of each year. Excluding instances where startups are removed from the data set (see appendix for details), the number of startups is cumulative year-over-year.

From January 2015 to January 2018, active <u>AI</u> startups increased 2.1x, while all active startups increased 1.3x. For the most part, growth in all active startups has remained relatively steady, while the number of AI startups has seen exponential growth.

See Appendix 10 for data and methodology.

AI startups (U.S., January 1995 — January 2018)
Source: Sand Hill Econometrics
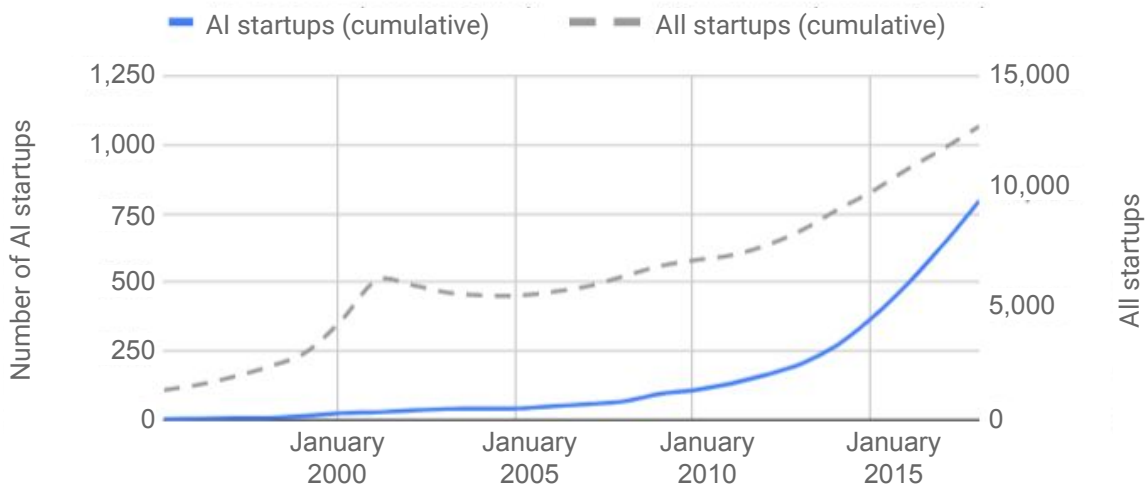
Note: The majority of the AI startups above develop AI systems. A minority use AI as an integral part of business, but do not develop the systems themselves. See appendix for more details.

Active AI startups in the U.S. increased 2.1x from 2015 to 2018
Meanwhile, active startups as a whole increased 1.3x.

Startups / Investments: VC funding

The graph below shows the amount of annual funding by Venture Capital firms (VCs) into active U.S. startups across all funding stages. The blue line (left-axis) shows funding of AI startups <u>only</u>, while the grey line (right-axis) shows funding of all venture-backed startups, including AI startups. The data are annual, and not cumulative year-over-year like the data on the previous page.

From 2013 to 2017, AI VC funding increased 4.5x while all VC funding increased 2.08x. The 1997 — 2000 boom in all VC funding can be explained by the dot-com bubble. The smaller booms in 2014 and 2015 reflect a period of relatively large economic growth.

See Appendix 10 for methodology.

Annual VC funding of AI startups (U.S., 1995 — 2017)
Source: Sand Hill Econometrics
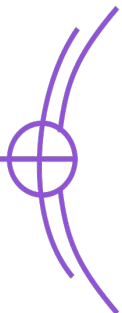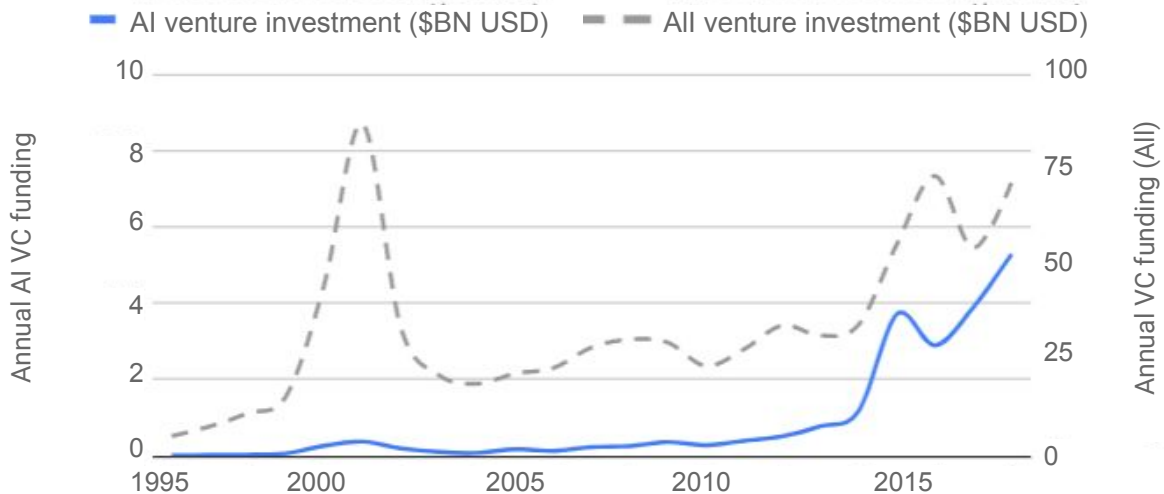
VC funding for U.S. AI startups increased 4.5x from 2013 to 2017
Meanwhile, VC funding for all active startups increased 2.08x.

Jobs: Openings by AI skill

The graphs below show the number of job openings per year by AI skill required, and the relative growth of job openings by AI skill required. AI skills are not mutually exclusive.

While ML is the largest skill cited as a requirement, deep learning (DL) is growing at the fastest rate — from 2015 to 2017 the number of job openings requiring DL increased 35x.

See Appendix 11 for methodology.

Job openings by AI skills required (2015 — 2017)
Source: Monster.com
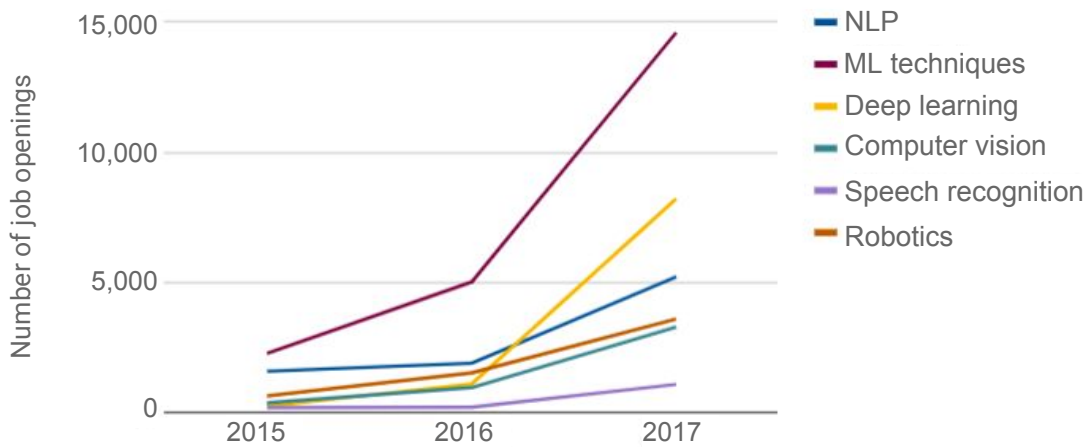
Legend: NLP, ML techniques, Deep learning, Computer vision, Speech recognition, Robotics

Growth of job openings by AI skills required (2015 — 2017)
Source: Monster.com

Legend: NLP, ML techniques, Deep learning, Computer vision, Speech recognition, Robotics
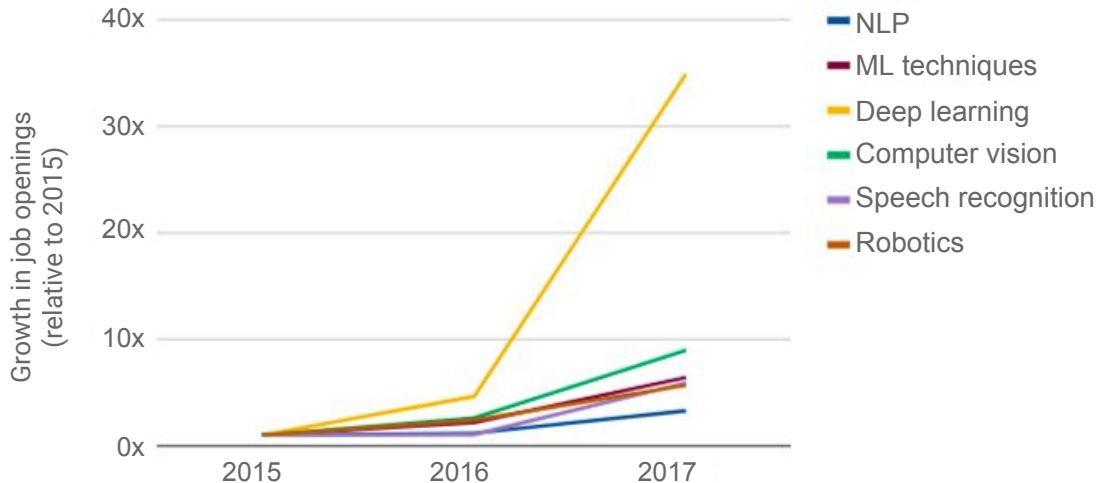
Note: While AI job openings are increasing across the board, they still represent a very low number of job openings for computer engineers.

Jobs: Gender diversity of applicants

The graph below shows male and female applicants for AI job openings in 2017. The data are by skill required, which is not mutually exclusive. The number of applicants does not imply hires or overall representation in industry.

On average, men make up 71% of the applicant pool for AI jobs in the U.S. Because the *Machine learning* requirement has the highest volume of applicants, the average is largely driven by *Machine Learning* job applicants. In addition to *Machine learning*, *Deep learning* and *Robotics* are more gender diverse, relative to other categories.

See Appendix 11 for data and methodology.

Job applicants by gender (2017)
Source: Gartner TalentNeuron
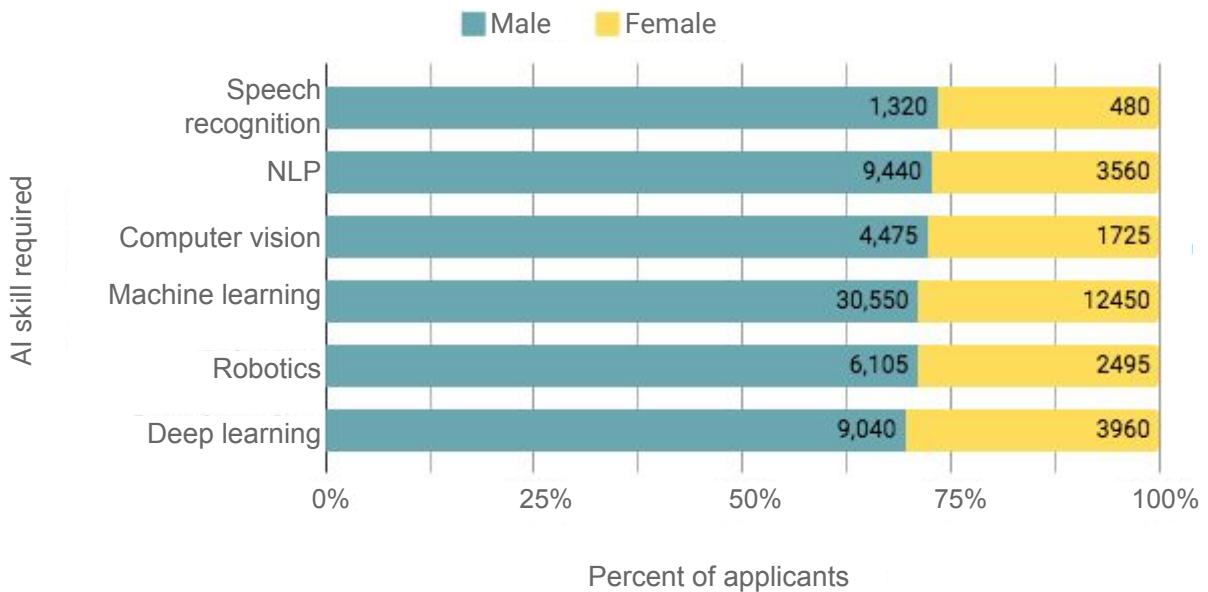
AI skill required	Male	Female
Speech recognition	1,320	480
NLP	9,440	3560
Computer vision	4,475	1725
Machine learning	30,550	12450
Robotics	6,105	2495
Deep learning	9,040	3960

Percent of applicants

On average, men make up 71% of the applicant pool for AI jobs in the U.S.
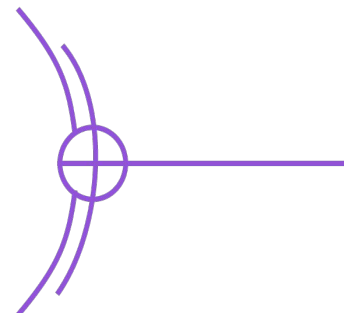
Patents

The graphs below show the number and growth of AI patents, by inventor region. AI patents were aggregated using IPC codes that fall into the *Cognition and meaning understanding* and *Human-interface technology* areas. Tracking patents over time is difficult. Please refer to the appendix for notes and nuances relating to this metric.

In 2014, about 30% of AI patents originated in the U.S, followed by South Korea and Japan, which each hold 16% of AI patents. Of the top inventor regions, South Korea and Taiwan have experienced the most growth, with the number of AI patents in 2014 nearly 5x that in 2004.

See Appendix 12 for data and methodology.

AI patents by inventor region (2004—2014)
Source: amplified
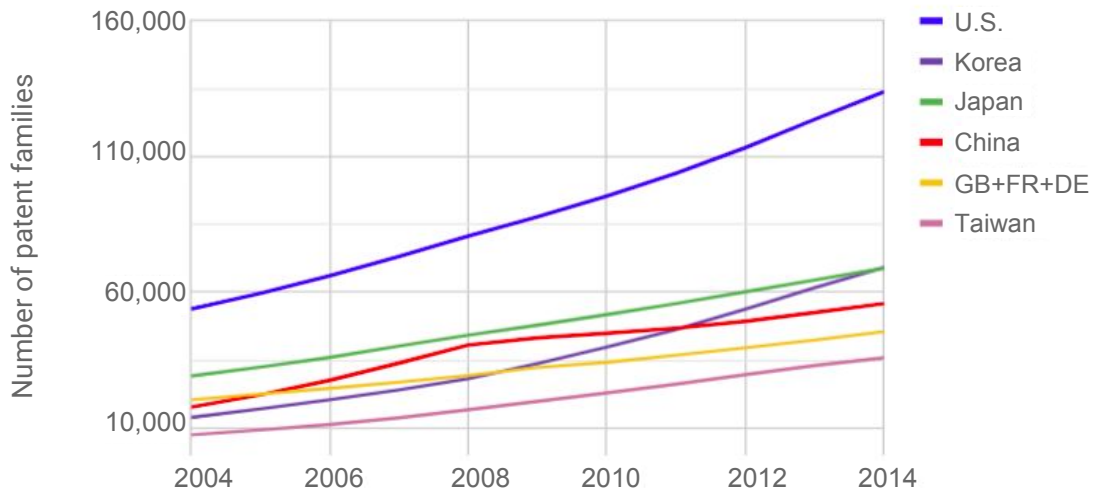
Growth of AI patents by inventor region (2004—2014)
Source: amplified
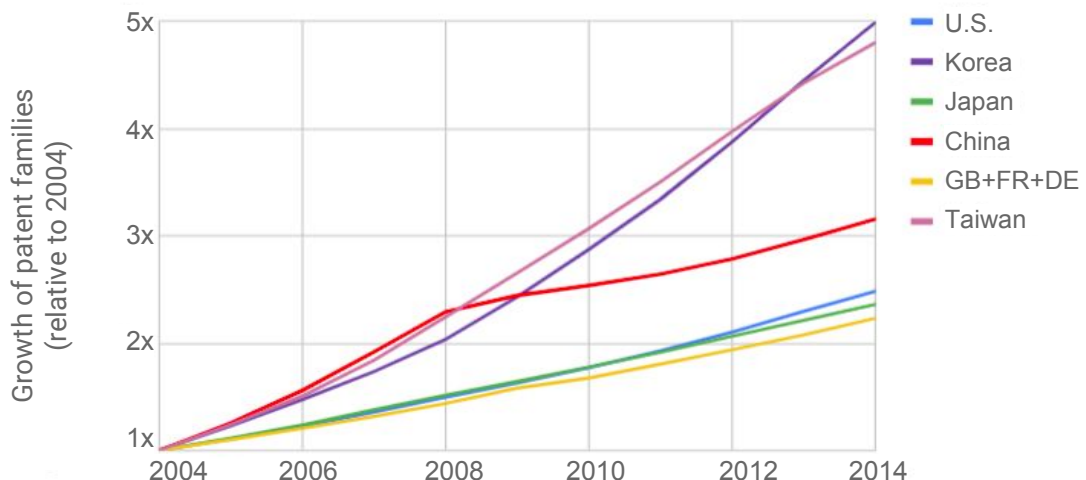
Note: GB + FR + DE refers to a combined number of patents from Great Britain, France, and Germany

AI adoption: Capabilities by region

The graphs below and on the following page show the results of a McKinsey & Company survey of 2,135 respondents, each answering on behalf of their organization. The graph displays the percent of respondents whose organizations have embedded AI capabilities in at least one function or business unit. Respondents can select multiple AI capabilities. See data for *Asia Pacific*, *India*, *Middle East* and *North Africa*, and *Latin America* on the .

While some regions adopt certain capabilities more heavily than others, AI capabilities are adopted relatively equally across regions. We look forward to tracking how company adoption changes over time.

See Appendix 13 for data and methodology.

Capabilities embedded in at least one company function (2018)
Source: McKinsey & Company

Percent of respondents
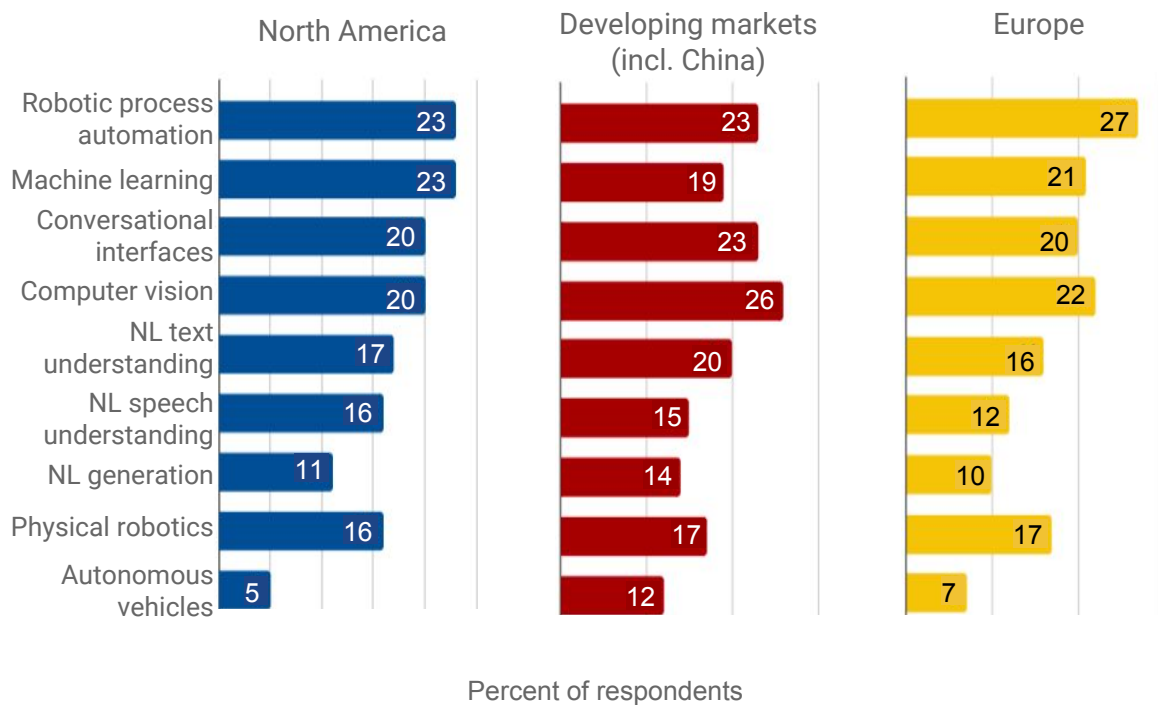
Note: The size of each bar is relative to the capabilities within each region; North America: N = 479; Developing markets (incl. China): N = 189 (China N = 35); Europe: N = 803

"…We found widespread adoption of different AI technologies across sectors, functions, and geographies around the world; about half of all companies had embedded AI into a corporate business process. However, it's still early; most had not yet adopted the complementary practices necessary to capture value from AI at scale…."
  - *Michael Chui, McKinsey & Company and the McKinsey Global Institute*

AI adoption: Capabilities by region (continued)

See description and data from *North America*, *Developing markets (incl. China)*, and *Europe* on the .

Capabilities embedded in at least one company function (2018)
Source: McKinsey & Company
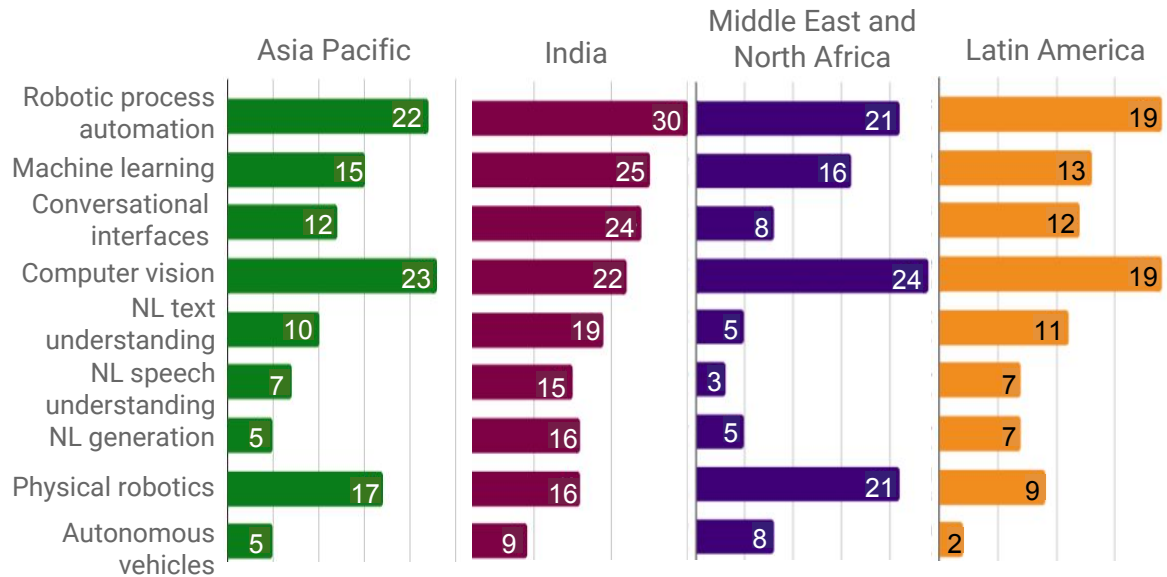
Capability	Asia Pacific	India	Middle East and North Africa	Latin America
Robotic process automation	22	30	21	19
Machine learning	15	25	16	13
Conversational interfaces	12	24	8	12
Computer vision	23	22	24	19
NL text understanding	10	19	5	11
NL speech understanding	7	15	3	7
NL generation	5	16	5	7
Physical robotics	17	16	21	9
Autonomous vehicles	5	9	8	2

Percent of respondents

Note: The size of each bar is relative to the capabilities within each region; Asia–Pacific: N = 263; India: N = 197; Middle East and North Africa: N = 77; Latin America: N = 127

AI adoption: Industry and function

The graphs below and on the following page show the results of a McKinsey & Company survey of 2,135 respondents, each answering on behalf of their organization. The graph shows the percent of respondents whose organizations have piloted or embedded AI capabilities within a particular business function. Respondents can select multiple functions. See data for *Manufacturing*, *Supply-chain management*, and *Risk* on the next page.

Organizations tend to incorporate AI capabilities in functions that provide the most value within their industry. For example, *Financial services* has heavily incorporated AI in *Risk*, while *Automotive* has done so in *Manufacturing*, and *Retail* has done so in *Marketing / sales*. This implies that the rate of AI progress for specific applications (e.g., *Manufacturing*) will likely correlate to uptake in industries where that specialization is particularly important.

See Appendix 13 for data and methodology.

AI adoption by industry and function (2018)
Source: McKinsey & Company
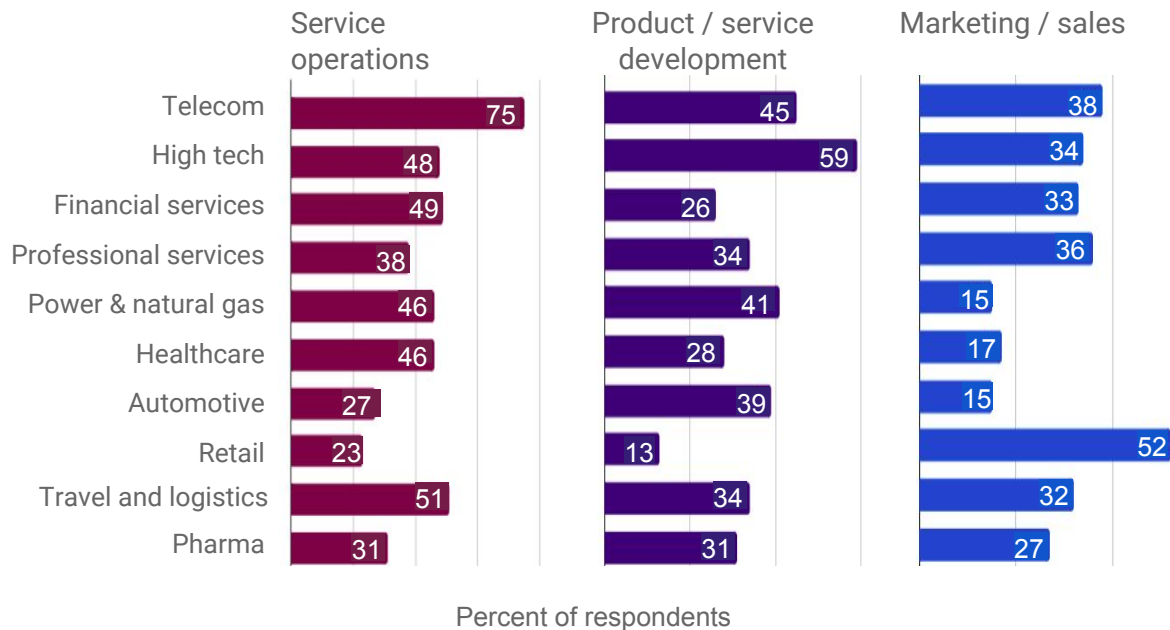
	Service operations	Product / service development	Marketing / sales
Telecom	75	45	38
High tech	48	59	34
Financial services	49	26	33
Professional services	38	34	36
Power & natural gas	46	41	15
Healthcare	46	28	17
Automotive	27	39	15
Retail	23	13	52
Travel and logistics	51	34	32
Pharma	31	31	27

Percent of respondents

Note: The size of each bar is relative to the industries within each function; Telecom: N = 77; High tech: N = 215; Financial services: N = 306; Professional services: N = 221; Electric power and natural gas: N = 54; Healthcare systems and services: N = 67; Automotive and assembly: N = 120; Retail: N = 46; Travel, transport, and logistics: N = 55; Pharma and medical products: N = 65.

AI adoption: Industry and function (continued)

See description and data from *Service operations*, *Product / service development*, and *Marketing / sales* on the .

AI adoption by industry and function (2018)
Source: McKinsey & Company
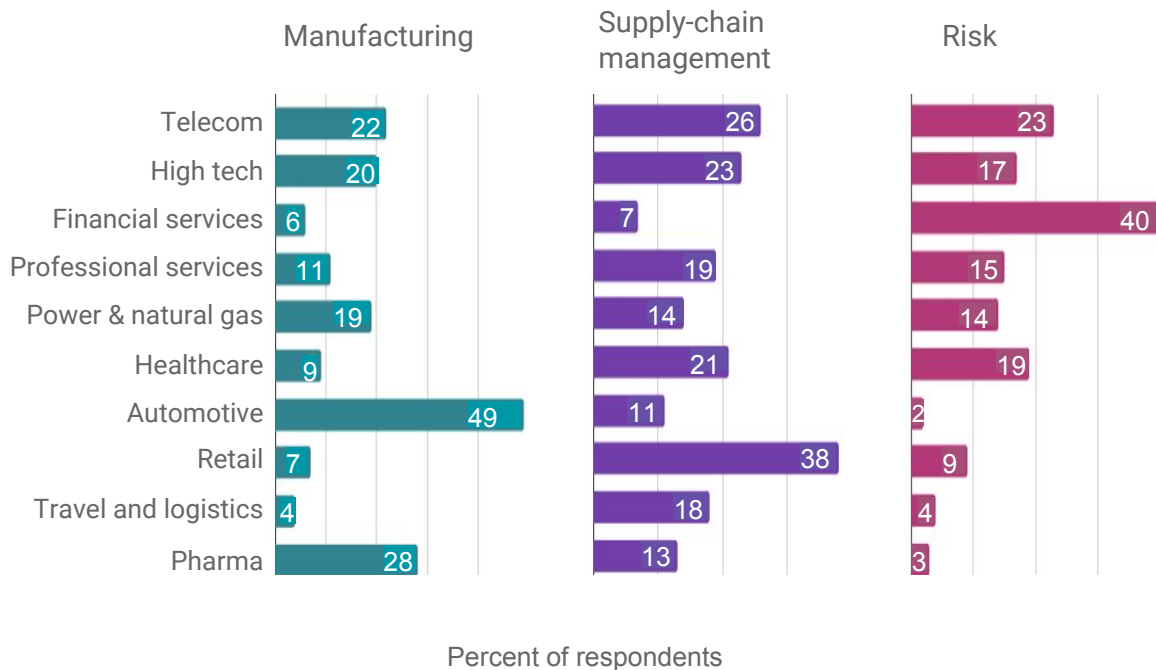
	Manufacturing	Supply-chain management	Risk
Telecom	22	26	23
High tech	20	23	17
Financial services	6	7	40
Professional services	11	19	15
Power & natural gas	19	14	14
Healthcare	9	21	19
Automotive	49	11	2
Retail	7	38	9
Travel and logistics	4	18	4
Pharma	28	13	3

Percent of respondents

Note: The size of each bar is relative to the industries within each function; Telecom: N = 77; High tech: N = 215; Financial services: N = 306; Professional services: N = 221; Electric power and natural gas: N = 54; Healthcare systems and services: N = 67; Automotive and assembly: N = 120; Retail: N = 46; Travel, transport, and logistics: N = 55; Pharma and medical products: N = 65.

Organizations adopt AI in business functions that provide the most value within their industry
This implies that the rate of AI progress for specific applications will likely correlate to uptake in industries where that specialization is particularly important.

Earnings calls mentions

The graphs below show *Artificial intelligence (AI)* and *Machine learning (ML)* mentions in company earnings calls, by industry. The first graph shows mentions only for companies in the IT sector, as this sector has a much larger scale of AI and ML mentions. The second graph shows *Artificial intelligence* mentions by industries other than IT. *Big data* and *Cloud* mentions are added to put *AI / ML* mentions in perspective. This analysis uses only companies publicly traded on the New York Stock Exchange.

AI and ML mentions increased in 2015 within the IT sector. For most other industries, this increase began in 2016. The *IT*, *Consumer Discretionary*, *Financial*, and *Health Care* sectors have the most mentions of AI on earnings calls.

See Appendix 14 for data and methodology.

Company earnings calls mentions — IT companies (2007—2017)
Source: Prattle
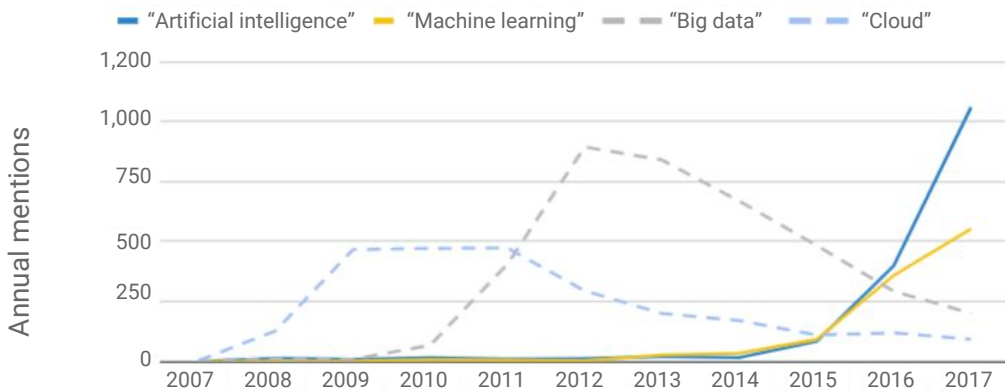
Company earnings calls mentions — Sum of other industries (2008—2017)
Source: Prattle
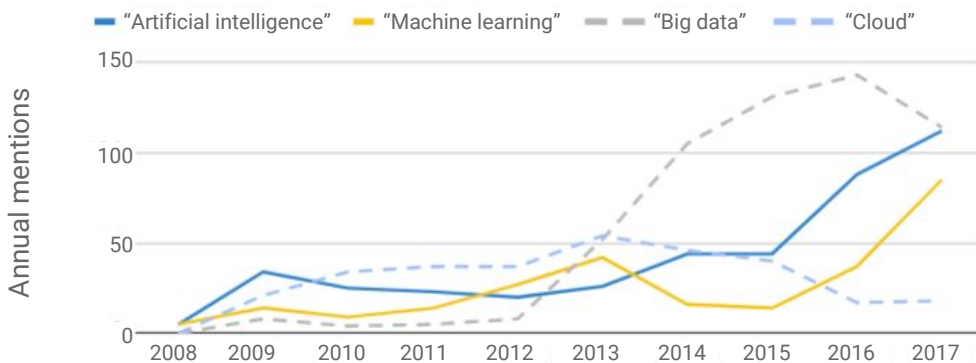
Top 10 companies — AI mentions (2017)
NVIDIA Corporation (93), LivePerson (50), Pegasystems (44), Facebook (40), Axon Enterprise (37), salesforce.com (36), Intel Corporation (35), Microsoft Corporation (34), Genpact Limited (34), Applied Materials (33)
Top 10 companies — ML mentions (2017)
Alphabet(57), Nasdaq (26), ServiceNow (21), Progress Software Corporation (21), Cadence Design Systems, Inc. (20), Splunk Inc. (18), Twitter (17), Overstock.com (14), Synopsys, Inc. (14), Aspen Technology (14)

Robot installations

The graphs below show annual installations of industrial robot units by region. The first visual shows the five largest regions by robot installations, while the second visual breaks out the 'Rest of world'.

China has seen a 500% increase in annual robot installations since 2012, while other regions, like South Korea and Europe, have increased 105% and 122%, respectively. Among smaller installers, Taiwan has the highest volume of annual installations today and has grown most from 2012—2017.

See Appendix 15 for data and methodology.
.

Robot Installations — Large regions (2012—2017)
Source: ifr.org
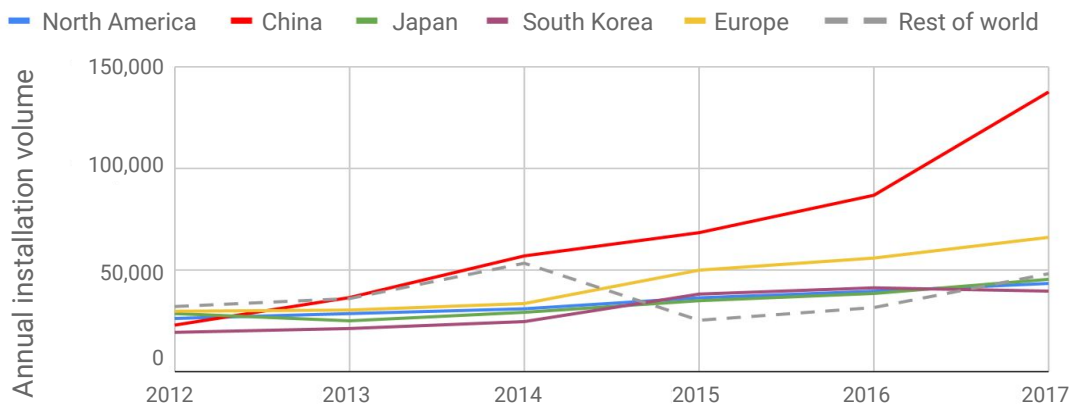
Robot Installations — Rest of world (2012—2017)
Source: ifr.org
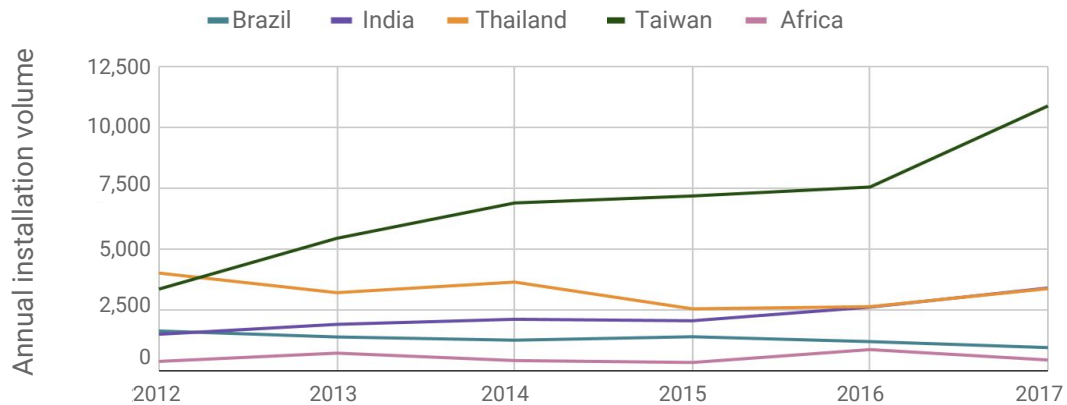
Note: The number of installations is domestic production plus imports minus exports.

GitHub stars

The graphs below show the number of times various AI and ML software packages have been starred on GitHub. This provides a rough measure of the popularity of various AI-programming frameworks.

Two recent trends are the growing popularity of frameworks backed by major companies *(i.e.,Tensorflow (Google), Pytorch (Facebook), mxnet (Amazon))*, and the significant popularity of TensorFlow relative to other languages.

See Appendix 16 for data and methodology.

Cumulative GitHub stars by AI library (2015—2018)
Source: GitHub
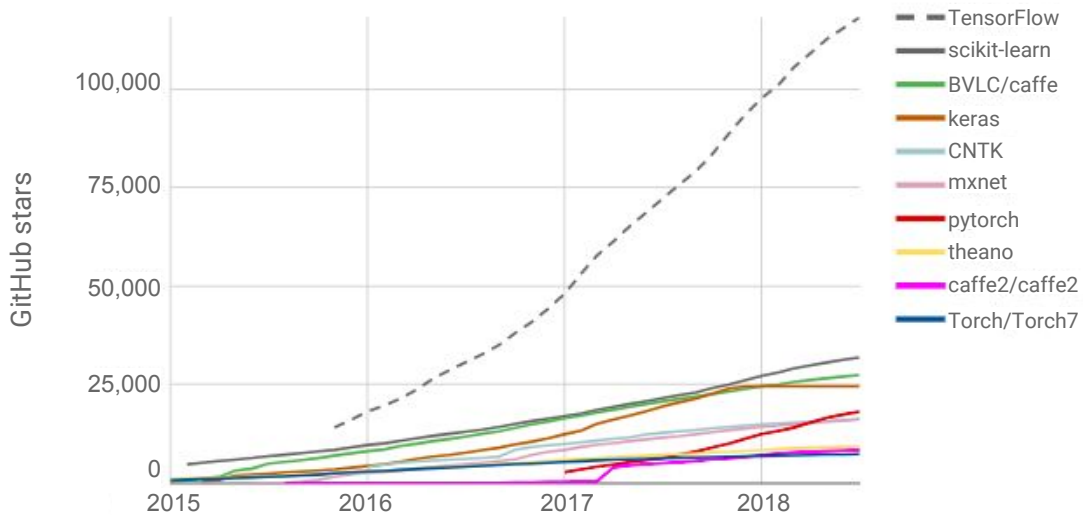
Cumulative GitHub stars by AI library — Excluding TensorFlow (2015—2018)
Source: GitHub
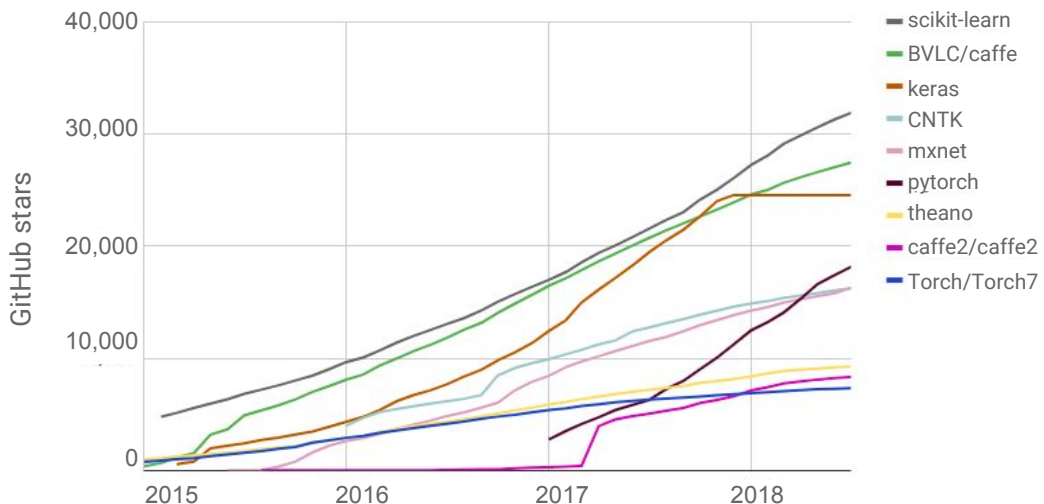
Sentiment of Media Coverage

The graph below shows the percent of popular media articles that contain the term *Artificial Intelligence* and that are classified as either *Positive*, *Negative*, or *Neutral* articles.

AI articles have become less neutral and more positive, particularly since early 2016, when articles went from 12% positive in January 2016 to 30% positive in July 2016. The percentage of positive articles has hovered near 30% since then.

See Appendix 17 for data and methodology.

Sentiment of articles referencing AI (2013—2018)
Source: TrendKite
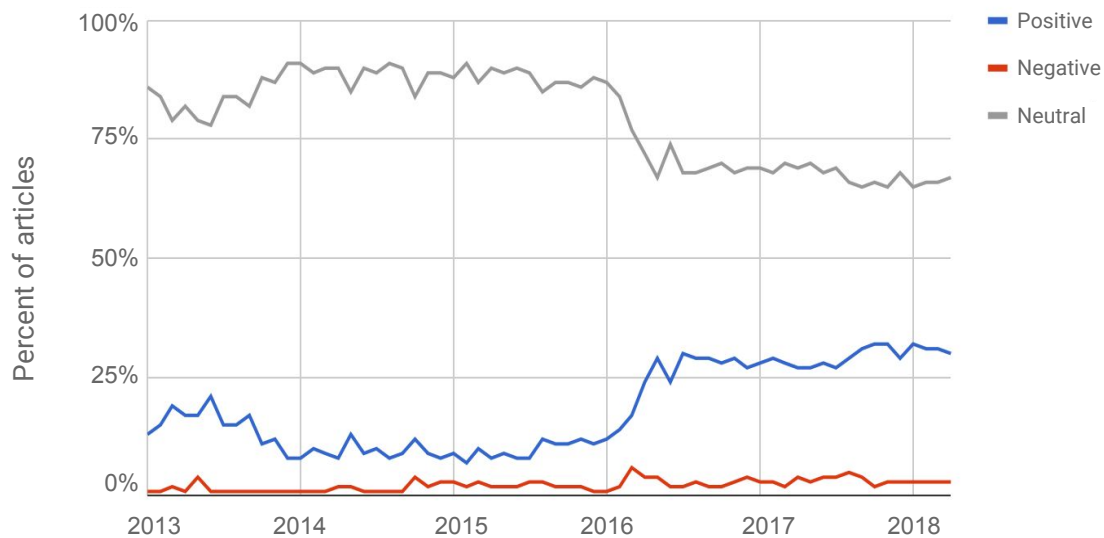
Articles on AI became 2.5x more positive from 2016 to 2018

Government mentions

The graphs below and on the following page show mentions of the terms *Artificial intelligence* and *Machine learning* in transcripts of the U.S. Congressional Record and the records of proceedings (known as Hansards) of the Parliaments of Canada and the United Kingdom.

Across the three governments, mentions of these terms have spiked since 2016. Additionally, in the three countries, *Machine learning* is rarely mentioned before 2016, and has remained a small fraction of total mentions, relative to *Artificial intelligence*.

Note that methodology differences make country-to-country comparisons difficult. For example, each count in the U.S. data indicates that *Machine Learning* or *Artificial Intelligence* was said at least once in a given <u>event or conversation</u>. Meanwhile, each count in the United Kingdom's data indicates that *Machine learning* or *Artificial intelligence* was said at least once in a given <u>comment</u>. So, rather than country-to-country comparisons, we suggest comparing trends over time within a country.

See the U.S below. Canada and the U.K. are in the following page. See Appendix 18 for data and methodology.

AI and ML mentions in U.S. Congress (1995—2018)
Source: U.S. Congressional Record website, McKinsey Global Institute analysis
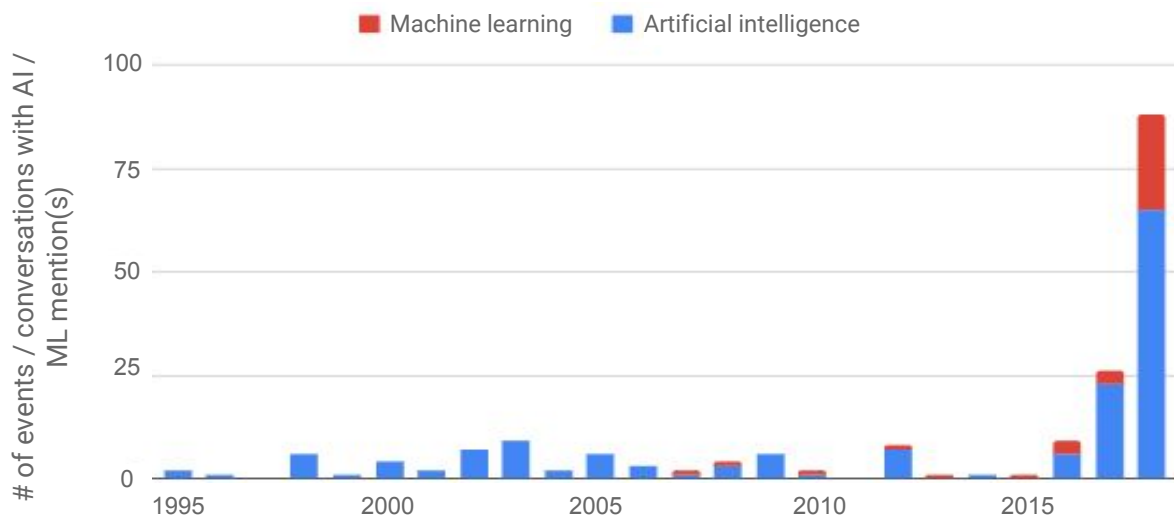
Note: The U.S. Senate and House of Representatives are included in this analysis. Data for 2018 is through 11/20/18.

Government mentions

AI and ML mentions in U.K. Parliament (1980—2018)
Source: Parliament of U.K. website, McKinsey Global Institute analysis
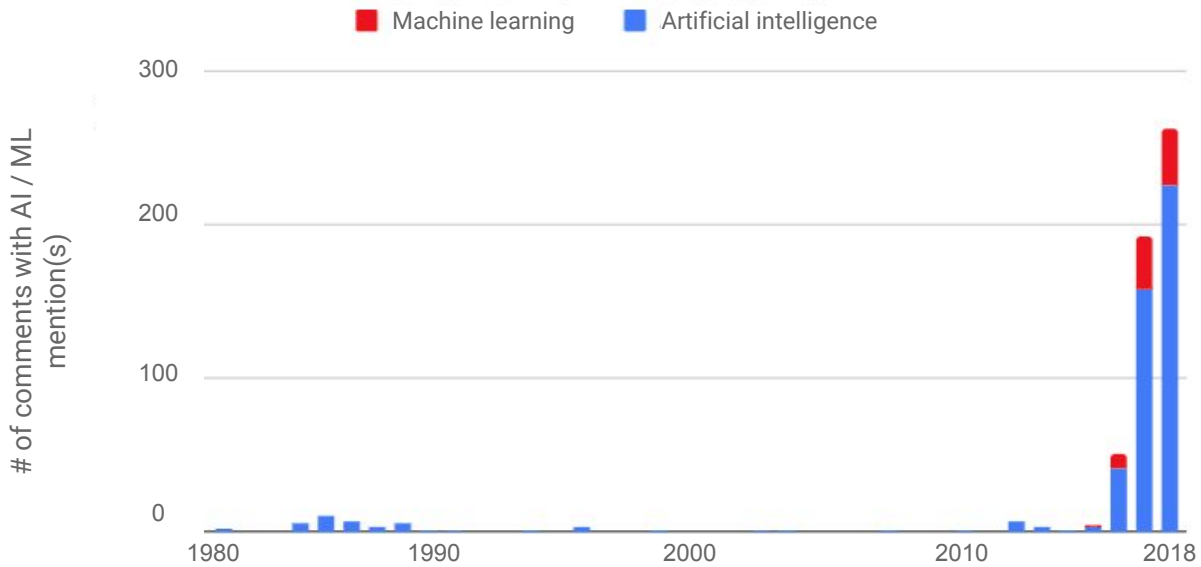
Note: The U.K. House of Commons, House of Lords, Westminster Hall, and Committees are all included in this analysis. Data for 2018 is through 11/20/18.

AI and ML mentions in Canadian Parliament (2002—2018)
Source: Parliament of Canada website, McKinsey Global Institute analysis
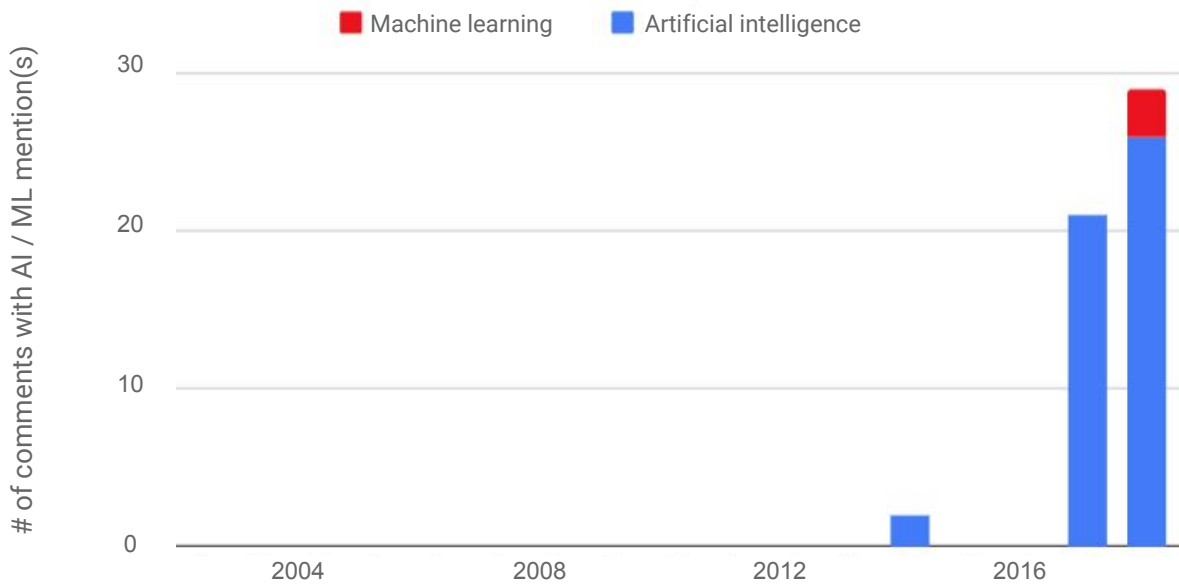
Note: Only the Canadian House of Commons is included in this analysis. Data for 2018 is through 11/20/18. Please note that this is an unofficial reproduction for non-commercial use. Readers seeking to use this chart should note that the Speaker's Permission does not extend to commercial purpose of financial gains

TECHNICAL PERFORMANCE

Object detection: ImageNet

The graph below shows ImageNet accuracy scores over time. The ImageNet competition ran until 2017 and scored models on a held-out competition-specific 'test' data set. Since the competition is no longer running, the AI Index has chosen to track continued progress in ImageNet through research papers, which test against the validation data set from the ImageNet 2012 corpus.

Because we are plotting results against the validation data set and without the competition, there is no longer a held-out test data set to assess models against. To give a sense of progress, we have highlighted prominent results against the validation set from the ImageNet 2012 corpus from 2012−2018. As this graph shows, ImageNet performance has continued to increase.

This metric also highlights a challenge inherent to modeling AI progress over time: if a certain research metric is built around a competition, then the retiring of that competition can make it challenging to derive true progress. However, due to the availability of open data sets, continuity can be assured given some finessing.

See Appendix 19 for data and methodology.

ImageNet (2010 −2018)
Source: ImageNet; see appendix
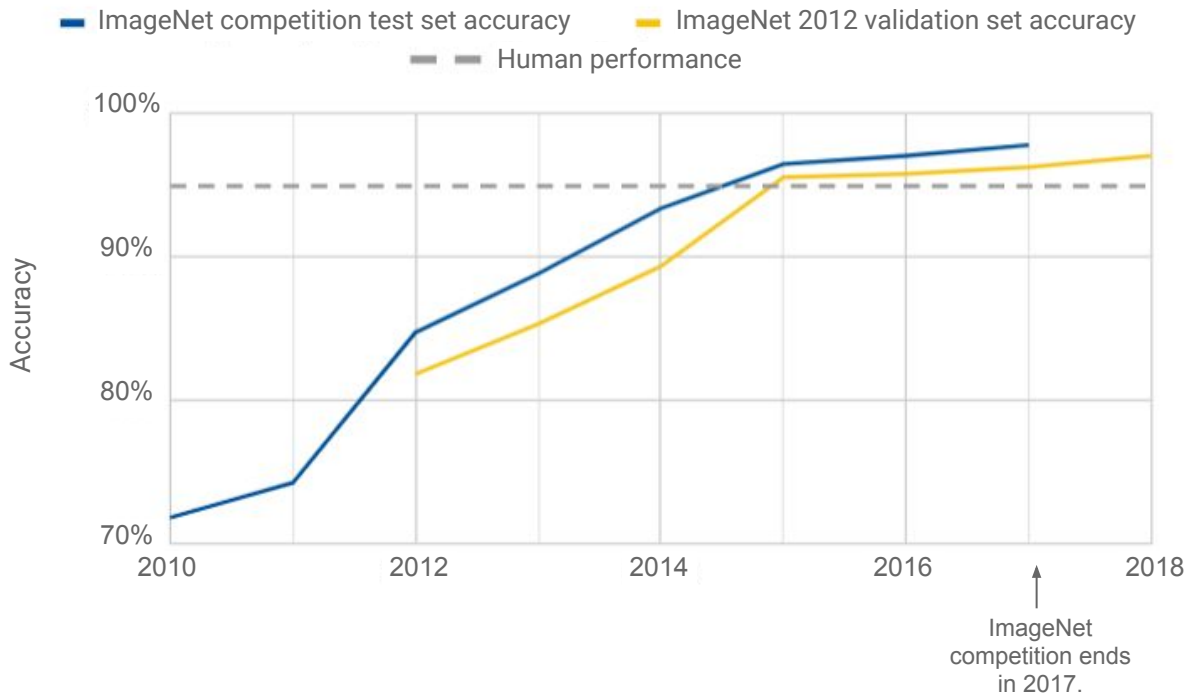
Object detection: ImageNet training time

The graph below shows the amount of time it takes to train a network to classify pictures from the ImageNet corpus (an image database) with a high degree of accuracy. This metric is a proxy for the time it takes well-resourced actors in the AI field to train large networks to perform AI tasks, such as image classification. Because image classification is a relatively generic supervised learning task, progress in this metric also correlates to faster training times for other AI applications. In a year and a half, the time required to train a network has fallen from about one hour to about 4 minutes.

The ImageNet training time metric also reflects the industrialization of AI research. The factors that go into reducing ImageNet training time include: algorithmic innovations and infrastructure investments (e.g., in the underlying hardware used to train the system, or in the software used to connect this hardware together)

See Appendix 19 for data and methodology.

ImageNet training time (June 2017 — November 2018)
Source: arXiv.org; see appendix for authors
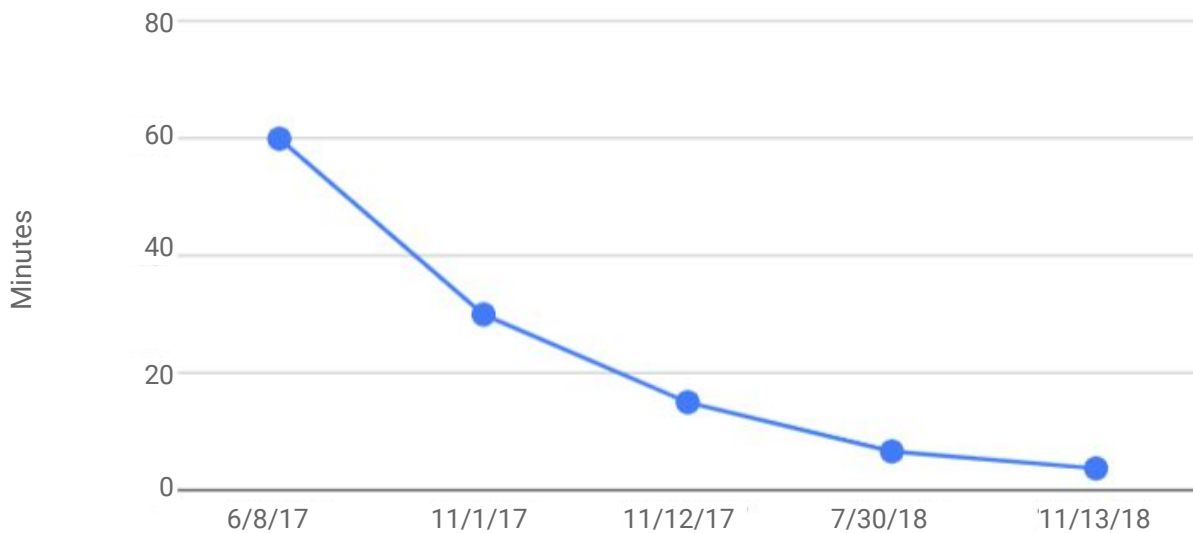
ImageNet training time became 16x faster between June 2017 and November 2018

Instance object segmentation: COCO

The graph below shows performance in the Common Objects in Context Challenge (COCO).

In 2017, the ImageNet challenge retired after computer vision algorithms obtained high performances in the object detection and image classification tasks provided by ImageNet. Since then, the research community moved on to harder computer vision tasks. The community shifted its focus toward vision tasks that require more sophisticated reasoning, such as localizing objects with pixel—level accuracy (known as object instance segmentation) and dividing a scene into regions with pixel-level precision (known as semantic segmentation).

The COCO metric shows performance in the detailed computer vision tasks mentioned above. The goal of the challenge is to build an algorithm that can localize each object of interest and delineate its boundaries accurately.

Since 2015, the highest average precision reached in the COCO challenge has increased by .2 points, or 72%.

See underlying data here, or learn more about the COCO challenge.

COCO instance segmentation challenge (2015—2018)
Source: COCO leaderboard

49
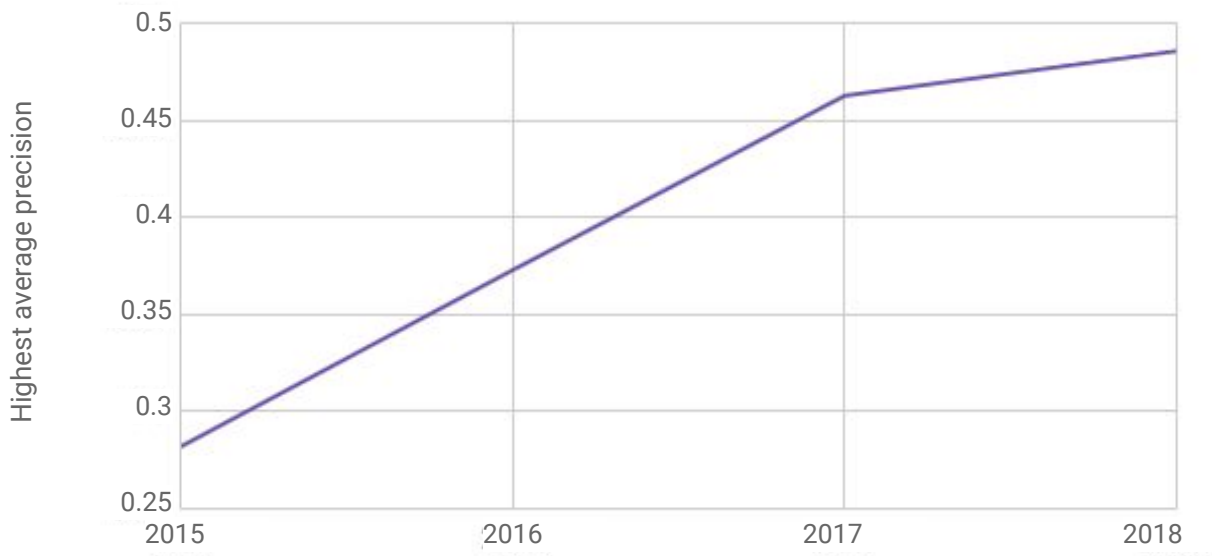
Parsing

The graph below shows the performance of AI systems on a task to determine the syntactic structure of sentences.

The parsing metric is the first step to understanding natural language in certain tasks, such as question answering. Originally done using the algorithms similar to those used for parsing programming languages, it is now almost universally done using deep learning. Since 2003, F1 scores for all sentences have increased by 9 percentage points (or a 10% increase).

See Appendix 20 for data and methodology.

Constituency parsing — Penn Treebank (1995—2018)
Source: See appendix
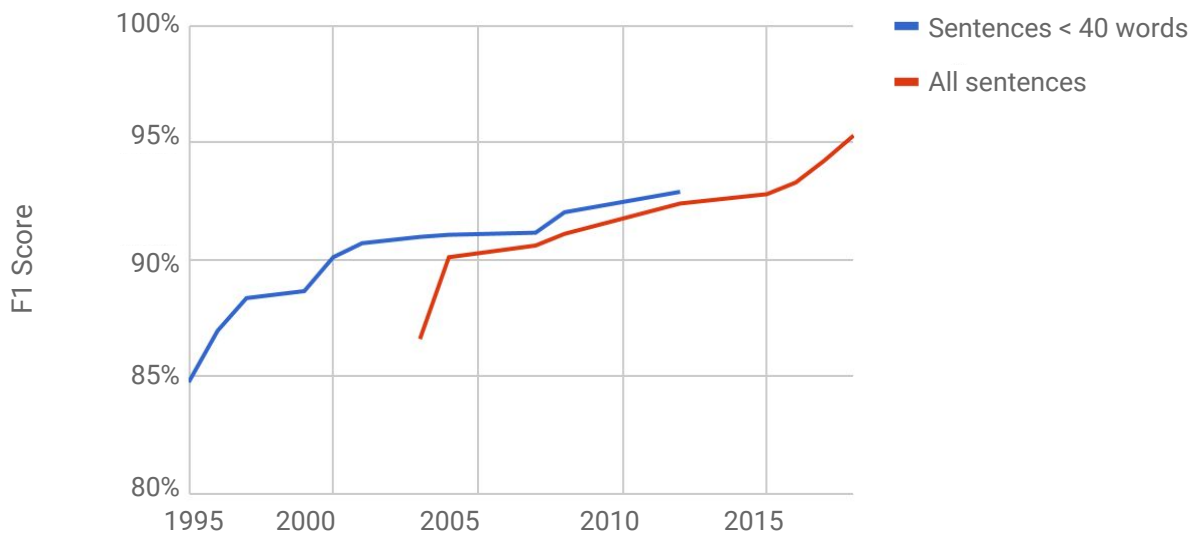
From 2003 to 2018, constituency parsing performance increased by 10%

Machine translation

The graph below shows the performance of AI systems on a task to translate the news from English to German and German to English.

BLEU scores for translations from English to German are 3.5x greater today than they were in 2008. Translations from German to English are 2.5x greater over that same time frame. Because each year uses different test sets, scores are not perfectly comparable across years (we believe this contributes to the drop in 2017 — see more in the appendix). Still, BLEU scores indicate progress in machine translation.

See Appendix 21 for data and methodology.

News translation — WMT competition (2008—2018)
Source: EuroMatrix
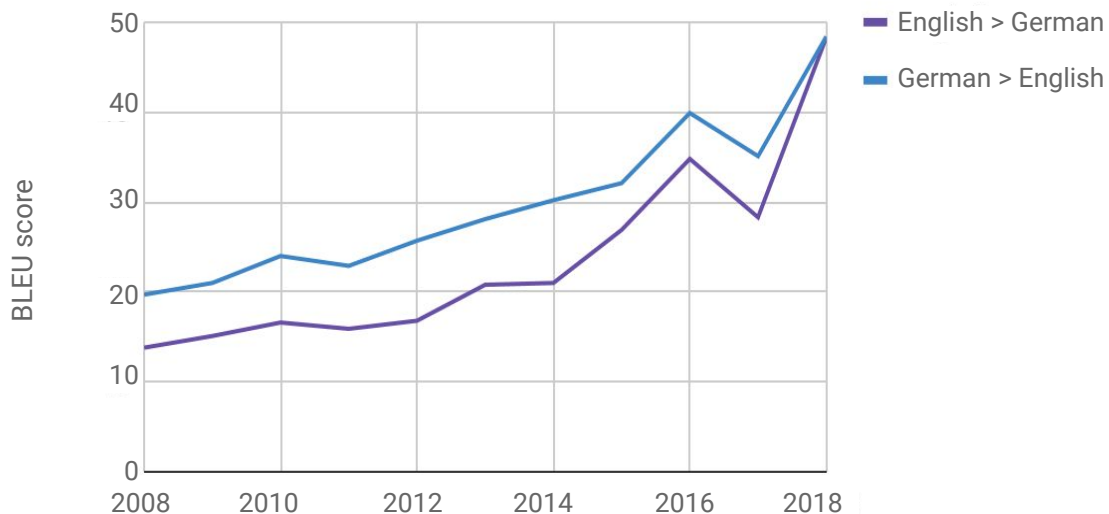
The English to German BLEU score is 3.5x higher today than in 2008

"…There are a couple of things that the top-performing systems do well. They exploit the transformer architecture for Neural MT and they exploit the data more effectively. The transformer architecture can give better results than earlier neural MT models (based on RNNs), but it requires careful choice of hyper-parameters and optimization…"
—Barry Haddow, University of Edinburgh

Commercial machine translation performance
Learn more about the state of commercial machine translation here.

51

Question answering: ARC

The graph below shows performance over time for AI2 Reasoning Challenge (ARC). The ARC data set contains 7,787 genuine grade-school level (US grades 3—9), multiple-choice science questions assembled to encourage research in advanced question answering. The questions are divided into a Challenge Set (2,590 questions) and an Easy Set (5,197 questions).

The Challenge Set contains only questions answered incorrectly by both a retrieval-based algorithm and a word co-occurrence algorithm. The questions are text-only, English language exam questions that span several grade levels as indicated in the files. Each question has a multiple-choice structure (typically 4 answer options). The questions are accompanied by the ARC Corpus, a collection of 14M unordered, science-related sentences including knowledge relevant to ARC. It is not guaranteed that answers to the questions can be found in the corpus.

The ARC benchmark was released in April 2018. Performance in 2018 has gone from 63% to 69% on the Easy Set and from 27% to 42% on the Challenge Set.

See Appendix 22 for data and methodology.

ARC leaderboard (April 2018—November 2018)
Source: Allen Institute for Artificial Intelligence

Note: This visual shows leading submissions connected by a trend line. More information on submission dates can be found in the appendix.

Question answering: GLUE

The graph below shows results from the GLUE benchmark leaderboard. General Language Understanding Evaluation (GLUE) is a new benchmark intended to test Natural Language Understanding (NLU) systems on a range of tasks and encourage the development of systems not tailored to specific tasks. It consists of nine sub-tasks — two on single sentences (measuring linguistic acceptability and sentiment), three on similarity and paraphrase, and four on natural language inference, including the Winograd Schema Challenge. The corpora sizes vary from less than 1,000 to over 400,000. The metrics include accuracy / F1 and the Matthews Correlation Coefficient.

Though the benchmark was only released in May 2018, performance has already improved to make up roughly half the gap between the first published baseline and the estimated non-expert human level of about 90%.

See Appendix 23 for data and methodology.

GLUE benchmark leaderboard (May 2018—October 2018)
Source: Gluebenchmark.com
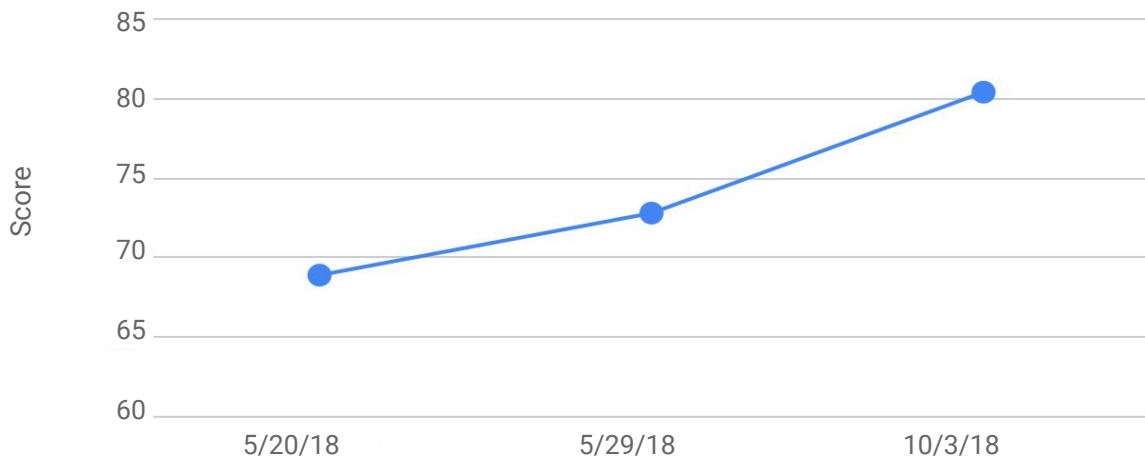
"…We haven't seen a large community emerge around GLUE yet, but I think the two major submissions we've had so far have been recognized as major milestones in representation learning for sentence understanding. Google's BERT—the current state of the art—has been cited eight times despite having been out for just over a month. It was a major focus of discussion in the corridors at the recent NLP conference EMNLP, and it's seeming like it will be the standard baseline to beat in near future work on any language understanding problem, whether or not that work is about pretrained representations…"
—Sam Bowman, NYU

Other Question Answering data sets / tasks
There has been a rapid expansion of QA data sets. Some other data sets not included in the 2018 AI Index: Squad 2.0; OpenBookQA; HotpotQA; MultiRC; NarrativeQA; DrQA; CoQA; QuAC; CWQ (rc setting)

OTHER MEASURES

DERIVATIVE MEASURES

The *Derivative Measures* section seeks to examine relationships between trends displayed in earlier pages.

The first measure, *Academia-Industry Dynamics*, plots the growth of select academia metrics alongside the growth of select industry dynamics. The second measure, the *AI Vibrancy Index,* combines academic and industry metric into one Index.

Academia-Industry Dynamics

To explore the relationship between AI-related activity in academia and industry, we first select a few representative measurements from the previous sections. In particular, we look at AI paper publishing from Scopus, combined enrollment in introductory AI and ML courses at several U.S. universities, and VC investments into AI-related startups.

These metrics represent quantities that cannot be compared directly. In order to analyze the relationship between trends, we normalize each measurement starting in 2010 and show growth rather than absolute numbers.

Academia-Industry dynamics (2010 — 2017)
Source: Sand Hill Econometrics, Scopus, university provided data
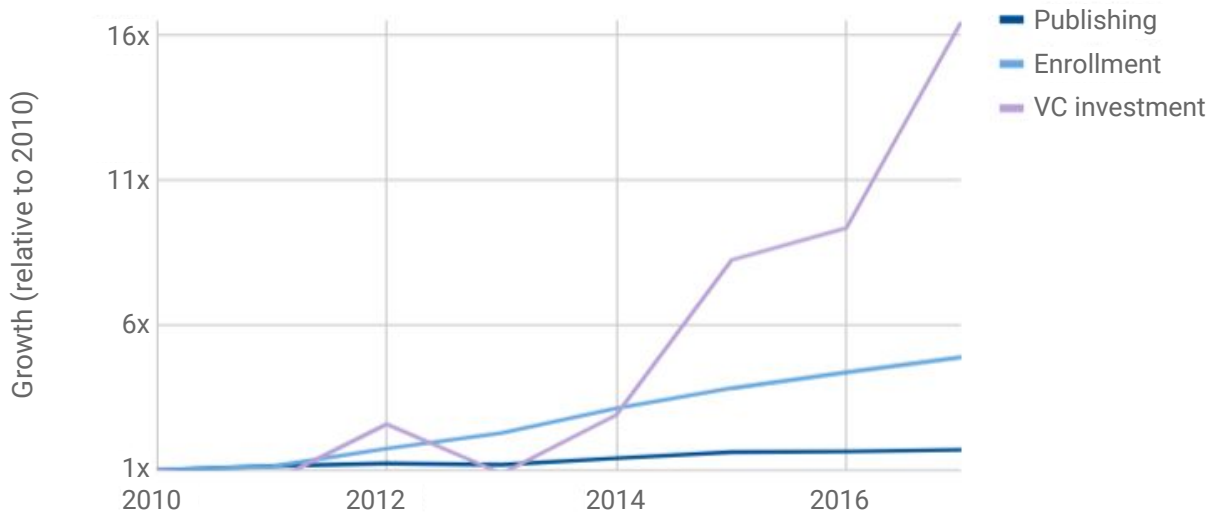
Note: The 2017 report included only Stanford University in the enrollment metric, this year we include the U.S. universities listed here.

The AI Vibrancy Index

The AI Vibrancy Index aggregates the three Academia-Industry metrics (publishing, enrollment, and VC Investment) into one measurement. We use the AI Vibrancy Index to quantify the liveliness of AI as a field. Like Academia-Industry Dynamics, the AI Vibrancy Index is normalized at 2010.

AI Vibrancy Index is largely driven by VC investment, as VC investment has grown the most relative to the other two metrics (see previous page). The three metrics are weighted equally. We encourage readers to download our data and adjust the metrics and their weights to create a Vibrancy Index that aligns most with their particular views of the metric's relevance.

Publishing: See publishing appendix
Enrollment: See enrollment appendix
VC investment: See VC appendix

AI Vibrancy Index (2010 — 2017)
Source: Sand Hill Econometrics, Scopus, university provided data

RECENT GOVERNMENT INITIATIVES

The resources listed below were compiled by the European Commission's Joint Research Center, and by Tim Dutton in his June 2018 Overview of National AI strategies.

USA	EUROPE *(EU, U.K., and France only — does not include several initiatives by other european countries)*	CHINA
October 2016: The Obama administration proposes a national AI R&D strategy with the goal of investing in research, developing methods for human-AI collaboration, addressing the safety, ethical, legal and, societal implications of AI, creating public data sets for AI training, and evaluating AI technologies through standards and benchmarks. The U.S. government also published the first policy report on AI: Preparing for the Future of Artificial Intelligence.	***EU*** **April 2018**: EU member states sign the *Declaration of Cooperation on AI*, where they agree to work together. EC also issues a *Communication on AI* with the objectives of (1) boosting the EU's research investments (2) preparing for socioeconomic changes (3) building an ethical and legal framework. The communication also dedicates 1.5B euros (1.7B USD) to support AI research during 2018—2020. This is in addition to the 2.6B euros from the Horizon 2020 program. The EC's goal is to invest 20B euros in AI research over the next 10 years.	Since 2014, the Chinese government has launched a series of key national AI initiatives with the goal of creating a $14.7B AI market in China by 2018 and ensure that China leads the world in AI by 2030.
May 2018: The Trump administration's *Summit on AI* announces its goals to (1) maintain American leadership in AI (2) support the American worker (3) promote public R&D, and (4) remove barriers to innovation. The *Select Committee on Artificial Intelligence* is created to advise the White House on AI R&D priorities and to consider the creation of partnerships with academia and Industry.	**June 2018**: the EC proposes a Digital Europe program with a budget of 9.2B euros (10.4B USD) for 2021—2027. The program focuses on advancing AI technology and ensuring the use of AI across the economy and society. The EC also proposes to develop common 'European libraries' of algorithms that are accessible to all.	**July 2015**: *Internet+* initiative focuses on intelligent manufacturing in China. AI-specific objectives include: increasing public support for development of AI; promoting the popularization of AI in areas like the smart home, smart car, and robots; building a large training database including voice, image, video, maps, research; and developing and industrializing key AI technologies such as computer vision, language processing, and human computer interaction. **2016—2020:** The robot industry development plan makes way for the development of intelligent industrial and service robots in China.

Note: See additional metrics on the following page.

The resources listed below were compiled by the European Commission's Joint Research Center, and by Tim Dutton in his June 2018 Overview of National AI strategies.

USA	EUROPE (EU, U.K., and France only — does not include several initiatives by other european countries)	CHINA
September 2018: DARPA announces $2B+ investment plan to overcome limitations on AI technology. The AI Next program begins. The Subcommittee on Information Technology of the U.S. House Committee on Oversight and Government Reform publishes a white paper on AI and its impact on policy: Rise of the Machines: Artificial Intelligence and its Growing Impact on U.S. Policy. Private companies play the central role in AI development / investment in the U.S. In 2017, private technology companies like Amazon and Alphabet invested $16.1B and $13.9B, respectively, in R&D. To put this in perspective, the total budget for the NSF, together with DARPA and DOT's investment in autonomous and unmanned systems totals $5.3 billion in the 2019 budget.	***France*** **2018:** France publishes three recent reports: (1) The French AI plan, which proposes a strategy on research, education, innovation, (2) A report produced by the Office of the Parliament for Science and Technology that focuses on social and regulatory issues, and (3) The Mission Villani report , which focuses on policy and reducing the brain drain, and also addresses the importance of diversity in AI research. ***United Kingdom*** **2016:** The U.K. Parliament House of Lords Select Committee on Artificial Intelligence began releasing a series of AI policy reports. **April 2018:** The government publishes its *AI Sector Deal* which invests 950M pounds (1.2B USD) to support research / education, and enhance the UK's data infrastructure.	**July 2017:** *New Generation AI Development Plan* strategizes AI development by 2030. It plans for China's AI technology to be on par with most advanced levels worldwide by 2020. By 2025 it states that China will have major breakthroughs in AI theory and AI will become the driving force for industrial upgrading and economic restructuring. And by 2030, China will become the world's major AI innovation center.

HUMAN-LEVEL PERFORMANCE MILESTONES

HUMAN-LEVEL PERFORMANCE MILESTONES

In the inaugural 2017 AI Index report, we included a timeline of circumstances where AI reached or beat human-level performance. The list outlined game playing achievements, accurate medical diagnoses, and other general, but sophisticated, human tasks that AI performed at a human or superhuman level. This year, we add four new achievements to that list. It is important not to over-interpret these results. The tasks below are highly specific, and the achievements, while impressive, say nothing about the ability of the systems to generalize to other tasks.

1980 Othello

In the 1980s Kai-Fu Lee and Sanjoy Mahajan developed BILL, a Bayesian learning-based system for playing the board game Othello. In 1989, the program won the U.S. national tournament of computer players, and beat the highest ranked U.S. player, Brian Rose, 56—8. In 1997, a program named Logistello won every game in a six game match against the reigning Othello world champion.

1995 Checkers

In 1952, Arthur Samuels built a series of programs that played the game of checkers and improved via self-play. However, it was not until 1995 that a checkers-playing program, Chinook, beat the world champion.

1997 Chess

Some computer scientists in the 1950s predicted that a computer would defeat the human chess champion by 1967, but it was not until 1997 that IBM's DeepBlue system beat chess champion Gary Kasparov. Today, chess programs running on smartphones can play at the grandmaster level.

2011 Jeopardy!

In 2011, the IBM Watson computer system competed on the popular quiz show Jeopardy! against former winners Brad Rutter and Ken Jennings. Watson won the first place prize of $1 million.

2015 Atari Games

In 2015, a team at Google DeepMind used a reinforcement learning system to learn how to play 49 Atari games. The system was able to achieve human-level performance in a majority of the games (e.g., Breakout), though some are still significantly out of reach (e.g., Montezuma's Revenge).

HUMAN-LEVEL PERFORMANCE MILESTONES

2016 — Object Detection in ImageNet

In 2016, the error rate of automatic labeling of ImageNet declined from 28% in 2010 to less than 3%. Human performance is about 5%.

2016 — Go

In March of 2016, the AlphaGo system developed by the Google DeepMind team beat Lee Sedol, one of the world's greatest Go players, 4—1. DeepMind then released AlphaGo Master, which defeated the top ranked player, Ke Jie, in March of 2017. In October 2017, a Nature paper detailed yet another new version, AlphaGo Zero, which beat the original AlphaGo system 100—0.

2017 — Skin Cancer Classification

In a 2017 Nature article, Esteva et al. describe an AI system trained on a data set of 129,450 clinical images of 2,032 different diseases and compare its diagnostic performance against 21 board-certified dermatologists. They find the AI system capable of classifying skin cancer at a level of competence comparable to the dermatologists.

2017 — Speech Recognition on Switchboard

In 2017, Microsoft and IBM both achieved performance within close range of "human-parity" speech recognition in the limited Switchboard domain.

2017 — Poker

In January 2017, a program from CMU called Libratus defeated four top human players in a tournament of 120,000 games of two-player, heads up, no-limit Texas Hold'em. In February 2017, a program from the University of Alberta called DeepStack played a group of 11 professional players more than 3,000 games each. DeepStack won enough poker games to prove the statistical significance of its skill over the professionals.

2017 — Ms. Pac-Man

Maluuba, a deep learning team acquired by Microsoft, created an AI system that learned how to reach the game's maximum point value of 999,900 on Atari 2600.

HUMAN-LEVEL PERFORMANCE MILESTONES

2018 — Chinese - English Translation

A <u>Microsoft</u> machine translation system achieved human-level quality and accuracy when translating news stories from Chinese to English. The test was performed on newstest2017, a data set commonly used in machine translation competitions.

2018 — Capture the Flag

A DeepMind agent reached human-level performance in a modified version of Quake III Arena <u>Capture the Flag</u> (a popular 3D multiplayer first-person video game). The agents showed human-like behaviours such as navigating, following, and defending. The trained agents exceeded the win-rate of strong human players both as teammates and opponents, beating several existing state-of-the art systems.

2018 — Dota 2

<u>OpenAI Five</u>, OpenAI's team of five neural networks, defeats amateur human teams at <u>Dota 2</u> (with <u>restrictions</u>). OpenAI Five was trained by playing 180 years worth of games against itself every day, learning via self-play. *(OpenAI Five is not yet superhuman, as it failed to beat a professional human team)*

2018 — Prostate Cancer Grading

Google developed a <u>deep learning system</u> that can achieve an overall accuracy of 70% when grading prostate cancer in prostatectomy specimens. The average accuracy of achieved by US board-certified general pathologists in study was 61%. Additionally, of 10 high-performing individual general pathologists who graded every sample in the validation set, the deep learning system was more accurate than 8.

WHAT'S MISSING?

WHAT'S MISSING?

The AI Index is still evolving, and many elements should be added before it is truly captures the state of AI. This will only happen with continued input and cooperation of the AI community. Here are three examples of suggestions by AI experts on directions in which the AI Index can be expanded.

FRANCESCA ROSSI
IBM, University of Padova *(on leave)*

Common sense reasoning and natural language understanding,
AI-powered question/answering systems are very useful, but they lack the capability to sustain a dialog. Deep natural language understanding is still a challenge, as well as common sense reasoning capabilities. Metrics that could measure both these capabilities could help advance in dialog-based AI systems, and also in many other aspects of AI. The AI index could add metrics to follow the advancements in tests and challenges related to these capabilities, such as: the Winograd Schema Challenge, The Aristo project at AI2 and the associated Kaggle competitions on science exams, and Dialog-based challenges such as the IBM Debater.

Cooperation with humans
It is also important to move from human vs machine to human+machine environments, if we want AI to augment human intelligence rather than replacing it. Metrics to show the advancements in this could be number of non-autonomous systems where humans are the final decision makers but are supported by machines.

AI Reasoning + Learning
Machine learning is very successful in many applications, since it can discover hidden correlation within huge amounts of data. This allows to make very accurate predictions. However, in many domains we need to go beyond correlation and derive causality information, in order to move from prediction to intervention and be able to shape the future. A possible metric about this could be the number of causality papers at AI conference (not just ML, also AI conferences like AAAI or IJCAI)

For this, it is enough to mine the titles of the papers. It is also important to monitor the work that combines ML with symbolic AI. A possible metric is the number of papers on ML+ symbolic AI combination.

RODNEY BROOKS
MIT

Robots with AI components
I want to address the robot shipments metric. The relevance of the robot shipments metric depends on the source. Many sources quote industrial robot shipments that have very little (or no) AI in them, which makes it a poor metric for progress in AI. Rethink Robotics is the one exception that I know of.

One could also look at robots which have an AI component, such as drones (which use SLAM, and other AI algorithms.) Home robots, such as Roomba, also have an AI component. Roomba would be interesting to track as it is numerically the most shipped robot ever. While there have been some recent failures in social home robots, any success in that space will rest heavily on AI

TOBY WALSH
UNSW Sydney and TU Berlin

Government spending and military use
Perhaps one of the most significant developments in AI in the last two years has been the increased investment announced by both governments and industry. We have seen, for instance, the UK government commit to spend GBP 1 billion, France to euro 1.5 billion, and Germany to 3 euro billion. At the same time, we've seen companies like Alibaba announce plans to invest $15 billion, and SoftBank's Vision Fund is focusing much of its $100 billion in AI.

The AI Index would benefit then from quantitative metrics that measure this increasing investment. This is often described as an "AI race" with countries like China declaring an ambition to seek economic and military dominance through the use of AI. To measure how the focus of AI research and development is shifting, such metrics might usefully break down investments by country. One other area of increasing concern is the use of AI by the military. There are some areas where this is to be welcomed like mine clearing. However, there are other areas, like the development of fully autonomous weapons, where many within the field have expressed grave concerns. The AI index would therefore also benefit from metrics which measure the development and uptake of such technologies within militaries around the world.

ACKNOWLEDGEMENTS

ACKNOWLEDGEMENTS

ACKNOWLEDGEMENTS

APPENDIX

APPENDIX 1
Papers on Scopus

See underlying data

Source
Elsevier's Scopus database of academic publications, which has indexed nearly 70 million documents. See more information about Scopus.

Methodology
It is important to note the difference between this methodology and the methodology described in the next page. Here we use the Scopus query tool to get the count of papers tagged with the keywords *Artificial Intelligence* and *Computer science*. This was done by the AI Index team.

Scopus tags its papers with keywords, publication dates, country affiliations, and several other metrics. For AI publications by year and region, we used the 'opensearch:totalResults' tool (which returns the paper count) to run the following query for each year / country combination:

AI by country query
title-abs-key(artificial intelligence) AND PUBYEAR = {} AND AFFILCOUNTRY({})

For annually published papers by topic, the following queries were used:

AI query
title-abs-key(artificial intelligence) AND PUBYEAR AFT {} AND PUBYEAR BEF {}
CS query
SUBJAREA(COMP) AND PUBYEAR AFT {} AND PUBYEAR BEF {}
All Scopus query
PUBYEAR AFT {} AND PUBYEAR BEF {}

All queries above were made for each year from 1996 to 2017.

Nuance
- The Scopus system is retroactively updated. As a result, the number of papers for a given query may increase over time.
- Members of the Elsevier team commented that data on papers published after 1995 would be most reliable (hence 1996 as a starting year for Scopus data).
- The Scopus paper count includes papers of all qualities and relevances to AI.

Nuances specific to AI publications by region
- Papers are double counted if they are tagged to multiple regions. This explains why top line numbers in a given year may not match last year's annual paper count.
- "Other" includes all other countries that have published AI paper(s) on Scopus.

Nuances specific to publications by topic
- The 2017 Index only showed AI papers within the CS category. This year, all papers tagged as AI were included, regardless of whether they fell into the larger CS category.
- Because AI papers on Scopus generally fall into the larger CS category, growth in CS papers largely includes the growth of AI papers.
- The "All papers" category includes all AI and all CS papers.

Elsevier methodology for papers on scopus

See underlying data

Source
Elsevier's Scopus database of academic publications, which has indexed nearly 70 million documents. This data was compiled by Elsevier. See more information about Scopus. See more information about Elsevier.

In depth methodology on paper indexing, affiliations, geographic coverage, and titles can be found on the Scopus Content Coverage Guide.

Methodology
It is important to note the difference between this methodology and the methodology described in the previous page. Here we outline Elsevier's methodology of counting AI papers, which uses a bottom-up approach with about 800 keywords. There are more details are available in Elsevier's recent AI report.

Defining AI
Elsevier established a set of about 800 keywords relevant to AI. These keywords were used to identify 600,000 AI papers from Scopus that became the basis of Elsevier's AI report. The keywords were determined as follows:

First, Elsevier used its Fingerprint Engine to absorb AI books, course syllabi, patents, and news articles. The Fingerprint Engine digested the documents and output a set of the most relevant weighted keywords. This produced 20,000 relevant terms. After a manual review of these terms, 797 unique keywords were selected. Elsevier expects to release them in the coming months. A sample list is available.

Still, several papers that contained one or several of the 797 AI keywords were not actually AI papers. This is due to false positives — for example, the term *ethical values* is overly broad and could be included in several non-AI papers. Similarly, the term *neural networks* belongs to non-AI fields, such as biology.

To address false positives, Elsevier added expert guidance though a manually selected training set of 1,500 AI publications on what is core or applied AI. They used this in a supervised classifier to eliminate false positive publications and to work with the size of the publication set of over 600,000 publications from Scopus.

Overall, Elsevier's approach was designed to capture the field of AI broadly, with multiple perspectives and keywords, while also incorporating the precision of the supervised classifier.

Country affiliations
An author's country affiliation is determined based on his or her primary organization, which is provided by authors of the papers. Global organizations will use the headquarters' country affiliation as a default, unless the author is specific in his/her organization description. For example, an author who inputs "Google" as their organization will be affiliated with the United States, one that inputs "Google Zurich" will be affiliated with Europe.

Papers are double counted when authors from multiple geographies collaborate. For example, a paper with authors at Harvard and Oxford will be counted once for the U.S. and once for Europe.

Europe is defined as EU44.

These rules also apply to sector affiliations.

Elsevier methodology for papers on scopus (continued)

AI subcategories (see visual)
Below you'll find a summary of Elsevier's method for determining AI subcategories on Scopus. To learn more about Elsevier's methodology, view Elsevier's recent AI report here.

To determine broad subcategories, Elsevier used an unsupervised clustering technique called the Louvain method. This approach maps the keywords into clusters and illustrates their connections, based on co-occurrence within the documents. Co-occurrence indicates that those clusters do not stand alone, but strongly relate to each other, e.g., neural networks in a computer vision document. They found that the AI field seems to cluster around the following areas:

Probabilistic Reasoning, Neural networks, Computer Vision, Search and Optimization, NLP and Knowledge Representation, Fuzzy logic, Planning and Decision making

Elsevier AI Resource Center has publish an interactive graph of the clusters, allowing users to browse individual connections and clusters, by region and over time.

Relative activity focus (see visual)
Papers are assigned subject areas by professional indexers, who subjectively categorize each paper into subject areas based on its scope and the content. Subject areas are not mutually exclusive.

Nuance
- The entire data collection process was done by Elsevier internally — the AI Index was not involved in the keyword selection process or the counting of relevant papers.
- These figures are approximately twice those included in the AI Index 2017 report which used Scopus but a different methodology, and about half as many as those reported in the China AI Development Report, which is based on data from Web of Science.
- The boundaries of AI are difficult to establish, in part because of the rapidly increasing applications in many fields, such as speech recognition, computer vision, robotics, cybersecurity, bioinformatics, and healthcare. But limits are also difficult to define because of AI's methodological dependency on many areas such as logic, probability and statistics, optimization, photogrammetry, neuroscience, and game theory — to name a few. Given the community's interest in AI bibliometrics, we believe it would be valuable if groups producing these studies would strive for a level of transparency in their methods which supported reproducibility of results, in particular on different underlying bibliographic databases.

APPENDIX 3
Papers on arXiv

See underlying data

Source

arXiv.org is an online archive of research articles in the fields of physics, mathematics, computer science, quantitative biology, quantitative finance, statistics, electrical engineering and systems science, and economics. arXiv is owned and operated by Cornell University. See more information on arXiv.org.

Methodology

Raw data for our analysis was provided by representatives at arXiv.org. The keywords we selected, and their respective categories, are below:

Artificial intelligence (cs.AI)

Computation and language (cs.CL)

Computer vision and pattern recognition (cs.CV)

Machine learning (cs.LG)

Neural and evolutionary computing (cs.NE)

Robotics (cs.RO)

Machine learning in stats (stats.ML)

For most categories, arXiv provided data years 1999 — 2017. For our analysis, we decided to start at the year 2010 in order to include *Machine Learning in Stats*, which did not exist on arXiv prior.

To see other categories' submission rates on arXiv, see arXiv.orgs submission statistics.

Nuance

- Categories are self-identified by authors — those shown are selected as the "primary" category. Therefore, it is worth noting that there is not one streamlined categorization process. Additionally, the *Artificial intelligence* or *Machine learning* categories may be categorized by other subfields / keywords.
- arXiv team members have shared that participation on arXiv can breed more participation — meaning that an increase in a subcategory on arXiv could drive over-indexed participation by certain communities.
- Growth of papers on arXiv does not reflect actual growth of papers on that topic. Some growth can be attributed to arXiv.org's efforts to increase their paper count, or to the increasing importance of dissemination by AI communities.

AAAI papers by country

See underlying data

Source

The Association for the Advancement of Artificial Intelligence (AAAI) hosts conferences every year, including the annual "AAAI conference". Raw data on 2018 AAAI paper submissions / acceptances by country was provided by AAAI representatives. Learn more about the AAAI conferences.

Methodology

We collected data on AAAI submissions / acceptance by country from the AAAI team. AAAI was the only conference where we were able to obtain this level of detail. The AI Index hopes to include equivalent data for other conferences in future reports.

Nuance
- Countries included in this analysis are those that submitted 10 or more papers to the AAAI conference.
- This data is from the 2018 conference, hosted in February 2018 in New Orleans, Louisiana. The landscape of submitted / accepted papers may look different for other years.
- Acceptance is largely limited due to space constraints, see below for a quote from AAAI on space limitations for the AAAI 2019 conference (1/27/19—2/1/19 in Honolulu, HI):

  "We had a record number of over 7,700 submissions this year [2019]. Of those, 7,095 were reviewed, and due to space limitations we were only able to accept 1,150 papers, yielding an acceptance rate of 16.2%." — AAAI

APPENDIX 5
U.S. AI and ML course enrollment

See underlying data

Source
Course enrollment data was collected directly from each university. Total student population was collected from school archives (typically housed on Office of the Registrar sites). The following universities are included in our analysis:

University of California-Berkeley, Stanford University, University of Illinois at Urbana–Champaign, University of Washington-Seattle, Carnegie Mellon University

Methodology
We requested enrollment in introductory AI and introductory ML courses over time at many leading computer science universities in the U.S. Several schools participated. Enrollment counts were not included in our analysis if the school did not include sufficient historical data, or if the data was overly nuanced.

Some schools provided enrollment by semester, and some provided it by year. In order to compare schools to each other, we collapsed semester enrollment data to show full academic years. Additionally, some schools had multiple courses that were considered "introductory" while others just had one. When appropriate and relevant, multiple courses were combined to show one "introductory AI" trend line.

For enrollment as a percent of the undergraduate population, each year's AI / ML enrollment was divided by the undergraduate population for that same year. This is a calculated field intended to show trends in enrollment on an even playing field across schools.

Nuance
- Nearly every school noted that enrollment, particularly in recent years, is a function of supply, rather than student demand. Our data shows the number of students that were successfully enrolled in a course, and does not account for waitlists or other demand metrics.
- Courses are generally open to undergraduates only, and can typically be taken by majors and non-majors. Some courses have changed their names over time, we show course names as of 2017 below. We also list any additional details / nuances that school administrators were able to provide on the enrollment data.
- Any nuance that was not mentioned by school administrators is not captured below.

AI courses:
Berkeley CS 188
Stanford CS 221
UIUC CS 440
- UIUC representatives attribute growth to larger classrooms / additional sections to meet some of the excess student demand.
UW CSE 415 / 416 (non-majors) & CSE 473(Majors)
- CSE 416 is new as of AY 2017 and accounts for some growth in AI course enrollment in 2017

ML courses:
Berkeley CS 189
- Representatives at Berkeley speculate that growth is due to a combination of novelty, subject interest, and growth in majors that allow Intro ML as a way to fill requirements.
Stanford CS 229
UIUC CS 446
UW CSE 446
CMU 10-701

International course enrollment

See underlying data

Source
Course enrollment data was collected directly from each university. The following universities are included in our analysis:

Tsinghua University (China), National Institute of Astrophysics, Optics and Electronics (Mexico), University of British Columbia (Canada), University of Toronto (Canada), University of Edinburgh (Scotland), University of Science and Technology of China (China), Shanghai Jiao Tong University (China), Pontificia Universidad Católica de Chile (Chile), TU Wien (Austria), Hebrew University (Israel), EPFL (Switzerland), MILA (Quebec, Canada)

Methodology — See methodology from Appendix 5.

Nuance
- Nearly every school noted that enrollment, particularly in recent years, is a function of supply, rather than student demand. Our data shows the number of students that were successfully enrolled in a course, and does not account for waitlists or other demand metrics.
- Unlike the U.S. schools studied, international schools significantly varied on whether courses were only open to undergraduates or not.
- Visual one shows growth in AI and ML courses combined. Visual two shows just AI course enrollment. We did this in order to show like for like data on each graph. In some cases, we had access to additional data on a school but did not show it because we wanted to have parallel information across schools. Additional data in located in the underlying data link in the top right corner.
- Some courses have changed their names over time, we show course names as of 2017 below. We also list any additional details / nuances that school administrators were able to provide on the enrollment data. Any nuance that was not mentioned by school administrators is not captured below.

INAOE — Courses: C141 (AI) and C142 (computational learning)
Notes: INAOE AI / ML enrollment is greatly affected by the number of students accepted into the INAOE graduate program as a whole. INAOE representatives say that there is a decreasing number of INAOE students, thus affecting AI / ML course enrollment.
USTC — Courses: USTC listed several introductory AI / ML courses across various departments including the Department of Computer Science and Technology, The Department of Automation, the Department of Information Science and Technology and the Department of Data Science.
University of Edinburgh — Courses: Intro applied ML (undergraduate and graduate students) and Informatics 2D — Reasoning and Agents (undergraduate only)
SJTU — Course: CS410 (undergraduate intro to AI)
PUC — Course: Intro to AI
Prior to 2017, the course was only taught once a semester. The large demand in 2017, relative to 2018, is due to the transition from one course to two courses.
Tsinghua — Courses**:** AI (60240052 & 40250182) and ML (00240332 & 70240403 & 80245013 & 80250943)
Open to undergraduates and graduate students
Toronto — Courses: AI (CSC384) and ML:(CSC411)
2016 was the first year that a summer AI course was open. decision to open two semesters of ML in 2015 — due to increased demand
UBC — Courses: AI and ML (CPSC 322, CPSC 340, CPSC 422)
TU Wien — Courses: Intro to AI (undergraduate) and ML (graduate)
Neither course is mandatory for CS students
Hebrew University — Courses: Intro to AI (67842) and Intro to ML (67577)
EPFL — Courses: Artificial Intelligence (undergraduate) and Intelligent Agents (graduate)
EPFL representatives let us know that there is a lot of variation each year due to scheduling and resources.
MILA — Machine Learning (graduate only, non-advanced)
MILA is a program for graduate students only. MILA representatives note that this ML course is not taken by core ML research students, but by students in other specialties who want to learn ML. Core ML students take more advanced courses. There was no course in 2014.

APPENDIX 7
Faculty diversity

See underlying data

Source
Faculty diversity was collected manually via AI department websites on September 21st, 2018. Schools selected are leading computer science universities with easily accessible AI faculty rosters.

Methodology
In order to get the gender diversity breakdown of AI faculty, professor names were collected on school websites, and then genders were assigned (see first nuance below) using both names and pictures. Please see below for more specific details on each school:

UC Berkeley — See faculty link
Includes *Assistant Professors*, *Primary*, *Secondary Faculty*
Stanford University — See faculty link
Includes *Faculty and Research Scientists* and *Affiliated Faculty*
University of Illinois at Urbana–Champaign — See faculty link
Includes *CS Faculty* and *Affiliate Faculty*
Carnegie Mellon University — See faculty link
Includes all faculty listed
University College, London — See faculty link
Includes all faculty under the *People* link
University of Oxford — See faculty link
Includes *Faculty* section only
ETH Zurich — See faculty link
Includes only those with *"Dr."* in their title

Nuance
- We assigned genders using professor names and pictures. In doing so, the AI index may have misgendered someone. We regret that we could not include non-binary gender categories into this analysis. We hope the metric is still useful in showing a broad view of gender representation in AI academia today, and look forward into expanding into other types of gender diversity in future reports.
- School data was pulled September 21st, 2018. School faculty could be altered by the time the 2018 AI Index report is published.
- Data is pulled from schools' AI faculty rosters and does not account for visiting professors or professors housed in other departments. Similarly, it will count a professor that is listed as an active member of AI faculty, even if that professor belongs to a different department.
- Gender representation in academia does not imply representation in industry (in fact, the proportion of women working in AI in industry may be lower).
- The metric provides a snapshot of representation today, and does not account for improvements over time, see below for a statement from a Stanford AI faculty member, Dr. James A. Landay:

  "We are very focused on hiring more diverse faculty. Most of the women on that list have been hired in just the last 2—3 years, so we have been making progress"
  - Dr. Landay, Stanford University

APPENDIX 8
Participation

Source

Conference attendance data was collected directly from conference / organization representatives. Data was collected from the following conferences:

AAAI — Association for the Advancement of Artificial Intelligence
AAMAS — International Conference on Autonomous Agents and Multiagent Systems
AI4ALL
ACL — Association for Computational Linguistics
CVPR — Conference on Computer Vision and Pattern Recognition
ICAPS — International Conference on Automated Planning and Scheduling
ICLR — International Conference on Learning Representations
ICML — International Conference on Machine Learning
ICRA — International Conference on Robotics and Automation
IJCAI — International Joint Conferences on Artificial Intelligence
KR — International Conference on Principles of Knowledge Representation and Reasoning
NeurIPS — Conference on Neural Information Processing Systems
UAI — Conference on Uncertainty in Artificial Intelligence
WiML — Women in Machine Learning workshop

Methodology

We defined *large conferences* as those with 2,000 or more attendees in 2018, and *small conferences* as those with fewer than 2,000 attendees in 2018. Conferences selected are those that lead in AI research and were also able to supply yearly attendance data.

AI4ALL and WiML were selected for their progress on AI inclusion and their availability of data. We look forward to adding more organizations / conferences that cater to underrepresented groups in future reports.

Nuance

Nuances specific to conferences

- Some conference organizers were only able to provide estimates of attendance — we have accepted estimates as accurate.
- Some conferences do not run annually, and some have skipped years.
- Several conference organizers have let us know that because conference venues are determined over a year in advance, the supply of spots are often limited. Therefore, the number of conference attendees doesn't necessarily reflect demand.

Nuances specific to AI4ALL / WiML

- It is important to note that several other formal and informal programs exist to support inclusivity in AI.
- Participation does not necessarily indicate progress in increasing the number of women and underrepresented groups in the field.

APPENDIX 9
Robot Software Downloads

See underlying data

Source

Data on ROS.org pageviews and robot software downloads comes from the ROS.org community metrics reports. See PDFs by year. Learn more about ROS.org.

Nuance

- ROS.org page views do not represent total pageviews, only pageviews in top countries. 90+ lower volume regions are not included.
- 2018 metrics are available on ROS.org, but are not included in the AI Index as they use a slightly different different methodology. See more about this on this thread.
- The monthly average was calculated using July of every year (except 2012 and 2013 which used June and August, respectively).
- ROS.org representatives note that robot software activity in non-english speaking countries could be underestimated, as metrics only account for websites in english.
- ROS.org representatives speculate that the dip in 2015 is likely due to sampling issues as well as fewer patch releases that month so it reflected in the download stats. See metrics that account for this here.

APPENDIX 10
Startups / Investments

Underlying data is not available.

Source
Sand Hill Econometrics, using the Dow Jones VentureSource database, provided the startups and VC investment data. The Crunchbase API was used to corroborate the list of AI companies in the SandHill database. See links to learn more about Sand Hill, VentureSource, and Crunchbase.

Methodology
We used the following keywords to obtain a list of AI-related companies on Crunchbase:

Category Name	Crunchbase Category UUID
Artificial Intelligence	c4d8caf35fe7359bf9f22d708378e4ee
Machine Learning	5ea0cdb7c9a647fc50f8c9b0fac04863
Predictive Analytics	ca8390d722c65bb5f87022f52f364b1b
Intelligent Systems	186d333a99df4a4a6a0f69bd2c0d0bba
Natural Language Processing	789bbbefc46e1532a68df17da87090ea
Computer Vision	a69e7c2b5a12d999e85ea0da5f05b3d3
Facial Recognition	0b8c790f03bcb2aba02e4be855952d6d
Image Recognition	af9307c9641372aeaac74391df240dd2
Semantic Search	bb1777e525f3b9f2922b0d5defe0d5bb
Semantic Web	ac34c4c6e430f66f44aeef8ca45b52bb
Speech Recognition	24dad27e6a49ccc1ec31854b73d60d16
Text Analytics	e7acbc9932c74c8809ed8fae63eb2559
Virtual Assistant	2eb0b56b6896e0f0545fc747a9ba6751
Visual Search	ad8f33a8c5d0b7d786f389aa2954c119

Crunchbase returned 4,477 AI companies. This list of organizations was then cross referenced with the list of all venture-backed companies in the VentureSource database — 1,096 of the 4,477 companies from Crunchbase matched VentureSource's list of AI companies. Sand Hill aggregated the matched list to show the number of companies and amount of funding over time.

The data on "All Startups" was collected using only the VentureSource database — the Crunchbase API was not involved.

Nuance
- The VentureSource data is proprietary and we regret that it cannot be made available to the public.
- All interaction with the VentureSource product was conducted by Sand Hill Econometrics.
- Companies are removed from the VentureSource database once they go public, get acquired, or get shut down, however, the funding rounds prior to the company's removal remain in the database.
- Both Crunchbase and VentureSource have difficulty discerning AI companies that develop AI systems vs those that only deploy them. Internally, we called these AI "Makers" vs "Users". We tested a random sample of 109 of the matched companies (this is 10% of the entire set of VentureSource matches) and found that 68 of them (or 62%) were AI "Makers". The remaining 38% of companies in our sample, while still considered AI companies, likely did not develop the AI systems themselves.
- Our list of AI categories in Crunchbase currently favor machine learning technologies.
- The number of active AI startups is recorded for January of every year.

APPENDIX 11
Jobs
See underlying data

Source
The visual on job openings by AI skill *(pg 33)* and the visual on gender diversity of AI applicants *(pg 34)* were each provided by the job site Monster.com. Monster.com used Garner TalentNeuron, an online talent market intelligence portal, in order to obtain the results. Find out more about Garner TalentNeuron or Monster.com.

Methodology
Monster.com
Monster.com used the Garner TalentNeuron database to search the terms below. A job was included in the count if the term was listed as a required skill in the job posting. When searching, each term was also paired with the keyword *Artificial Intelligence*. For example, *Robotics* was only counted if the job description included the both terms *Artificial Intelligence* **and** *Robotics*. Jobs from Monster.com are U.S. only. See keywords below:

Artificial intelligence +
Natural Language Processing, Machine Learning Techniques, Deep Learning, Computer Vision, Speech Recognition, Robotics

Nuance
- Jobs are double counted if they include more than one relevant AI skill; Some skills could imply a requirement for skills that are not actually listed in the description.

Nuances specific to gender diversity of AI applicants
- Applicants do not necessarily imply hires, or overall representation in industry.
- We regret that we could not include non-binary gender categories into this analysis. We hope the metric is still useful in showing a broad view of gender representation in AI academia today, and look forward into expanding into other types of gender diversity in future reports.

APPENDIX 12
Patents

See underlying data

Source
Patent data was provided by amplified, a patent searching service.

Methodology
All methodology and underlying data is detailed by amplified here.

Nuance
- Our visuals do not show years 2015—2017 because those years do not contain complete patent counts due to the long lead time between filing and publishing.

While there have been several attempts to track AI patents over time and across regions, there is no standard methodology. As a result, our patent data may look different than those of other organizations.

A notable example of this is Tsinghua University's *China AI Development Report*, released summer 2018, which shows China having the greatest number of total published patents, followed by the U.S. and Japan. Still, it is unclear which patent codes and definitions were used to get to this result, and the actual number of patents by each region is not easily visible in the diagram.

Because the field is so difficult to define, the AI Index is fully transparent with the patent codes and keywords being used in this analysis. And we encourage readers to challenge our patent methodology so that it can eventually become fully reflective of the field.

AI adoption survey

See underlying data

Source
This survey was written, filled, and analyzed by McKinsey & Company (McKinsey). You can find additional results from McKinsey's AI adoption survey here.

Methodology
McKinsey gave this statement about their methodology:
The survey was conducted online and was in the field from February 6, 2018, to February 16, 2018. It garnered responses from 2,135 participants (out of 20,466 people who were invited to participate) who represent the full ranges of regions, industries, company sizes, functional specialties, and tenures within McKinsey's Online Executive Panel. Of them, 1,657 say their companies are piloting at least one artificial intelligence capability, out of nine that the survey asked about, in one or more functions or business units. All survey participants are members of the online panel, a group of more than 30,000 registered users of McKinsey.com who opted in to participate in proprietary McKinsey research and represent a worldwide sample of executives, managers, and employees at all levels of tenure. 108 countries are represented in the survey sample; to adjust for differences in response rates, the data are weighted by the contribution of each respondent's nation to global GDP.

McKinsey provided the AI Index with the percent of respondents selecting each answer choice to the following questions:

Q1. How would you describe your organization's current use of the following AI capabilities (Robotic process automation, Machine learning, Virtual agents or conversational interfaces, Computer vision, Natural language text understanding, Natural language speech understanding, Natural language generation, Physical robotics, Autonomous vehicles)? *Select one:*
1. Not used at all
2. Piloted in at least one function or business unit
3. Embedded in standard business processes in at least one function or business unit
4. Embedded in standard business processes across multiple functions or business units
5. Don't know

Q2. In which of the following business functions has your organization deployed AI? [randomize] (Select all that apply)
1. Marketing and sales
2. Strategy and corporate finance (e.g., capital-allocation decisions)
3. Risk (i.e., risk management, fraud, and debt)
4. Human resources
5. Product and/or service development
6. Supply chain management
7. Manufacturing
8. Service operations (e.g., field services, customer care, back office)
9. None of the above
10. Don't know

For Q1, we asked McKinsey to cut the responses by geographic location. The AI Index then summed answers 3 and 4 to get the percent of respondents that have embedded the AI capability in *at least one* function or business unit. For Q2, McKinsey provided cuts by geography, and cuts by industry. Unlike Q1, Q2 was limited to respondents whose company has embedded or piloted any AI capability. The number of relevant respondents is in the note under each visual.

Nuance
- Survey respondents are limited by their perception of their organizations' AI adoption.

Corporate initiatives: Company earnings calls

See underlying data

Source

All data on earnings calls mentions was provided by Prattle, an investment research company specializing in text analytics and sentiment analysis. Prattle sources earnings call transcripts from FactSet.

Methodology

Prattle provided us with a count of earnings calls mentions over time for the words *Artificial Intelligence* and *Machine Learning,* and the same data on the terms *Cloud* and *Big Data* for additional context and benchmarking. Results were provided by company, and included each company's industry and size.

To pull language from the call transcripts, Prattle used variants of regular expressions (also known as *regex*), to search each text for strings and language patterns used in discussions on AI. The terms *Artificial Intelligence* and *Machine Learning* also included mentions of their abbreviations *AI* and *ML*. Prattle then removed false positives (for example, because *ML* is also the abbreviation for *milliliter*, Prattle removed counts of *ML* that had numeric characters preceding the term).

Sector classifications were done using GICS, the standard groupings used by S&P. The size categorizations (these not used in the report but are available in the underlying data) are based on market cap classifications, which are calculated by rattle using data on outstanding shares and sales prices. The thresholds provided by Prattle are as follows:

mkt_cap <= '500000000' THEN 'micro'::text
mkt_cap <= '4000000000' THEN 'small'::text
mkt_cap <= '20000000000' THEN 'mid'::text
mkt_cap IS NOT NULL THEN 'large'::text
else Null

All analysis of the data was done by the AI Index.

Nuance
- We observe that the mentions of both Big data and Cloud have steep initial curves, then trail off. This can be interpreted as the normalization of the technology.
- Mentions are not necessarily a proxy for company investment in or adoption of AI.
- Only public companies were included in this analysis.

APPENDIX 15
Robot installations

See underlying data

Source
Data was pulled directly from the International Federation of Robotics' (IFR) 2014, 2015, and 2017 World Robotics Reports. See links to the reports below. Learn more about IFR.

2012 data
Source: 2014 World Robotics Report — Executive Summary (Table 1)
2013—14 data
Source: 2015 World Robotics Report — Executive Summary (Table 1)
2015—16 data
Source: 2017 World Robotics Report — Executive Summary (Table 4.1)

Methodology
The data displayed is the number of industrial robots installed by country. Industrial robots are defined by the ISO 8373:2012 standard.

See more information on IFR's methodology.

Nuance
- It is unclear how to identify what percentage of robot units run software that would be classified as "AI" and it is unclear to what extent AI development contributes to industrial robot usage.
- This metric was called *robot imports* in the 2017 AI Index report

GitHub Stars

See underlying data

Source

We used the GitHub archive stored on Google BigQuery.

Methodology

The visual in the report shows the number of stars for various GitHub repositories over time. The repositories include:

apache/incubator-mxnet, BVLC/cafe, cafe2/cafe2, dmlc/mxnet, fchollet/keras, Microsoft/CNTK, pytorch/pytorch, scikit-learn/scikit-learn, tensorflow/tensorflow, Theano/Theano, Torch/Torch7

GitHub archive data is stored on Google BigQuery. We interfaced with Google BigQuery to count the number of "WatchEvents" for each repository of interest. A sample of code for collecting the data over the course of 2016 is below:

```
 SELECT
        project,
        YEAR(star_date) as yearly,
        MONTH(star_date) as monthly,
        SUM(daily_stars) as monthly_stars
FROM (
SELECT
        repo.name as project,
        DATE(created_at) as star_date,
        COUNT(*) as daily_stars
FROM
        TABLE_DATE_RANGE(
                [githubarchive:day.],
                TIMESTAMP("20160101"),
                TIMESTAMP("20161231"))
WHERE
        repo.name IN (
                "tensorflow/tensorflow",
                 "fchollet/keras",
                "apache/incubator-mxnet",
                 "scikit-learn/scikit-learn",
                "cafe2/cafe2", "pytorch/pytorch",
                 "Microsoft/CNTK", "Theano/Theano",
                "dmlc/mxnet", "BVLC/cafe")
        AND type = 'WatchEvent'
GROUP BY project, star_date
)
GROUP BY project, yearly, monthly
ORDER BY project, yearly, monthly
```

See nuance on the following page

86

GitHub Stars (continued)

Nuance

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

There are other ways to retrieve GitHub Star data. The star-history tool was used to spot-check our results.

While Forks of GitHub project are also interesting to investigate, we found that the trends of repository Stars and Forks were almost identical.

APPENDIX 17
Public interest

See underlying data

Source

All data was retrieved from TrendKite. The TrendKite service indexes general media articles and they employ a sentiment analysis classifier that categorizes articles as "positive", "negative", or "neutral". Learn more about TrendKite.

Methodology

We used the query below to identify AI articles. We adjusted it to remove a source that introduced a disproportionate amount of irrelevant articles with negative sentiment. Last year, the proportion of neutral articles was implied but not included in the visual. This year, we have included neutral.

Query
"Artificial Intelligence"
AND NOT "MarketIntelligenceCenter.com's"

NOT site_urls_ll:(
        "individual.com"
 OR "MarketIntelligenceCenter.com")

Filters
• Only included English-language articles
• Removed press releases
• Removed financial news
• Removed obituaries

APPENDIX 18
Government mentions

See underlying data

Sources, Methodologies, and nuances, by country:
Data collection and analysis was performed by the McKinsey Global Institute (MGI).

Canada (House of Commons):
Data was collected using the Hansard search feature on Parliament of Canada website. MGI searched for the terms *"Artificial Intelligence"* and *"Machine Learning"* (quotes included) and downloaded the results as a CSV. The date range was set to "All debates." Data is as of 11/20/2018. Data are available online from 08/31/2002.

Each count indicates that *Artificial Intelligence* or *Machine Learning* was mentioned in a particular comment or remark during the proceedings of the House of Commons. This means that within an event or conversation, if a member mentions *AI* or *ML* multiple times within their remarks, it will appear only once. However if, during the same event, the speaker mentions *AI* or *ML* in separate comments (with other speakers in between) it will appear multiple times. Counts for *Artificial Intelligence* and *Machine Learning* are separate, as they were conducted in separate searches. Mentions of the abbreviations *"AI"* or *"ML"* are not included.

United Kingdom (House of Commons, House of Lords, Westminster Hall, and Committees)
Data was collected using the Find References feature of the Hansard website of the UK Parliament. MGI searched for the terms *"Artificial Intelligence"* and *"Machine Learning"* (quotes included) and catalogued the results. Data is as of 11/20/2018. Data are available online from 1/1/1800 onwards. Contains Parliamentary information licensed under the Open Parliament Licence v3.0.

Like in Canada, each count indicates that *Artificial Intelligence* or *Machine Learning* was mentioned in a particular comment or remark during a proceeding. Therefore, if a member mentions *AI* or *ML* multiple times within their remarks, it will appear only once. However if, during the same event, the same speaker mentions AI or ML in separate comments (with other speakers in between) it will appear multiple times. Counts for *Artificial Intelligence* and *Machine Learning* are separate, as they were conducted in separate searches. Mentions of the abbreviations *"AI"* or *"ML"* are not included.

United States (Senate and House of Representatives)
Data was collected using the advanced search feature of the US Congressional Record website. MGI searched the terms "*Artificial Intelligence*" and *"Machine Learning"* (quotes included) and downloaded the results as a CSV. The "word variant" option was not selected, and proceedings included Senate, the House of Representatives, and Extensions of Remarks, but did not include the Daily Digest. Data is as of 11/20/2018, and data is available online from the 104th Congress onwards (1995).

Each count indicates that *Artificial Intelligence* or *Machine Learning* was mentioned during a particular event contained in the Congressional Record, including the reading of a bill. If a speaker mentioned AI or ML multiple times within remarks, or multiple speakers mentioned AI or ML within the same event, it would appear only once as a result. Counts for *Artificial Intelligence* and *Machine Learning* are separate, as they were conducted in separate searches. Mentions of the abbreviations *"AI"* or *"ML"* are not included.

ImageNet

Return to ImageNet (pg 47)

See underlying data: (1) accuracy scores (2) training time

ImageNet accuracy scores

Source
Data on ImageNet accuracy was retrieved through the LSRVC leaderboard for years 2010- 2017. Validation data set performance for 2016 to 2018 was obtained through an arXiv literature review.

To highlight progress here, we have taken scores from the following papers:
ImageNet Classification with Deep Convolutional Neural Networks (2012)
Visualizing and Understanding Convolutional Networks (Nov 2013)
Going Deeper with Convolutions (Sept 2014)
Deep Residual Learning for Image Recognition (Dec 2015)
PolyNet: A Pursuit of Structural Diversity in Very Deep Networks (Nov 2016)
Squeeze-and-Excitation Networks (Sep 2017)
GPipe: Efficient Training of Giant Neural Networks using Pipeline Parallelism (Nov 2018)

The estimate of human-level performance is from Russakovsky et al, 2015. Learn more about the LSVRC ImageNet competition and the ImageNet data set.

Methodology
For ImageNet competition data (2010—2017), we collected scores from the leaderboards for each LSVRC competition hosted on the ImageNet website. Validation data set performance was obtained through a literature review. Only the highest performing scores from each year were recorded.

ImageNet training time

Source
ImageNet training time was collected through an arXiv literature review. The following papers were used:

June 8th, 2017: 60 minutes
November 1st, 2017: 30 minutes
November 12th, 2017: 15 minutes
July 30th, 2018: 6.6 minutes
November 13th, 2018: 3.7 minutes

APPENDIX 20
Parsing

See underlying data

Source

The Wall Street Journal portion of the Penn Treebank is a data set of sentences annotated with a constituency-based parse tree for each sentence. Section 23 of this data set has become the primary test set for research into automatic parsers. Learn more about Penn Treebank.

Methodology

Automatic parsers are evaluated by comparing the constituents of automatically generated parses to the constituents in gold parses from a test set. The precision and recall of the generated constituents are combined and reported as the F1 score. We report the F1 score of parsers on sentences in section 23 of the WSJ portion of the Penn Treebank. We report these scores for sentences of length <40 words and on the entire set of sentences where each are available. Learn more about constituency-based parse trees on the wikipedia page for parse trees.

We consulted the leaderboard maintained at nlpprogress.com for english constituency parsing. An additional literature review was conducted to confirm the listed scores and search for newer state-of-the-art results. New data points were combined with data from the 2017 AI Index report. Reported F1 scores of parsers we identified include ensembles, not just single models.

Nuance

- In the early days of automatic parsing research, parsers were typically evaluated only on sentences of length <40 words and length <100 words for computational and methodological reason. We have recorded the F1 scores of systems on sentences of length <40 words and on all sentences in the corpus when available.

Machine translation (BLEU)

See underlying data

Source
Data was pulled from EuroMatrix, which provides a matrix of BLEU scores for Workshop on Machine Translation (WMT) competitions.

Methodology
Each year WMT hosts a news translation task and provides new training and test data sets. Teams of attendees submit the translation systems they have built to participate in the translation task.

The main metric used by WMT aims at ranking the competing entries and does not allow year-over-year comparisons. It is also very labor intensive. We have fallen back on BLEU, an automatic method that does a rough comparison of the system translation to a number of human-generated translations. It is a modified version of precision, between 0 and 1, where higher is better. One can also calculate the average BLEU score of a machine translation system on a corpus of translation pairs. For each year, we have recorded the highest average BLEU score achieved by a system submitted to that year's news translation task for English to German and German to English.

EuroMatrix has recorded the BLEU scores of submissions to the news translation task since 2006 on the English to German and German to English language pairs. We selected the BLEU score of a top performing system each year, specifically using BLEU (11b), which defines a protocol for tokenizing sentences. When possible, we selected the score of a system that had a high-ranking BLEU score in both pairs as the representative for that year.

Nuance

- BLEU can be computed automatically and it has been shown to correlate with human judgment of translation quality. However, the metric cannot be used across corpuses and it can misleading to compare BLEU scores between systems.
- While the graphs trends upward generally, we can see one way this metric is flawed in 2017, when the BLEU scores fell significantly compared to the 2016 scores (though, the 2017 scores are still higher than the 2015 scores). It is unlikely that the performance of MT systems declined compared to 2016, but the evaluation scheme presented here is not perfect. However, looking at trends over larger time periods, the BLEU score can give an indication of progress in the area of machine translation.

APPENDIX 22
ARC

See underlying data

Source
AI2 Reasoning Challenge (ARC) is hosted by the Allen Institute for Artificial Intelligence. ARC performance data was retrieved from the ARC leaderboards. Find leaderboards for the easy set and the challenge set in the corresponding links.

Methodology
Participants download the ARC data set and submit to the leaderboard through the Allen Institute website.

Examples of questions from the Easy development corpus:

> Which technology was developed most recently? (A) cellular telephone (B) television (C) refrigerator (D) airplane [Grade 4]

> A student hypothesizes that algae are producers. Which question will best help the student determine if this is correct? (A) Do algae consume other organisms? (B) Which organisms consume algae? (C) Do algae use sunlight to make food? (D) Could an ecosystem survive without algae? [Grade 8]

Examples from the Challenge development corpus:

> Juan and LaKeisha roll a few objects down a ramp. They want to see which object rolls the farthest. What should they do so they can repeat their investigation? (A) Put the objects in groups. (B) Change the height of the ramp. (C) Choose different objects to roll. (D) Record the details of the investigation. [Grade 4]

> High-pressure systems stop air from rising into the colder regions of the atmosphere where water can condense. What will most likely result if a high-pressure system remains in an area for a long period of time? (A) fog (B) rain (C) drought (D) tornado [Grade 8]

Each question is worth one point. Models are allowed to give multiple answers, in which case a model that gives N answers gets 1/N points if one of its N answers is correct, and 0 otherwise. The overall score is the average of the scores of the individual questions.

The AI Index has only collected scores that beat scores from previous submissions. If a submission is lower than any of the previous submissions, it is not included in our visual.

Read more about the rules and submission guidelines here.

APPENDIX 23
GLUE

See underlying data

Source
GLUE benchmark data was pulled from the GLUE leaderboard. Learn more about the GLUE benchmark here.

Methodology
Participants download the GLUE tasks and submit to the leaderboard through the GLUE website. Scores are calculated for each task based on the task's individual metrics. All metrics are scaled by 100x (i.e., as percentages). These scores are then averaged to get the final score. For tasks with multiple metrics (including MNLI), the metrics are averaged.

On the leaderboard, only the top scoring submission of a user is shown or ranked by default. Other submissions can be viewed under the expanded view for each user. Competitors may submit privately, preventing their results from appearing. The AI Index visual does not include any private submissions. MNLI matched and mismatched are considered one task for purpose of scoring.

The AI Index has only collected scores that beat scores from previous submissions. If a submission is lower than any of the previous submissions, it is not included in our visual.

Read more about the rules and submission guidelines here.