OCTOBER 2020

Privately Held AI Companies by Sector

CSET Data Brief

AUTHOR
Santiago Mutis

Introduction

Artificial intelligence has become a strategic priority for the United States due to its potential to enhance productivity across several areas, including research and development, manufacturing, communications, health care, and transportation. The National Institute of Standards and Technology has characterized AI as one of the "Industries of the Future"[1] and encouraged the development of the field to expand research and commercial applications. As investment in AI grows,[2] companies and governments increasingly seek recommendations for bolstering their economic and national security.

In order to understand the commercial AI landscape, this brief identifies "AI companies" and classifies them into sectors to provide a picture of where such companies are clustering, their regional impact on the creation of businesses and jobs, how to optimize local educational pipelines, and which international AI hubs could be partnered with to boost the U.S. economic potential. It uses sectoral analysis to link companies to aggregate economic indicators, identify geographic distribution patterns, and generate comparative recommendations at a local and national level.

Selecting AI Companies

"AI companies" are defined in this paper as privately held, for-profit enterprises.[3] whose products and services rely significantly on artificial intelligence, or who produce hardware specifically designed to develop or implement artificial intelligence.[4] Removing publicly traded corporations from the dataset concentrates the analysis on AI-focused businesses and excludes companies that have an important role in AI but are highly diversified (e.g., Alphabet, Amazon, Facebook or Microsoft).[5]

Classifying Privately Held AI Companies Into Sectors

To determine a company's industrial sector, the author used the North American Industry Classification System,[6] a comprehensive business establishment and industry classification standard widely used by governments and organizations in the United States, Canada, and Mexico. NAICS codes are built with different levels of disaggregation and use a hierarchical system that ranges between two and six digit codes, allowing for diverse and contrasting levels of analysis. Each business establishment is

assigned one NAICS code based on the services or industrial operations performed on a single physical location. This enables more accurate geographic comparisons. Finally, roughly half of the companies in the dataset contain at least one NAICS code, which reduces the significance of the overall sample. More information on the use of NAICS codes in this analysis can be found in Appendix 2.

Based on the method described, the author classified U.S.-based, private AI companies into 89 different industries using four-digit NAICS codes (i.e. NAICS level 4). As shown in Table 1, the leading AI industries emerged as Software Publishers (49 percent of the companies) and Computer Systems Design and Related Services (17 percent), consistent with expectations for highly AI-focused businesses. These industries differ in that the former offers software as a product in a packaged version (e.g., Application Software, Computer Games, or Programming Languages), and the latter offers software design services to a client to meet particular needs (e.g., Computer Facilities Management Service, Software Consulting Service, or Systems Integration Design Service).

Table 1: Top 10 industries of private AI companies.

No.	Industry (NAICS description)	NAICS code	% Total
1	Software Publishers	5112	49%
2	Computer Systems Design and Related Services	5415	17%
3	Other Information Services	5191	5%
4	Data Processing, Hosting and Related Services	5182	4%
5	Navigational, Electromedical, and Control Instruments Manufacturing	3345	3%
6	Scientific Research and Development Services	5417	2%
7	Management, Scientific, and Technical Consulting Services	5416	2%
8	Advertising, Public Relations, and Related Services	5418	1%
9	Manufacturing and Reproducing Magnetic and Optical Media	3346	1%
10	Semiconductor and Other Electronic Component Manufacturing	3344	1%

As a hierarchical classification system, each digit of a NAICS code offers progressively narrower categories. The first two digits determine the economic sector, the third labels the subsector and the fourth the industry group.

This difference between these two industry groups extends to the aggregate two-digit NAICS code, where a distinction is made between the Information industry and the Professional Services industry. As is shown in Figure 1, 59 percent of the private AI companies analyzed are classified under the

Information code (NAICS code 51) and 24 percent under the Professional Services code (NAICS code 54).

Interestingly, the third industry group with a significant number of privately held AI companies—representing 10 percent of the sample—is the Manufacturing sector (NAICS code 33). This sector is composed of specific sub-industries like manufacturing of Navigational Instruments, Optical Instruments, and Semiconductor and Other Electronic Components. The number of private AI manufacturing companies is notable: it is seven times greater than private AI Finance and Insurance companies, eight times greater than private AI Healthcare companies, and nine times greater than private AI Retail companies. Policy and research targeting AI companies should account for the outsized role of AI manufacturing organizations in the AI field writ large. This finding shows no significant differences based on geography.

Figure 1: Top industries of private AI companies.

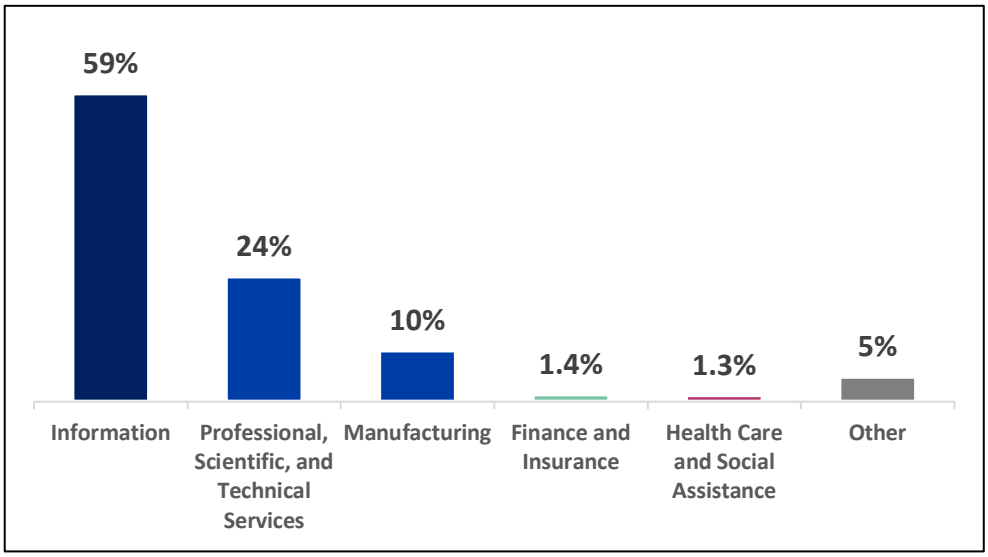

Figure 1 shows the top industries at a NAICS level 2 into which AI companies are grouped.

Geographic Distribution of Privately Held AI Companies

Based on NAICS code, we can map the geographic distribution of U.S.-based private AI companies. The detailed state-level distribution shows a concentration of AI companies in the following cities/metropolitan areas:[7] California (50.5 percent of total establishments), mainly around the San Francisco Bay Area, New York City (11 percent), Boston/Cambridge (9 percent), Austin (5 percent), and Seattle/King County metro (3 percent).[8]

Figure 2: Geographic distribution of AI companies (percentage of total).

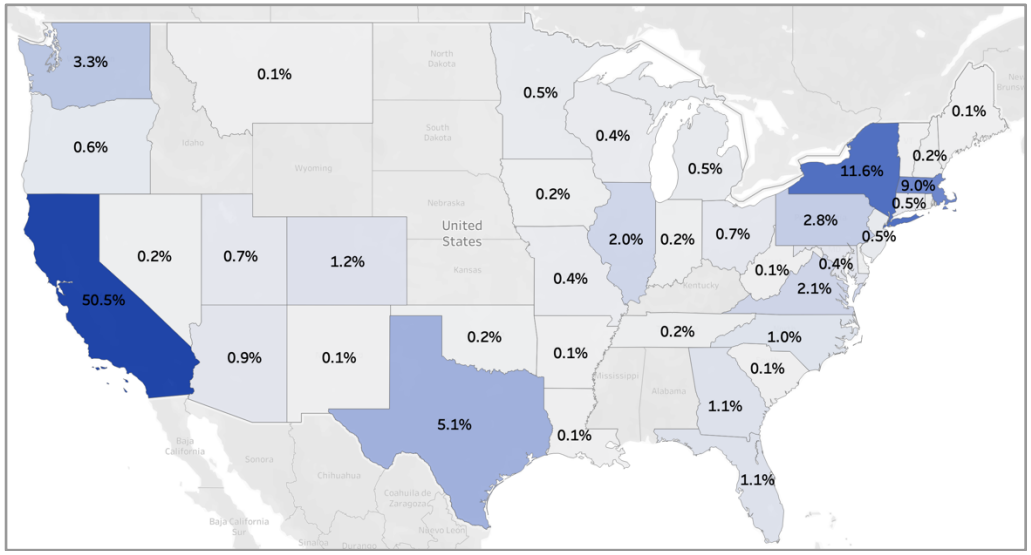

Figure 2 shows the percentage total of AI companies by state.

Additionally, we can adjust the data to account for population size and have a better sense of the intensity of concentration of AI companies. This measure shows a very similar geographic distribution to the overal share.

Figure 3: Intensity measure of AI companies (per 100,000 people).

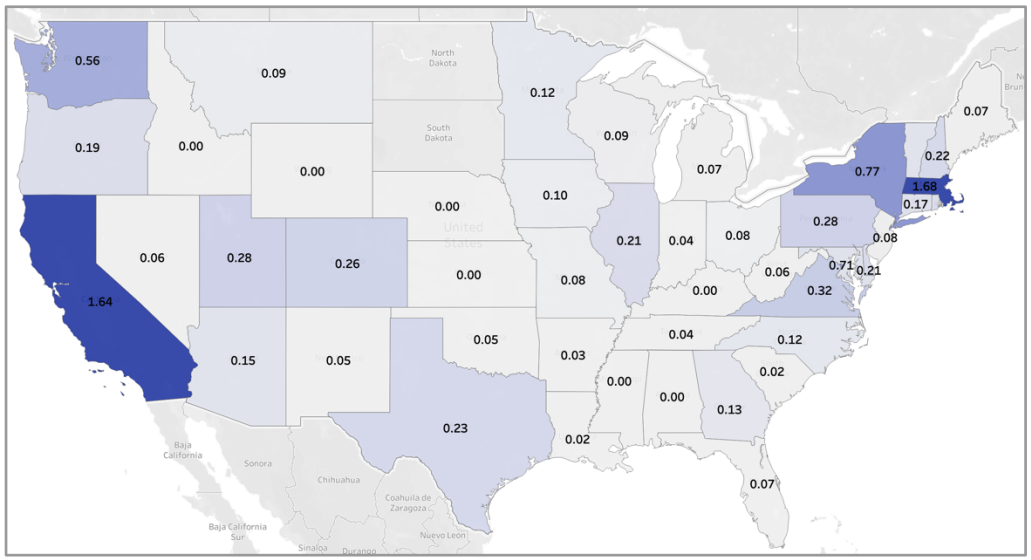

Figure 3 shows the number of AI companies per 100,000 people by state.

Policy Implications

As artificial intelligence adoption continues to expand and its value increases,[9] privately held AI companies will grow in diverse industries, impacting the economy, labor markets, and national security.

The manufacturing sector disproportionately generates economic activity,[10] so understanding how AI companies affect manufacturing will be crucial to leveraging the benefits of this technology for workers and the economy. With more than 10 percent of AI companies focused on manufacturing, this field offers a high-value focus for policymakers. The strength of the manufacturing sector also carries national security implications, as it determines the strength of the industrial base that produces specialized defense-critical products.[11] Ensuring a sustained pipeline of skilled labor with the capacity to support the specialized AI industry needed for advanced manufacturing will be critical to the competitiveness of the country, particularly in a globally competitive and high-paying AI sector.

Policymakers should also note the geographic distribution of AI companies. While AI-focused companies are concentrated in Silicon Valley, they represent only half of the national total. Distributing efforts to grow and support other AI clusters could encourage competition and diversify talent, use cases, and initiatives. Resources could be provided to AI research communities, specialized educational institutions, and industry and labor alliances outside of the main AI cluster in order to strengthen a wider range of locations and industries.

Finally, policymakers should request a detailed analysis of the economic impact of AI, focusing on the real-world effects of AI companies on adjacent industries and markets. This can be achieved by combining existing aggregate economic data from public sources and financial markets to determine the true impact of AI companies. With such a resource in hand, policymakers can accurately establish and possibly regulate the effect of AI organizations on variables like economic growth, employment, and business creation in different parts of the country.

Acknowledgments

Thank you to Igor Mikolic-Torreira, James Dunham, Zachary Arnold, Autumn Toney, Ilya Rahkovsky, Alexandra Vreeman, Theodoros Evgeniou, Nathan Goldschlag, and especially to Melissa Flagg and Dewey Murdick for helpful comments.

© 2020 by the Center for Security and Emerging Technology. This work is licensed under a Creative Commons Attribution-Non Commercial 4.0 International License.

To view a copy of this license, visit https://creativecommons.org/licenses/by-nc/4.0/.

Appendix 1: AI Company Selection

The author based the selection of AI companies on a CSET paper, "Tracking AI Investment: Initial Findings from the Private Market," which used financial data from Crunchbase and Refinitiv to identify AI companies.[2] Researchers looked across company descriptions in both of these sources for keywords and keyword combinations that indicate activity consistent with our definition of "AI companies."

This provides us with a list of 5,300 AI companies that was gathered in January 2020 by running a keyword search query through business descriptions in these commercial datasets. This data brief was constructed with an initial dataset of companies that is expanding and it will be updated with new data as CSET moves forward. Full details on the definitions, data, queries, keywords, and methodology can be found in the paper "Tracking AI Investment." [2]

Appendix 2: Industry Code Classification

We first identified a company's nationality to narrow down its sector. To do this, we used the organization's headquarters declared in commercial datasets as a proxy to determine nationality. We focused on the 2,600 U.S.-based companies from the AI companies list, then identified 1,300 U.S. companies from the AI companies list containing at least one NAICS code in the Crunchbase or Refinitiv datasets, giving priority to the Crunchbase code if both existed.

NAICS codes were selected for this analysis for a number of reasons:

They are built with different levels of disaggregation allowing for diverse and contrasting levels of analysis.

NAICS are built around establishments, not companies. An establishment is typically a single physical location, classified to an industry according to the primary business activity taking place there. NAICS codes do not offer guidance on the classification of enterprises (companies) composed of multiple establishments, but do provide geographical presence details.

The U.S. Census Bureau assigns and maintains only one NAICS code for each establishment based on its primary activity. However, other federal agencies assign more than one NAICS code per establishment, allowing businesses to register up to 10 codes for federal contractors. Nonetheless, in

the datasets used for this analysis there is only one NAICS code per company.

There is public economic data from federal agencies like the Census Bureau, the Bureau of Economic Analysis, and the Bureau of Labor Statistics categorized by NAICS codes.

NAICS codes can be used comparatively beyond North America through crosswalks that link industry codes to other widely used international sectoral codes like the International Standard Industrial Classification (ISIC).

NAICS codes were selected over other alternatives like the Global Industry Classification Standard (GICS), the Industry Classification Benchmark (ICB), and the Standard Industry Classification (SIC) because of their level of detail, general applicability, and up-to-date classification revision.

Endnotes

[1] Dr. Walter Copan, "Industries of the Future," Testimony to the Senate Committee on Commerce, Science and Transportation, 116th Congress, First Session, January 15, 2020, https://www.nist.gov/speech-testimony/industries-future.

[2] Zachary Arnold, Ilya Rahkovsky, Tina Huang, "Tracking AI Investment: Initial Findings from the Private Markets" (Center for Security and Emerging Technology, September 2020). https://cset.georgetown.edu/research/tracking-ai-investment/.

[3] Privately held AI companies are for-profit enterprises whose shares are not traded on a stock exchange, but can include state-supported companies.

[4] The methods used to identify companies whose "products and services rely significantly on artificial intelligence, or who produce hardware specifically designed to develop or implement artificial intelligence" can be found in Appendix 1 and its subsequent references.

[5] For different analytical purposes it would be crucial to include the top publicly traded corporations (like Google, Amazon, Facebook or Microsoft) as they lead in such dimensions as AI PhD hires, AI publications or AI patents. However, for the purpose of this report, excluding them creates a more focused analysis.

[6] United States Census Bureau, "North American Industry Classification System (NAICS) Frequently Asked Questions (FAQs)," Department of Commerce, November 13, 2019, https://www.census.gov/eos/www/naics/faqs/faqs.html.

[7] The five states with the highest number of privately held AI companies are California with 648 businesses, New York with 149 businesses, Massachusetts with 116 businesses, Texas with 66 businesses, and Washington with 43 businesses.

[8] A second group of cities has a significant number of private AI companies. This group is made up of New York City with 140 businesses, Boston/Cambridge with 97 businesses, Austin with 46 businesses, and Seattle/King County metro with 42 businesses.

[9] Tim Fountaine, Brian McCarthy and Tamim Saleh, "Building the AI-Powered Organization." Harvard Business Review (July 2019) https://hbr.org/2019/07/building-the-ai-powered-organization.

[10] Robert E. Scott, "The Manufacturing Footprint and the Importance of U.S. Manufacturing Jobs." Economic Policy Institute (January 22, 2015) https://www.epi.org/publication/the-manufacturing-footprint-and-the-importance-of-u-s-manufacturing-jobs/.

[11] Joel Yudken, "Manufacturing Insecurity: America's Manufacturing Crisis and the Erosion of the U.S. Defense Industrial Base." Cornell University ILR School (September 2010) https://digitalcommons.ilr.cornell.edu/cgi/viewcontent.cgi?article=1122&context=laborunions.