JANUARY 2021

The U.S. AI Workforce

Understanding the Supply of AI Talent

CSET Issue Brief

CSET
CENTER *for* SECURITY *and* EMERGING TECHNOLOGY

AUTHORS
Diana Gehlhaus
Santiago Mutis

Table of Contents

Executive Summary

Having access to the right talent is critical to maintaining a competitive edge in artificial intelligence. In the United States, policymakers are actively discussing legislative proposals to grow and cultivate a globally competitive domestic AI workforce. However, little data is available on the U.S. AI workforce and associated talent pipelines outside of the PhD segment.

Yet having access to good workforce data is critical to actually "winning" the competition for AI talent. This brief provides two contributions to better understand the U.S. AI workforce: (1) a definition of the AI workforce based on the government occupational classification system, identifying 54 occupations that either participate or could participate in AI product and application development, and (2) a preliminary assessment and characterization of the supply of AI talent, which consisted of 14 million workers in 2018 (about 9% of total U.S. employment).

Our definition of the AI workforce enables supply-side analysis that is more comprehensive than other commonly used sources, because it is linked to the federal occupation classification system. While many supply-side analyses of the AI workforce rely on sources such as LinkedIn, we use data from the U.S. Census Bureau. Our definition also enables greater analytic consistency across federal government and other datasets that link to this classification system, such as Burning Glass.

Key initial findings regarding the supply of U.S. AI workers include:

- The technical component of the AI workforce struggles with diversity, where a majority of workers are male and not representative in terms of race and ethnicity.
- Four-year college is a common pathway for many AI jobs; however, a sizeable share do not have four-year degrees, particularly in non-technical occupations.
- Degrees in engineering and computer science are among the top fields of study for technical AI occupations; however, non-technical degrees such as business are also common across AI occupations.
- While technical occupations garner much attention, the large number of non-technical occupations in the AI workforce suggests an approach to AI workforce policy that includes a range of education and training pathways.

This brief is the first in a three part series. The second paper will discuss U.S. AI labor market dynamics, while the third paper will provide actionable policy recommendations. Additional future research related to this series will explore topics such as the perceived rise of AI-related certifications and broader manpower and personnel policy implications for the DOD and national security community.

Introduction

There is a desire to understand the U.S. artificial intelligence (AI) workforce and the associated talent pipelines in the national security policymaking community. The National Security Commission on Artificial Intelligence (NSCAI) and the U.S. Congress, along with other national security and emerging technology policy experts are actively proposing legislation to enhance the AI workforce.[1] The Department of Defense (DOD) and its Joint Artificial Intelligence Center (JAIC) have explicitly expressed a need to understand the state of the U.S. AI workforce relative to the expected demand from both the government and the Defense Industrial Base sector.

However, policymakers are limited by the available data and analyses of a workforce not clearly or consistently defined. An exception is the PhD segment of the workforce, many of whom are non-U.S. citizens.[2] However, that is just one small—but important—segment of the AI workforce.[3] Not only is it unknown what the rest of the AI workforce consists of, but what the characteristics of these workers are, and the state of supply relative to demand.

This paper provides a first effort at mapping the U.S. AI workforce, and characterizes the supply of AI talent in the United States. We first provide an occupation-based definition of the AI workforce and categorize these occupations to correspond to their roles in the AI development process. We next provide several facts and figures to describe the people in these occupations, using U.S. government-collected statistics.

This paper is the first in a series of papers on the supply of the domestic AI workforce. The series will culminate in a data-driven approach to U.S. AI workforce policy and is designed as follows: Paper 1 (this Brief) provides a primer on what the domestic AI workforce is and who is in it; Paper 2 will provide labor market dynamics on the domestic AI workforce, looking at measures related to talent supply and demand over time; Paper 3 will be a policy report with actionable recommendations in the short-, medium-, and long-term to more effectively grow and cultivate the domestic AI workforce. The series is depicted in Figure 1.

Figure 1: Domestic AI Workforce Line of Research

Source: CSET.

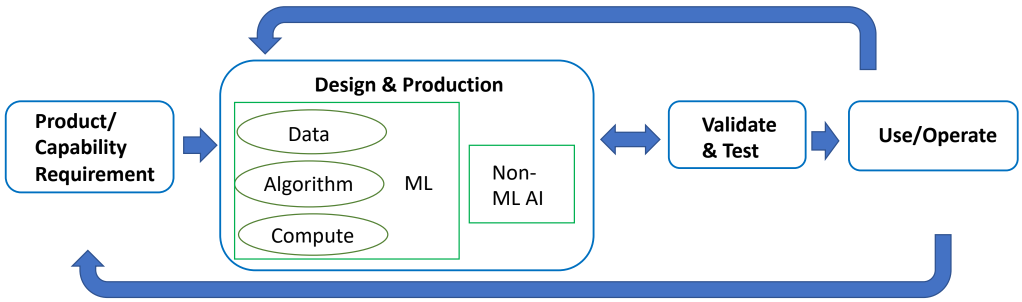
Defining the AI Workforce

Understanding the AI workforce requires clearly defining it. Here we explain our definition of the AI workforce. We use a model of the AI development process, a framework for categorizing the various roles and responsibilities, and official occupational databases administered by the federal government.

The first step in defining AI workers is to understand what roles and responsibilities are required to formulate, design, implement, and deliver AI applications (including machine learning (ML) applications). Here we include all types of AI applications, from playlist suggestions and voice recognition to traffic route mapping, healthcare diagnostics, and autonomous drone delivery.

Although the design and use of AI applications vary widely, we believe each follows a generic development process. Figure 2 visualizes this process in stages, and overlays the roles and responsibilities involved. It shows AI is a "team sport," with people performing different functions who possess a range of knowledge, skills, and abilities.

Figure 2: The AI Development Process



Source: CSET. "AI" denotes Artificial Intelligence; "ML" denotes Machine Learning.

The phases of the AI development process are as follows. First, there is the inception of a product or capability requirement, whether a new business product or identified operational need. Second, there is the design, development and production of the identified product or application. Third, there is verification, validation, testing, and evaluation (VVT&E) of the AI product or application. Fourth, the product or application is fielded and moves into the stage of operation and maintenance. In reality, there is much iteration, which is generically captured by the bi-directional blue arrow within the development process and by the "Use/Operate" arrow looping back to the beginning of the development and capability requirements process.

The people involved in the AI development process, as depicted in Figure 2, comprise our universe for AI workers. We define "AI Workforce" as the set of occupations that include people who are qualified to work in AI or on an AI development team, or have the requisite knowledge, skills, and abilities (KSAs) such that they could work on an AI product or application with minor training.

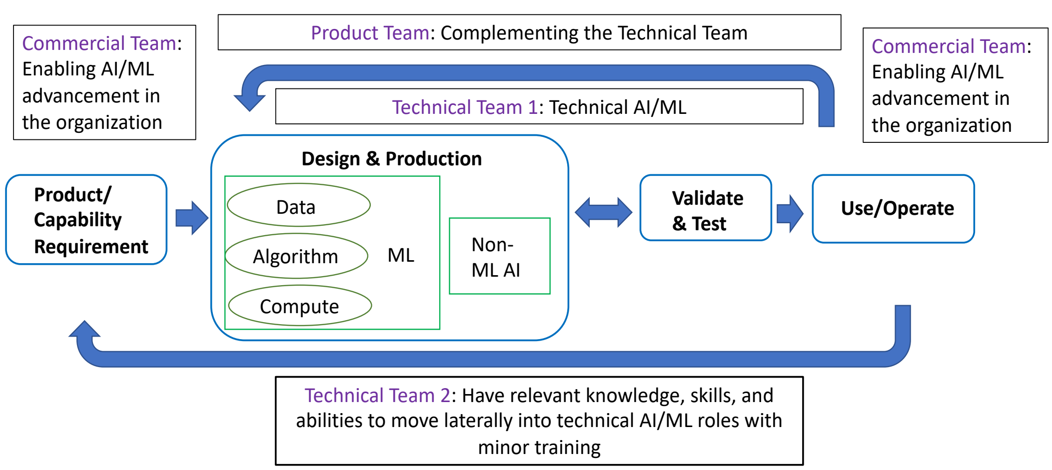
Given the large variety of required roles and responsibilities, a good understanding of the AI workforce distinguishes the different types of AI workers. There are the technical AI workers who are employed in occupations at the center of the AI development process, such as computer scientists, data architects, and software engineers. Other workers are in occupations that also play a critical role as part of the product team, such as program managers and compliance attorneys. A final category of workers is necessary at the institution or organizational level, in occupations that perform commercial functions such as business analysis, marketing and sales, and acquisition and procurement.

For our analysis of the AI workforce, we distinguish AI workers through four occupational categories:

(1) Technical Team 1 (Tech 1): occupations that are or could be actively working in AI, needed to provide technical inputs into AI applications, or could laterally move into an AI development role.
(2) Technical Team 2 (Tech 2): occupations that have the related KSAs to perform technical roles on an AI team, either as is or with some minimal additional training.
(3) Product Team: occupations that complement AI technical occupations in product development (such as project or product managers and legal compliance officers).
(4) Commercial Team: occupations that provide support for the scaling, marketing, or acquisition of AI at the organizational level.

Figure 3 overlays the four categories of AI workers on the AI development process shown in Figure 2.

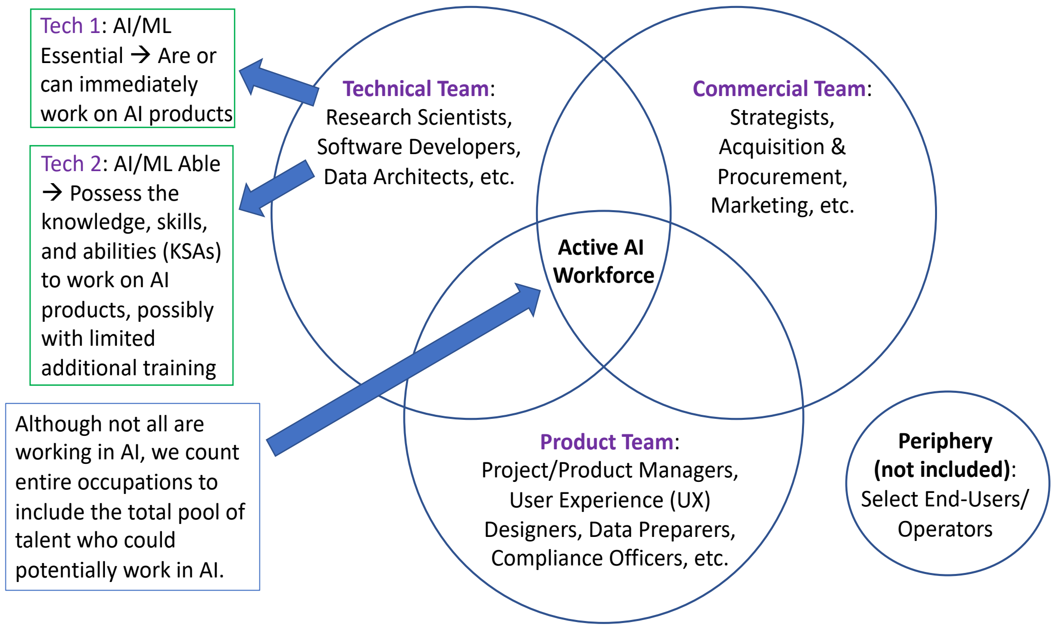
Figure 3: Roles & Responsibilities in the AI Development Process



Source: CSET. "AI" denotes Artificial Intelligence; "ML" denotes Machine Learning.

Using these categories, Figure 4 maps our AI occupation framework. It is possible, depending on organization, that some occupations could overlap across categories. However, for this brief, we assign a primary category for analysis. We note that use of the term "Technical Team" without distinction is jointly referring to Tech 1 and Tech 2 occupations. We also note we are not including peripheral occupations that use or operate AI which are not involved in the development process depicted above.

Figure 4: Framework for AI Occupations



Source: CSET. "AI" denotes artificial intelligence; "ML" denotes machine learning. Each occupation is assigned to one category; here we denote the intersection as the set of workers from each category working actively in AI. (Figure not to scale.)

We count entire occupations, regardless of the share actively working in AI, because we are interested in the total pool of possible AI talent—the set of people that have the requisite KSAs to work in AI or on an AI development team with minor training. Although it is not possible to estimate the exact shares working actively in AI, it is likely that some occupations will have higher shares than others.

Our framework for defining and categorizing AI occupations is a first iteration.* It considers all AI occupations without distinction for type of AI application or type of AI organization (e.g., AI developer, AI consumer). For

* As this research progresses, we may update or enhance this framework with what we learn. For example, we are conducting interviews with organizations engaged in AI to learn about their AI workforce.

a given company or application of AI, the composition and allocation of which AI occupations are needed will vary.

Identifying AI Occupations: Method Overview

Here we provide a brief overview of the methods used to identify and classify the set of AI occupations that comprise the AI Workforce. Appendix 1 of this paper provides a more detailed discussion.

To identify the set of AI occupations, we used a multi-step process to scan and review government-administered occupational databases. We started with a list of AI and AI-related keywords as identified by Burning Glass, a proprietary job posting aggregation database, such as "machine learning," "software," "modeling," and "mathematical."[4] We next scanned O*NET Online, a Department of Labor-administered database of occupation titles,[5] detailed work activities, and tasks, to see where these keywords matched.[*]

We manually reviewed the results in several iterations, looking at each occupation and associated tasks to see if and how it aligned within the process and framework presented in Figures 3 and 4. For example, both Web and Digital Interface Designers and Graphic Designers have tasks that involve creating designs, interfaces, prototypes, and layouts using market research, software, and esthetic design concepts to enhance product usability. These tasks included keywords such as "software," "product," and "user experience." It is likely both of these occupations comprise many "user experience" (UX) designers, a core part of many AI product teams. We included both occupations accordingly.

For occupations that were not able to be determined by the keyword and tasks analysis, we employed a database of over 33,000 job titles by occupation code provided by the U.S. Census Bureau to adjudicate. We considered job titles as a proxy for the types of workers being classified into a given occupation, which we used to make a final determination.[†]

As O*NET occupation codes are based on the Standard Occupational Classification System (SOC) taxonomy of occupations,[6] our final list is also

---

[*] Each O*NET occupation has between 4 and 40 associated tasks, with 20 tasks on average.
[†] This determination was made manually, assessing the types of roles (job titles) being classified in the occupation to the roles and responsibilities in our framework.

organized by SOC code.[*] A final step for our analysis was to map SOC codes to Census codes, a taxonomy maintained by the U.S. Census Bureau that is closely linked to the SOC.[†] This mapping enables our analyses to exploit both Labor Department and Census Bureau data, which is not possible from other commonly used data sources for supply-side analyses such as LinkedIn. It also enables consistency across datasets, both federal government and others that link to this classification system, such as Burning Glass.

Once we had a set of occupations, we assigned each to one of the four categories above. We manually determined the category assignment using the framework provided in Figure 3 along with existing literature describing various AI team compositions.[7] (We certainly acknowledge there are limitations to our approach which are listed in Appendix 1.)

For example, Operations Research Analysts had the following task containing the keyword "mathematical": "Formulate mathematical or simulation models of problems, relating constants and variables, restrictions, alternatives, conflicting objectives, and their numerical parameters." This task relates directly to the technical work conducted in designing and developing AI, and we assigned the occupation as Tech 1 accordingly.

U.S. AI Workforce

This process yielded a result of 54 detailed SOC-based occupations. The full list is provided in Appendix 1. As a metric, we assessed the average number of keywords in our AI occupation list. We found AI occupations had five times the average number of keyword hits returned per occupation of non-AI occupations.[‡]

---

[*] There are approximately 974 O*NET occupations and 867 detailed SOC occupation codes.

[†] We considered both the 2010 SOC and 2018 SOC, along with the associated versions of Census Occupation Codes for this analysis. This required crosswalking all versions.

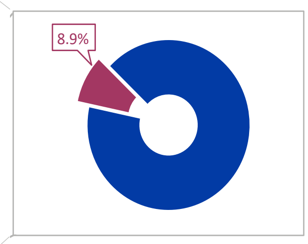
[‡] Individual tasks could have multiple keyword hits, each of which is counted separately. Our analysis showed AI occupations had an average of 95.2 keyword hits per occupation compared to 19.5 for non-AI occupations.

Size and Characteristics of the U.S. AI Workforce

We estimate the 54 occupations that comprise the U.S. AI workforce had 14 million workers in 2018. This translates into approximately 9% of all employed workers, as shown in Figure 5.* Tech 1 occupations, which include most computer and mathematical occupations (technical AI/ML), are the largest segment of the AI Workforce followed by Product Team occupations, Tech 2 occupations (AI-able), and Commercial Team occupations.†

Figure 5: AI Occupations Comprised 9% of Total U.S. Employment in 2018

	2018 Employment	Share of Total Employed
Technical Team 1	4,759,090	3.0%
Technical Team 2	3,006,580	1.9%
Product Team	4,350,740	2.8%
Commercial Team	1,908,340	1.2%
Total	**14,024,750**	**8.9%**



Source: American Community Survey 2018, CSET.

The remainder of this section provides several facts and figures about the supply of AI workers across the four categories. We compare that to the supply of all employed U.S. workers, where appropriate, for context. All data comes from the Census Bureau's American Community Survey (ACS) 2018 microdata and all shares are adjusted using the appropriate survey weights provided by the Census Bureau.[8]

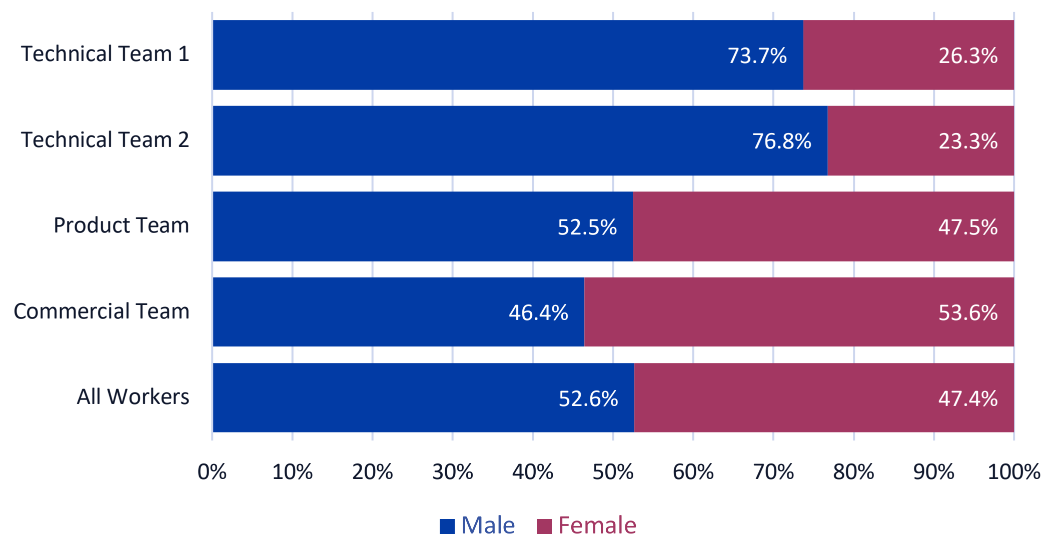
* As noted in the methodology, these estimates intentionally include all people working in AI occupations.

† Future research will explore how Tech 1 occupations have grown over the last few years, and the potential short-, medium-, and long-term implications for workforce development. For example, Current Population Survey data shows employment of software developers increased from 1.0 million in 2010 to 1.8 million in 2019, an increase of 77%.

First, we consider the demographic composition of the AI workforce.* The data shows a lack of racial and gender diversity, particularly in Technical Team occupations. The impact this could have—on everything from the quality of research, to increased prevalence of algorithmic bias, to the long-term deterioration of talent pipelines—is well-documented.[9]

Figure 6 shows the share of employment in each category by gender. Technical Team occupations have a far higher share of males than females, while Product and Commercial Team occupations share are a closer match to the share of all employed workers. In fact, Commercial Team occupations have a slightly higher share of female employment.

Figure 6: Technical Team Occupations are Mostly Male



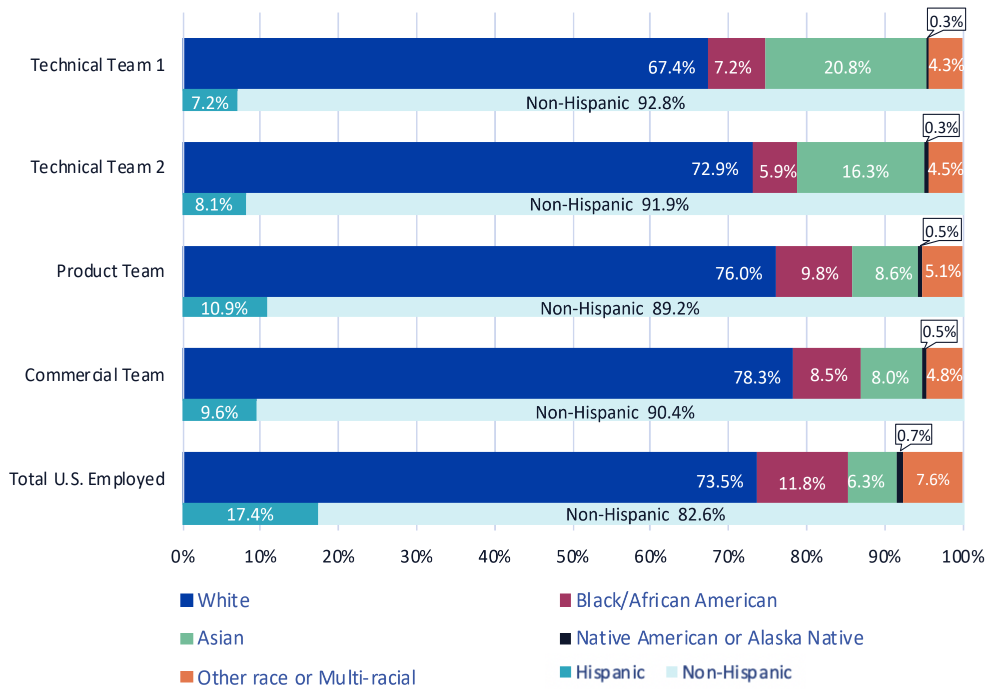
Source: American Community Survey 2018, CSET.

Figure 7 shows the racial distribution of workers employed in AI occupations along with Hispanic ethnicity. The share of African Americans and Hispanics employed in AI occupations across categories is lower than the average for all employed workers, with Tech 1 and Tech 2 occupations having the lowest shares. Alternatively, the share of workers of Asian descent in Tech 1 and Tech 2 occupations is notably higher. The share of White workers for Tech 2 and Product Team occupations are roughly in line with the average for all

* We also looked at nativity, or citizenship. This is reported in Appendix 2.

workers. However, notably the share is less than average for Tech 1 occupations and higher than average for Commercial Team occupations.

Figure 7: Technical Team Occupations Have a High Share of Asian Workers



Source: American Community Survey 2018, CSET.

Figure 8 shows the average age of AI workers. Workers in Technical and Commercial Team occupations have an average age in line with that for all workers. However, Product Team workers are older than the total employment average. Given almost 40% of this category consists of Management Analysts and Project Management Specialists, occupations of particular interest to the federal government, this could suggest a need to focus on training younger workers for these fields.*

* Management Analysts have an average age of 46.3.

Figure 8: Product Team Workers are Notably Older

	Mean Age
Technical Team 1	41.6
Technical Team 2	41.9
Product Team	43.5
Commercial Team	41.9
Total U.S. Employed	41.9

Source: American Community Survey 2018, CSET.

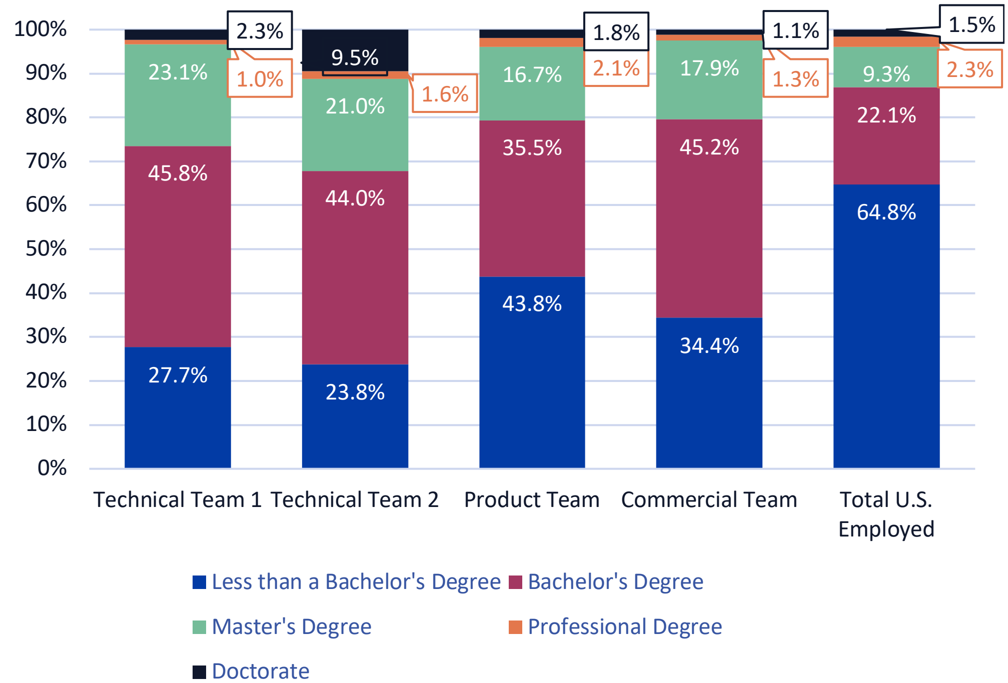
Although the average age for Tech 1 occupations is similar to that of all employed workers, some key occupations in this category are notably younger. For example, the average age of Software Developers is 39.1. The average age of mathematical science workers, which includes Statisticians and Data Scientists, is 38.9.

The next three figures consider educational attainment of the AI workforce. Figure 9 shows the educational attainment of the AI workforce by category.

Across all four categories of AI occupations, the majority of workers have at least a Bachelor's degree. Technical Team occupations have the highest shares with at least a Bachelor's degree, at over 70%, and over 25% have a graduate degree.* Product Team occupations have the most varied educational attainment, although over half have at least a Bachelor's degree. Commercial Team occupations fall in between, with roughly ⅔ of workers having at least a Bachelor's degree.

* About 3.6% of the AI Workforce has a doctorate degree. Tech 2 occupations have a notably higher share of PhD holders, likely because of the high concentration of science occupations.

Figure 9: Bachelor's Degrees Comprise the Highest Share by Education



Source: American Community Survey 2018, CSET.

One implication of the prevalence of college degrees—for Technical Team occupations in particular—is that many AI and AI-related jobs have responsibilities that require a college credential. Under this interpretation, it would follow that any strategy aimed at growing and cultivating the pipeline of AI talent must include more youth enrolling in and completing four-year college.

However, a counterpoint to the implied need for a four-year degree is that two factors unrelated to skill requirements could be driving some of these shares. First, at least in some Technical Team occupations, a larger share of younger people is filling these rapidly growing roles. And second, more young people are attending four-year college than previous generations, regardless of what jobs they go into.[10] In fact, research suggests about 40% of recent college graduates are underemployed, taking jobs that do not require a college credential, and that the bottom half of recent college graduates are experiencing a decrease in wage premiums.[11]

Moreover, even with a high share having four-year college or graduate degrees, a notable share of the AI workforce has less than a Bachelor's degree. For example, about 44% of workers in Product Team occupations have less than a Bachelor's degree. A third of workers in Commercial Team occupations have less than a Bachelor's degree, as do about a quarter of Technical Team occupation workers. The implication is that while many AI and AI-related jobs do likely require a four-year college degree, many may not. It follows that any AI workforce strategy should consider both baccalaureate and sub-baccalaureate pathways.

Looking at the composition of degrees for AI occupations suggests that at least some technical expertise is needed, particularly for Technical Team occupations.[*] Figure 10 lists the top 5 fields of study at the Bachelor's degree level by share. In both Tech 1 and Tech 2 occupations, engineering and computer science degrees comprised about half of all majors.

Aside from the high concentration of technical degrees in Technical Team occupations, Figure 10 also shows a second important takeaway: many workers with at least a Bachelor's degree still come from a range of undergraduate majors. Even in Technical Team occupations, this includes non-technical degrees in business and social sciences, along with fine arts in Product Team occupations and communications in Commercial Team occupations. Business is in the top 5 in each occupational category.

[*] Future work will address whether this technical expertise can only be obtained through a four-year college education.

Figure 10: AI Workers Have a Mix of Undergraduate Majors

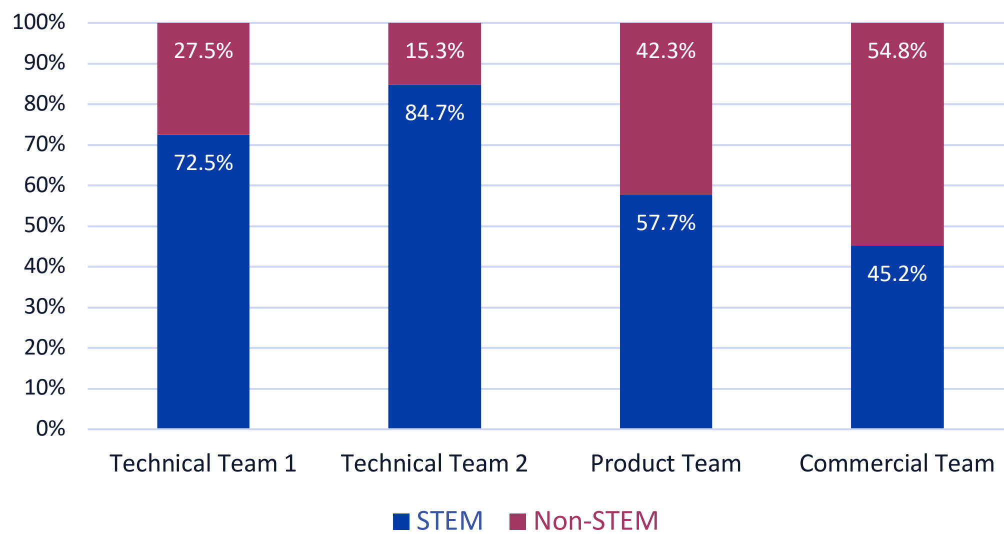
Technical Team 1		Technical Team 2	
Computer and Information Sciences	28.0%	Engineering	44.6%
Engineering	19.6%	Biology and Life Sciences	9.4%
Business	16.6%	Physical Sciences	8.2%
Social Sciences	5.6%	Business	7.6%
Mathematics and Statistics	4.2%	Computer and Information Sciences	7.0%
Other	25.9%	Other	23.2%
Product Team		**Commercial Team**	
Business	20.5%	Business	35.3%
Engineering	9.9%	Communications	10.8%
Medical and Health Sciences and Services	9.8%	Social Sciences	9.0%
Social Sciences	7.8%	Engineering	7.4%
Fine Arts	7.8%	Fine Arts	4.3%
Other	44.3%	Other	33.2%

Source: American Community Survey 2018, CSET.

Similar to having the widest range of educational attainment, Product Team occupations also had the most disparate composition of college majors. The top five majors accounted for just 56% of total degrees, relative to at least two-thirds for the other categories. (The next five majors comprised an additional 24%—biology, computer and information sciences, psychology, communications, and physical sciences respectively.)

In terms of STEM degrees, Product Team and Commercial Team occupations had the fewest. Figure 11 shows the share of undergraduate degrees in STEM fields. Across all majors, Commercial Team occupations had the fewest workers earning a degree in STEM fields, with fewer than half. This compares to over 70% in Tech 1 occupations, 85% in Tech 2 occupations, and almost 60% in Product Team occupations.*

Figure 11: Commercial Team Occupations Have the Fewest STEM Degrees



Source: American Community Survey 2018, CSET.

AI workers also earn more than the average U.S. worker. Figure 12 shows the average annual wage and salary income for AI workers in 2018. Tech 1 occupations have the highest earnings, on average, although all four categories earn more than the national average. It is likely some of the higher earnings, particularly for Technical Team occupations, are driven by higher educational attainment. In fact, we can see this clearly when looking at average earnings by educational attainment, which is provided in Appendix 2. However, on net, Tech 2 occupations have more education than Tech 1 occupations and a lower average salary. The implication is that additional factors, such as high demand for the KSAs held by Tech 1 workers, are driving higher than average salaries.

* Using the definition of STEM fields of study from the U.S. Department of Homeland Security and translating it into Census undergraduate degree categories: https://www.ice.gov/sites/default/files/documents/Document/2016/stem-list.pdf.

Figure 12: AI Workers Earn More on Average

	Mean Wage & Salary Income
Technical Team 1	$ 95,010
Technical Team 2	$ 87,380
Product Team	$ 69,220
Commercial Team	$ 78,360
Total U.S. Employed	$ 50,340

Source: American Community Survey 2018, CSET.

There is also a sizable range in the average earnings of AI workers. Software Developers, for example, earned about $112,000 annually on average, and Computer Research and Information Scientists earned about $102,000 annually on average. In contrast, both Web and Digital Interface Designers and Graphic Designers earned about $43,000 annually on average, less than the average for total employed.

Finally, Figure 13 shows the employment composition by industry, looking at the top 5 industries for each category by two-digit industry code.* Although there is a range of industries represented, Professional, Technical, and Scientific Services is a top employer in all four categories. Within this sector, Computer Systems Design and Related Services (NAICS 5415) was the top employer for Tech 1 and Tech 2 occupations, while Management, Scientific, and Technical Consulting Services (NAICS 5416) was the top employer for Product Team and Commercial Team occupations. Metal, Machinery, and Equipment Manufacturing (NAICS 33)† and Finance and Insurance (NAICS 52) are also top employers across groups. Healthcare (NAICS 62) is a top employer only for Product Team and Commercial Team occupations,

* The North American Industry Classification System (NAICS) is the industry taxonomy used by the federal government.

† There is no formal differentiation between NAICS 31, 32, and 33, which comprise the manufacturing sector. Our designation for NAICS 33 is a rough approximation of this manufacturing segment given the description of detailed industries to facilitate interpretation. A similar designation was made for NAICS 32 in Figure 11. See https://www.census.gov/cgi-bin/sssd/naics/naicsrch for more.

potentially suggesting a more delayed adoption of advanced technologies relative to other top sectors. However, should organizations in this industry increase adoption, this also suggests they have many of the key team players.

Figure 13: AI Workers are in a Range of Industries

Technical Team 1			Technical Team 2		
NAICS	Industry	Share	NAICS	Industry	Share
54	Professional, Scientific, and Technical Services	37.8%	33	Metal, Machinery, and Equipment Manufacturing	27.6%
52	Finance and Insurance	12.2%	54	Professional, Scientific, and Technical Services	25.2%
92	Public Administration	7.8%	92	Public Administration	7.6%
33	Metal, Machinery, and Equipment Manufacturing	7.1%	61	Educational Services	7.4%
51	Information	6.6%	32	Materials Manufacturing	4.8%*
N/A	Other	28.6%	N/A	Other	27.4%

Product Team			Commercial Team		
NAICS	Industry	Share	NAICS	Industry	Share
54	Professional, Scientific, and Technical Services	26.7%	54	Professional, Scientific, and Technical Services	20.8%
62	Health Care and Social Assistance	25.2%	33	Metal, Machinery, and Equipment Manufacturing	12.6%
33	Metal, Machinery, and Equipment Manufacturing	7.7%	92	Public Administration	7.5%
92	Public Administration	6.0%	62	Health Care and Social Assistance	6.5%
52	Finance and Insurance	5.5%	52	Finance and Insurance	6.3%
N/A	Other	28.8%	N/A	Other	46.3%

*NAICS 51—Information—also had 4.8% of Tech 2 employment.
Source: American Community Survey 2018, CSET.

Perhaps most interestingly, federal, state, and local government (Public Administration, NAICS 92) is also a top employing industry across all AI workforce categories. Within the government, the national security and international affairs community is the top employer across AI occupations.*

One possible implication is that the federal government may have more AI talent than previously believed, given the large amount of literature suggesting such talent is scarce in government and challenging to recruit and retain.[12] It may be that this talent is working on non-AI projects or is not employed in roles that enable participation in AI development and/or deployment. In that case, with some minor upskilling or training, the government could refocus this talent to AI and other emerging technologies. Future research will explore this further.

* The American Community Survey provides employment estimates at the detailed industry level. NAICS 928, National Security and International Affairs, is a detailed industry code within NAICS 92.

Conclusion

There is a need for data on the domestic AI workforce to better inform workforce development and national security policy. Currently, little information is available on the supply of AI talent in the United States, and no real consensus exists on how to appropriately define the AI workforce.

This paper is the first in a three-part series on the U.S. AI workforce. The second paper will consider labor market dynamics in terms of supply and demand for the domestic AI workforce. The third paper will assess what federal, state, and local government policy levers are available to facilitate a sufficient domestic AI workforce pipeline.

This first brief provides two contributions to advance our understanding of the U.S. AI workforce: (1) a definition of the AI workforce based on the government occupational classification system, identifying 54 occupations that either participate or could participate in AI product and application development, and (2) a preliminary assessment of the supply of AI talent, which consisted of 14 million workers in 2018 (about 9% of total employed).

Our definition of the AI workforce enables analysis of the full set of knowledge, skills, and abilities necessary to design, develop, and execute AI: to research cutting-edge AI algorithms; to apply these algorithms to real-world problems; to design practical AI applications; and to develop, commercialize, and deploy actual AI products. By linking our definition to an established occupational classification system, we can systematically analyze the AI workforce with the wealth of existing labor statistics. We can also standardize AI workforce analytics across research products and update these analytics as new data becomes available.

We find that, consistent with existing literature, the AI workforce struggles with diversity, particularly in Technical Team AI occupations. The findings presented here support the assertion that AI workers are less diverse than the total population; for example, Technical Team occupations are mostly male and few are African American or Hispanic. More needs to be done to promote diversity and inclusion in AI and AI-related fields, to sustain talent pipelines and ensure the competitive standing of U.S. human capital long-term.

In terms of credentialing, four-year college is a critical pathway to many AI jobs, especially for Technical Team occupations. Moreover, technical degree programs in engineering and computer science are common training vehicles for Technical Team AI talent. However, it is also important to note that many AI-related jobs in Product and Commercial Team occupations do not require a four-year degree. Even for workers with four-year college degrees, many come from non-technical fields of study. Any strategy to build and sustain the AI workforce should therefore consider the range of education and training pathways.

Additional future research related to this series will explore other topical issues on domestic AI talent pipelines and career pathways, such as the perceived rise of AI-related certifications. It will also include an examination of broader manpower and personnel policy implications for the DOD and national security community.

Acknowledgments

We would like to thank Igor Mikolic-Torreira, Remco Zwetsloot, Dewey Murdick, Ilya Rahkovsky, James Dunham, Tessa Baker, Matt Mahoney, Richard Fry, Ron Hodge, and Jonathan Rotner for their reviews and suggestions. We would also like to thank Anthony Martinez and Jesse Lastunen for their technical advice, Kayla Goode for visualizations, and Melissa Deng, and Alex Friedland for editorial support.

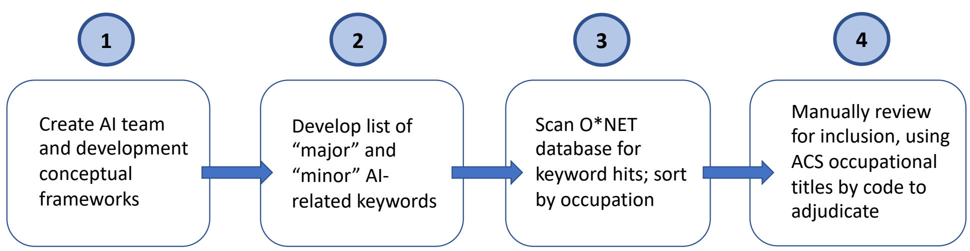
Appendix 1: AI Workforce Methodology

We developed a definition of the AI workforce in two parts: (1) we identified the set of AI occupations, and (2) we assigned each of these occupations a category based on where the occupation fit in the AI development process.

For (1), we identified what an "AI occupation" is in four steps as visualized in Figure A1. For (2), we categorized each occupation using the framework provided in Figure 3, along with existing literature describing various AI team compositions.

For our data sources, we used Burning Glass (a proprietary job posting data provider), O*NET Online (an occupational database administered by the U.S. Department of Labor), and the Standard Occupational Classification System (SOC) taxonomy of occupations, as maintained by the U.S. Bureau of Labor Statistics.*

Figure A1: We Identify AI Occupations in Four Steps



Source: CSET.

For the first step in identifying AI occupations, we developed a list of AI and AI-related keywords. The initial list was set using the Burning Glass AI and AI-adjacent keyword list[13] and supplemented with literature on AI team composition.[14] To maximize inclusion in our search results, we separated out keyword phrases into components (e.g., included "artificial intelligence" along with "artificial" and "intelligence").

We designated each keyword as major" or "minor" depending on relevance to AI. For example, words that could have many non-AI related uses were designated "minor" (e.g., "data" or "computer"). Those designated "major"

* We considered both the 2010 SOC and 2018 SOC, along with the associated versions of Census Occupation Codes for this analysis. This required crosswalking all versions.

will have a greater importance in our review of results because of the likely greater relevance to AI (e.g., “machine learning”). The full set of keywords by “major” or “minor” assignment is in Figure A2 below.

Figure A2: AI and AI-related Keywords by “Major” or “Minor” Designation

artificial intelligence	Major	artificial	Minor
computer vision	Major	intelligence	Minor
automate	Major	computer	Minor
data mining	Major	data	Minor
data science	Major	economic	Minor
machine learning	Major	economics	Minor
user experience	Major	automation	Minor
neural net	Major	IT automation	Minor
deep learning	Major	machine	Minor
reinforcement learning	Major	medical research	Minor
supervised learning	Major	research	Minor
unsupervised learning	Major	signal	Minor
cluster	Major	statistic	Minor
mathematical	Major	processing	Minor
modeling	Major	test	Minor
software	Major	compute	Minor
mathematics	Major	Product	Minor
natural language	Major	engineer	Minor
robotic	Major	program	Minor
scripting	Major		
systems design	Major		
DevOps	Major		

Source: CSET.

With this keyword list, we scanned the O*NET Online database. To be as inclusive as possible, we scanned for hits in occupation titles, job tasks, and detailed work activities.* We created a pivot table to sort all keyword return search results by SOC code, looking at the associated job task for any title or detailed work activity (DWA) hit to keep all entries comparable (tasks are also the most detailed in terms of descriptive information).

We next manually reviewed each occupation to determine inclusion as an AI occupation. We first considered the keywords included in the search results for the occupation: Were there any major keywords? How many major or minor keywords were included and which ones? How many associated tasks?

Since our list of keywords was intentionally broad to ensure the greatest level of inclusion, our delineation of "major" versus "minor" keywords assisted in the initial screen. We gave greater weight, qualitatively, to occupations that included major keywords. Whereas we were looking for evidence to include occupations that contained only minor keywords, we were looking for evidence to not include occupations that contained major keywords.

For occupations with only minor keywords, we made an initial cut given what roles and responsibilities (KSAs) we were looking for in the AI Workforce Framework (depicted in Figures 3 and 4 in the main report). For example, some occupations were clearly not relevant, such as retail salespersons. For each occupation that passed our initial scan, we extracted the 3-6 most relevant occupational tasks given our definition of the AI workforce, noting any major keywords where applicable (since those had greater significance).

With this list, we manually reviewed each occupation a second time for inclusion. We asked two questions: (1) Does this fit into any of the roles and responsibilities identified in the AI Workforce Framework? And (2) If not, do people working in this occupation have related knowledge, skills, and abilities (KSAs) such that they could perform the tasks for occupations in the Framework, with minor training? We believe occupations included in (2) are an important part of the AI talent pipeline, because they can migrate to AI occupations with relative ease.[15] For example, among their tasks, Medical and Health Services Managers "develop and maintain computerized record

* Each O*NET occupation has between 4-40 distinct tasks, with an average 20 tasks. The list of associated tasks is derived from surveys administered by the Department of Labor to persons employed in the occupation.

management systems to store and process data" and "maintain awareness of advances in computerized diagnostic and treatment equipment, data processing technology, and government regulations." And among their tasks, Aircraft Mechanics and Service Technicians "modify aircraft structures, space vehicles, systems, or components" and "fabricate defective sections or parts, using metal fabricating machines, saws, brakes, shears, and grinders." With some minor training, we believe Medical and Health Services Managers could manage the development of medically-oriented AI applications, and that Aircraft Mechanics and Service Technicians could build or repair AI-enabled hardware such as machines, devices, and equipment.

We assigned each occupation as "Yes," "No," or "Maybe" depending on our assessment. In the cases of "Yes" and "No" the delineation was clear based on the keywords and tasks: Computer Research Scientists and Software Developers were a "Yes" while Interpreters and Translators and Musicians were a "No." However, some occupations were less clear. For these we needed additional information. We relied upon a list of all job titles by occupation provided by the Census Bureau to assess which types of workers were being classified in that occupation, and how that answered (1) or (2) above.

The process yielded a result of 54 detailed SOC-based occupations. A full list of occupations, categorization, and 2018 employment is provided in Figure A3.

We acknowledge openly that not all individuals employed in these occupations are working on AI. However, it is a set of people that are qualified to work in AI or on an AI development team, or have the requisite KSAs such that they could work on an AI product or application with minor training.

Given the "overcounting" of people working directly in AI, there were another three occupations we did not include because of additional ambiguity in available data. These occupations were too aggregated in Census codes to isolate the detailed occupation of interest. These are postsecondary teachers (of which we are interested in a few disciplines), an "all other" managers category (of which we are interested in a few types), and sales representatives, wholesale and manufacturing (of which we are only interested in technical and scientific products).*

Figure A3: AI Occupations by 2018 Employment and Type

2018 Census Title	2018 Employment	Type*	2018 Census Title	2018 Employment	Type*
Computer and Information Research Scientists	31,090	T1	Industrial Engineers, including Health and Safety	225,220	T2
Computer and Information Systems Managers	583,820	T1	Mechanical Engineers	291,600	T2
Computer Hardware Engineers	49,020	T1	Medical Scientists	145,000	T2
Computer Network Architects	105,380	T1	Physical Scientists, All Other	288,320	T2
Computer Occupations, All Other	828,510	T1	Web Developers	113,050	T2
Computer Programmers	432,220	T1	Aircraft Mechanics and Service Technicians	185,810	PT
Computer Systems Analysts	545,250	T1	Clinical Laboratory Technologists and Technicians	338,710	PT
Database Administrators and Architects	119,140	T1	Computer, Teller, and Office Machine Repairers	151,090	PT
Information Security Analysts	110,280	T1	Data Entry Keyers	312,220	PT

* Looking at data from the Bureau of Labor Statistics' Occupational Employment Statistics (OES), which does have employment at the more detailed level for these occupations, we found that our occupations of interest comprised less than 25% of the occupation's total employment.

2018 Census Title	2018 Employment	Type*	2018 Census Title	2018 Employment	Type*
Mathematical Science Occupations, All Other (inc. Data Scientists)	163,670	T1	Electrical and Electronic Engineering Technologists and Technicians	85,070	PT
Mathematicians	Other math**	T1	Electrical and Electronics Repairers, Industrial and Utility	33,300	PT
Network and Computer Systems Administrators	215,420	T1	Graphic Designers	316,450	PT
Operations Research Analysts	153,700	T1	Legal Support Workers, All Other	57,900	PT
Software Developers	1,350,240	T1	Management Analysts	907,970	PT
Software Quality Assurance Analysts and Testers	71,340	T1	Medical and Health Services Managers	730,700	PT
Statisticians	Other math**	T1	Natural Sciences Managers	24,780	PT
Aerospace Engineers	133,400	T2	Other Engineering Technologists and Technicians, Except Drafters	370,020	PT
Architectural and Engineering Managers	177,200	T2	Project Management Specialists	726,120	PT
Astronomers and Physicists	10,650	T2	Statistical Assistants	31,850	PT
Atmospheric and Space Scientists	9,560	T2	Web and Digital Interface Designers	78,750	PT
Bioengineers and Biomedical Engineers	14,680	T2	Business Operations Specialists, All Other	348,270	CT
Biological Scientists	83,810	T2	Logisticians	155,570	CT

2018 Census Title	2018 Employment	Type*	2018 Census Title	2018 Employment	Type*
Computer Support Specialists	681,150	T2	Market Research Analysts and Marketing Specialists	348,980	CT
Economists	26,230	T2	Marketing Managers	530,990	CT
Electrical and Electronics Engineers	203,510	T2	Purchasing Agents, Except Wholesale, Retail, and Farm Products	284,510	CT
Engineers, All Other	560,690	T2	Purchasing Managers	202,070	CT
Geoscientists and Hydrologists, Except Geographers	42,520	T2	Sales Engineers	37,960	CT

*T1 = Technical Team 1; T2 = Technical Team 2; PT = Product Team; CT = Commercial Team
**The Census Bureau aggregated employment into "Mathematical science occupations, all other" due to the small sample size.
Source: American Community Survey 2018, CSET.

Limitations

The taxonomy of the AI Workforce presented here is a first attempt at defining AI occupations. We will continue to iterate on the Framework and taxonomy as this research progresses. We note the following limitations in our approach:

- We may have not considered the full universe of relevant keywords.
- We are counting selected occupations in full even though only some percentage are likely working in AI. We consider our analysis instead as the people with the requisite skills, or the full domestic AI workforce.
- Final determinations of inclusion and status designation are at the team's discretion and judgement.
- The Census Bureau assigns people to occupations by their job title; however, in cases of uncertainty, they assign based on what the respondent said were their job duties. People could have the same title and be put into different SOC codes.

- The 2010 SOC classifies a fair number of occupations of interest as residual occupations (e.g., "Engineers, all other"). O*NET has 8-digit level codes which provides greater detail on why we are including some 6-digit SOC codes that would not be intuitive. Some sub-sets of a given 6-digit SOC code may be AI occupations.
- There is not perfect matching between Census occupation codes and O*NET/SOC codes, especially in the 2018 update.

Appendix 2: Selected Detailed Results

This appendix provides full tables for Figures 6 and 8 in the main report (Tables B1 and B2, respectively), along with two additional tables for citizenship (Table B3) and average wage by educational attainment (Table B4).

Table B1: Race and Hispanic Ethnicity by AI Occupation Category

	Technical Team 1	Technical Team 2	Product Team	Commercial Team	Total U.S. Employed
White	67.4%	72.9%	76.0%	78.3%	73.5%
Black/African American	7.2%	5.9%	9.8%	8.5%	11.8%
Asian	20.8%	16.3%	8.6%	8.0%	6.3%
Native American or Alaska Native	0.3%	0.3%	0.5%	0.5%	0.7%
Other race or Multi-racial	4.3%	4.5%	5.1%	4.8%	7.6%
Hispanic	7.2%	8.1%	10.9%	9.6%	17.4%

Source: American Community Survey 2018, CSET.

Table B2: Educational Attainment by AI Occupation Category

	Technical Team 1	Technical Team 2	Product Team	Commercial Team	Total U.S. Employed
HS or Less	5.9%	5.5%	12.8%	10.3%	33.4%
Some College	13.9%	11.0%	19.7%	16.7%	22.2%
Associate's Degree	8.0%	7.4%	11.3%	7.4%	9.1%
Bachelor's Degree	45.8%	44.0%	35.5%	45.2%	22.1%
Master's Degree	23.1%	21.0%	16.7%	17.9%	9.3%
Professional Degree	1.0%	1.6%	2.1%	1.3%	2.3%
Doctorate	2.3%	9.5%	1.8%	1.1%	1.5%
Less than a Bachelor's Degree	27.7%	23.8%	43.8%	34.4%	64.8%
Bachelor's Degree or Higher	72.3%	76.2%	56.2%	65.6%	35.2%

Source: American Community Survey 2018, CSET.

Table B3: Citizenship by AI Occupation Category

	Technical Team 1	Technical Team 2	Product Team	Commercial Team	Total U.S. Employed
U.S. Citizen	73.5%	76.6%	85.4%	87.2%	82.7%
Naturalized Citizen	12.8%	12.4%	9.2%	7.6%	8.8%
Not a Citizen	13.7%	11.0%	5.4%	5.2%	8.5%

Source: American Community Survey 2018, CSET.

Table B4: Mean Wage and Salary Income by Educational Attainment

	Technical Team 1	Technical Team 2	Product Team	Commercial Team	Total U.S. Employed
Less than a Bachelor's Degree	$69,900	$59,120	$48,980	$52,130	$33,720
Associate's Degree	$73,470	$65,580	$54,830	$56,580	$44,160
Bachelor's Degree	$96,570	$86,090	$72,580	$82,640	$66,930
Master's Degree	$114,520	$106,050	$98,050	$111,130	$84,460
Professional Degree	$105,270	$109,360	$130,110	$115,790	$139,810
Doctorate	$153,250	$114,230	$118,840	$113,300	$108,430

Source: American Community Survey 2018, CSET.

Endnotes

[1] National Security Commission on Artificial Intelligence, "Interim Report and Third Quarter Recommendations," October 2020, https://drive.google.com/file/d/1jg9YlNagGI_0rid-HXY-fvJOAejlFIiy/view?usp=sharing.
[2] Remco Zwetsloot, "Strengthening the U.S. AI Workforce" (Center for Security and Emerging Technology, September 2019), https://cset.georgetown.edu/research/strengthening-the-u-s-ai-workforce/.
[3] Element AI, "2019 Global AI Talent Report," accessed September 13, 2020, https://www.elementai.com/news/2019/2019-global-ai-talent-report.
[4] Melissa Flagg and Autumn Toney, "U.S. Demand for AI-Related Talent" (Center for Security and Emerging Technology, August 2020), https://cset.georgetown.edu/research/u-s-demand-for-ai-related-talent/. A full list of keywords is included in Appendix 1 of this report.
[5] O*NET Resource Center, "The O*NET Content Model," U.S. Department of Labor, Employment and Training Administration, accessed September 13, 2020, https://www.onetcenter.org/content.html; O*NET Resource Center, "About O*NET," U.S. Department of Labor, Employment and Training Administration, accessed September 13, 2020, https://www.onetcenter.org/overview.html.
[6] U.S. Bureau of Labor Statistics, "Standard Occupation Classification," U.S. Department of Labor, accessed September 13, 2020, https://www.bls.gov/soc/#:~:text=All%20workers%20are%20classified%20into,groups%2C%20and%2023%20major%20groups.
[7] Tim Fountaine, Brian McCarthy, and Tamim Saleh, "Building the AI-Powered Organization," *Harvard Business Review*, July-August 2019 issue, https://hbr.org/2019/07/building-the-ai-powered-organization.
[8] "American Community Survey," IPUMS USA, accessed September 13, 2020, www.ipums.org.
[9] Sara Harrison, "Five Years of Tech Diversity Reports—and Little Progress," *WIRED*, October 1, 2019, https://www.wired.com/story/five-years-tech-diversity-reports-little-progress/.
[10] National Center for Education Statistics, "Table 303.70. Total Undergraduate Fall Enrollment in Degree-granting Postsecondary Institutions, by Attendance Status, Sex of Student, and Control and Level of Institution: Selected Years, 1970 through 2029," U.S. Department of Education, accessed September 13, 2020, https://nces.ed.gov/programs/digest/d19/tables/dt19_303.70.asp.
[11] "The Labor Market for Recent College Graduates," Federal Reserve Bank of New York, accessed September 13, 2020, https://www.newyorkfed.org/research/college-labor-market/college-labor-market_underemployment_rates.html. See also: Alexandre Tanzi and Katia Dmitrieva, "In Hot U.S. Jobs Market, Half of College Grads Are Missing Out," *Bloomberg*, February 14, 2020, https://www.bloomberg.com/news/articles/2020-02-14/in-hot-u-s-jobs-market-half-of-college-grads-are-missing-out.

[12] Jory Heckman, "Commission Looks to AI Reserve Corps, Service Academy to Fill Federal Workforce Gaps," *Federal News Network*, July 22, 2020, https://federalnewsnetwork.com/artificial-intelligence/2020/07/commission-looks-to-ai-reserve-corps-service-academy-to-fill-federal-workforce-gaps/.
[13] Flagg and Toney, "U.S. Demand for AI-Related Talent."
[14] "2020 AI Talent Report: Current Landscape & Market Trends" (TalentSeer, accessed September 13, 2020), https://www.talentseer.com/2020-ai-talent-report#:~:text=With%20AI%20technology%20predicted%20to,AI%20research%2C%20development%20and%20commercialization.&text=This%20report%20provides%20an%20in,the%20talent%20market%20in%202020. See also: Fountaine, McCarthy, and Saleh, "Building the AI-Powered Organization."
[15] Penny Mealy, R Maria del Rio-Chanona, and J. Doyne Farmer, "What You Do at Work Matters: New Lenses on Labour," March 18, 2018, http://dx.doi.org/10.2139/ssrn.3143064.