Data Brief

# Identifying the AI Development Workforce

**Authors**
Sonali Subbu Rathinam
Ronnie Kinoshita

CSET CENTER *for* SECURITY *and* EMERGING TECHNOLOGY

June 2026

## Executive Summary

Assessing the artificial intelligence workforce is critical for developing effective training pipelines and policy. Assessment requires reliable measures of AI talent demand and supply. Most existing methodologies adopt an occupation- or skills-based approach to identify the AI workforce, both of which can overestimate the size of the AI workforce as they blur together three different labor markets: (1) people building AI systems, (2) people adopting AI tools in other roles, and (3) workers whose tasks are exposed to AI-enabled change. In this report, we apply a more precise definition of the *AI development workforce* to assess the U.S. demand and supply of people who design, train, fine-tune, scale, and deploy AI systems.

We define AI development jobs as roles that require specialized knowledge, skills, and abilities and directly contribute to the technical development of AI systems.[1] Using this definition, we developed a machine learning (ML) classifier that identifies AI development roles in job postings data. We ran our classifier over a dataset of U.S. job postings from January 2010 to February 2026 to estimate the demand for AI development roles. We then estimated the size of the AI development workforce using a dataset of U.S. employee profiles. We found:

- Approximately **1.6 million AI development job postings** in the United States since 2010, including 331,445 postings in 2025.
- Approximately **519,000 AI development workers** in the United States as of March 2026.
- AI development roles are a small portion of the total U.S. workforce, accounting for less than 1% of both total labor demand and employment.
- AI development is concentrated in highly technical occupations, although among these occupations the proportion of roles that directly support AI development varies widely.

These estimates are substantially smaller than previous counts of the AI workforce because we disaggregated AI development roles from the broader workforce adopting AI tools in other roles and workers whose tasks are exposed to AI-enabled change. The technical workforce that develops AI systems is smaller and more specialized than previously understood. Our future publications and talent tracking tool, PATHWISE, will incorporate our AI development jobs definition and provide an empirical basis on which to develop workforce and education policy recommendations to bolster the AI workforce.

## Introduction

The AI workforce is foundational to crafting any comprehensive national AI strategy. Governments design policy to grow it, companies compete for it, and researchers track it to gauge national capacity. Given its importance to national competitiveness and effective workforce development, it is essential to clearly define and measure this population.

Existing efforts to measure the AI workforce adopt an occupation- or skills-based approach to identify the AI workforce. Although well-established and replicable, both approaches can overestimate the AI workforce, including an overly broad set of AI-related roles.[2] These approaches miss cutting-edge technical roles that do not fit into existing occupation or skills buckets. We developed a methodology to specifically identify the workforce driving the design, training, and deployment of AI systems. Our focus on workers involved in technical AI development provides a more precise AI workforce analysis and enables more targeted policy analyses and recommendations.

We define AI development jobs as roles that require specialized knowledge, skills, and abilities (KSAs) and that directly contribute to the technical development of AI systems—for example, systems built around ML models, including natural language processing (NLP), computer vision, and large language models (LLMs). Using this definition, we quantitatively identified the demand for, and current employment of, the technical AI workforce in the United States. In this report, we describe our methodology for identifying AI development jobs and the current employment of AI workers, and present initial findings. Future publications on this data will include analyses as well as targeted workforce and education policy recommendations for the AI workforce.

## U.S. Job Postings and Profiles Data

We used U.S. job posting and worker profile data provided by Lightcast.[*] Each job posting contains the full job description, including job duties and responsibilities. Each worker profile includes a worker's employment and education history and acquired skills. Lightcast enhances both datasets by mapping them to federal workforce

---

[*] We use Lightcast data as it is representative of the U.S. labor market demand, particularly for high-skill occupations (Cammeraat & Squicciarini 2021, Deming & Kahn 2018).

taxonomies such as the U.S. Bureau of Labor Statistics' Standard Occupational Codes (SOC) and Lightcast's in-house occupation and skills taxonomies.

## Measuring Demand for AI Talent

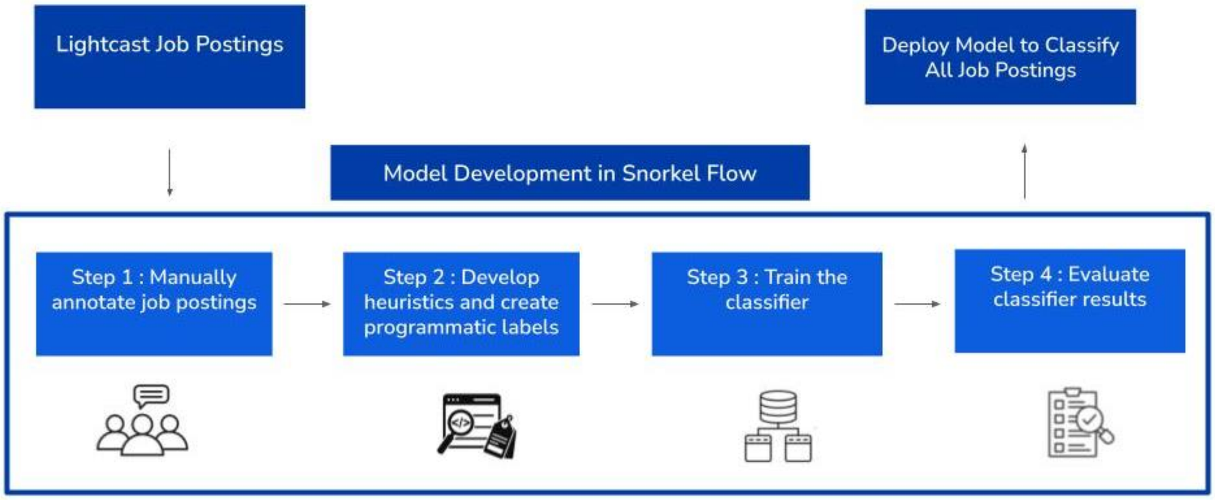
An important aspect in understanding the AI workforce is defining what constitutes an AI job. As outlined in recent CSET analysis, we define AI development jobs as roles that require specialized KSAs and that directly contribute to the technical development of AI systems.* Technical development work includes the development of models, AI-enabled applications, and the software scaffolds they are dependent on. The technical development of a model refers to the engineering work needed to design, train, fine-tune, scale, or deploy the model. Technical development of the software scaffold refers to the engineering work needed to design or code the software. Nontechnical roles, such as project managers, can directly contribute to these technical development tasks and therefore are included as relevant roles to AI development.

Using this definition, we developed a text-based classification model to identify AI development roles from the job postings dataset. First, we curated an expert-labeled set of job postings for training and validation. Because manual labeling at scale is time-consuming, we used Snorkel Flow, a model development platform that expedites labeling of training data using weak supervision.[3] We implemented the following steps, depicted in Figure 1, to a random subset of 56,000 U.S. job postings. This subset includes two samples: a random sample of all job postings, and a random sample of postings that mention either AI or ML in the job description. We included the second sample to ensure sufficient coverage of both AI-relevant and nonrelevant roles during annotation.

---

* Examples of specialized KSAs include PyTorch, TensorFlow, Keras, management of MLOps infrastructure, model benchmarking, etc.

**Figure 1.** Model Development Steps for AI Development Job Posting Classification




**Step 1: Manually annotate job postings.** We sought to identify roles in the AI workforce responsible for building AI and exclude roles that primarily use AI tools. Therefore, we used the job title and the posting's text, particularly the duties and responsibilities sections, for labeling, rather than relying solely on the listed job skills. First, subject-matter experts analyzed and manually labeled 1,023 job postings as AI development roles or not; 230 of the 1,023 were labeled as AI development roles. These postings served as a set of accurately labeled training data for the model. The decision matrix in Table 1 was used by experts to label job postings, particularly when resolving unique cases.

**Table 1.** Decision Matrix for Labeling AI Development Job Postings

| | | Does the role require technical AI KSAs? | | |
|---|---|---|---|---|
| | | Yes | Maybe | No |
| Does the role directly contribute to the technical development of AI systems?* | Yes | Included | Included | Included |
| | Maybe | Included | Excluded | Excluded |
| | No | Excluded | Excluded | Excluded |

* Note: An "AI system" includes both models and AI-enabled applications.

**Step 2: Develop heuristics and create programmatic labels.** Using our expert labeled job postings and our definition of AI development jobs, we developed heuristics to guide the next step in the labeling process. The heuristics were encoded as programmatic label functions using weak supervision and applied to the job title and job posting text. We used keyword- and prompt-based label functions to label our job postings.[*] See the text box below for examples of label functions. In total, we encoded over 60 label functions for weak labeling of our training data. Snorkel's generative model then aggregated the outputs of the label functions, and learned from the agreements and disagreements to provide an initial "weak" label for every posting in our training dataset as either an AI development or not an AI development role.

**Keyword-based label function:** A job posting with the phrase "build machine learning pipelines" is an AI-development role but a posting with a job title like "attorney" is not related to AI development.

**Prompt-based label function:** A business analyst posting with primary duties of assisting a data science team in developing an AI product is an AI development posting, but a business analyst posting with primary duties of building dashboards with conceptual knowledge of ML is not related to AI development.

**Step 3: Train the classifier.** Using our label functions, we labeled 50,000 job postings which are used as input to train the classifier. We trained a logistic regression model on Snorkel Flow, where the output was a Yes/No prediction of whether the posting was for a relevant role or for a role not related to AI development. We overcame class imbalance during model training as models in Snorkel Flow automatically adjust weights based on the class distribution in the hand annotated set of job postings.

**Step 4: Evaluate classifier results.** We evaluated the performance of our classification model on the remaining job postings. These served as the test data for our model.[†] Similar to the training process, we hand-labeled 108 job postings (45 of which were labeled as AI development jobs). We found that the classifier achieves a 91.7% F1 on the test data, indicating that our model is reliable at predicting both AI development

[*] We used the GPT-4o-mini and Gemini Flash via Vertex AI models.

[†] We used a small sample of job postings for testing, as the Snorkel test sets can be orders of magnitudes smaller than the training set.

and not-AI development postings. Table 2 reports the precision (ratio of correctly identified postings), recall (ratio of how many AI-relevant postings were labeled as so), and F1 (combined statistic for the precision and recall) for each of our two classes.

**Table 2.** AI Development Job Classification Model Evaluation

| Class | Number of Postings in Test Data | Precision | Recall | F1 |
|---|---|---|---|---|
| AI development | 63 | 86.0% | 95.6% | 90.5% |
| Not AI development | 45 | 96.6% | 88.9% | 92.6% |

Given strong performance, we deployed the model to classify all Lightcast job postings from January 2010 through February 2026. This was implemented by exporting the entire model application as an MLflow file from Snorkel Flow, which was then scaled to all job postings in Google Cloud Platform. The model deployment code is available in this GitHub repository.

Our method has limitations. One is that job postings can be misclassified, reflecting Type I (false positive) and Type II (false negative) errors. This mainly occurs as our underlying data is not perfect. Many of the postings lacked job titles, and had irrelevant html segments and website links in the text. We specifically analyzed postings that were incorrectly classified, and found that they were all distinct from one another. Therefore, we did not add any additional label functions during evaluation to avoid overfitting.

## Measuring AI Development Talent Employment

The demand for AI talent is only part of the story. We also wanted to estimate the size of the current AI development workforce. In line with previous work, we assume that the share of AI development job postings within each occupation is similar to the share of current AI employment in that occupation.[4] While this assumption has not been tested directly for AI roles, previous work provides evidence to support its use and a similar approach is used in related studies assessing workforce skills and tasks.

Accordingly, we calculated AI employment per occupation by multiplying the share of AI development job postings within a given occupation by total employment for that occupation. We then calculated total U.S. AI development employment by summing the number of AI employees across all occupations. We used Lightcast's occupation taxonomy, specifically the specialized occupation group, as it offers more granularity for technical roles than federal SOC classification codes. We note that our analysis does not include employees without a digital presence, as Lightcast primarily aggregates profiles from online sources.[*]

## Results: U.S. AI Development Workforce Demand

Using our classifier, we identified 1.6 million U.S. AI development job postings from January 2010 to February 2026. This represents approximately 0.3% of the full set of 484,600,000 U.S. job postings over that period. For 2025, we identified 331,445 AI development job postings, which represented approximately 0.8% of job postings that year. The demand for AI development workers represents a small share of overall labor demand.

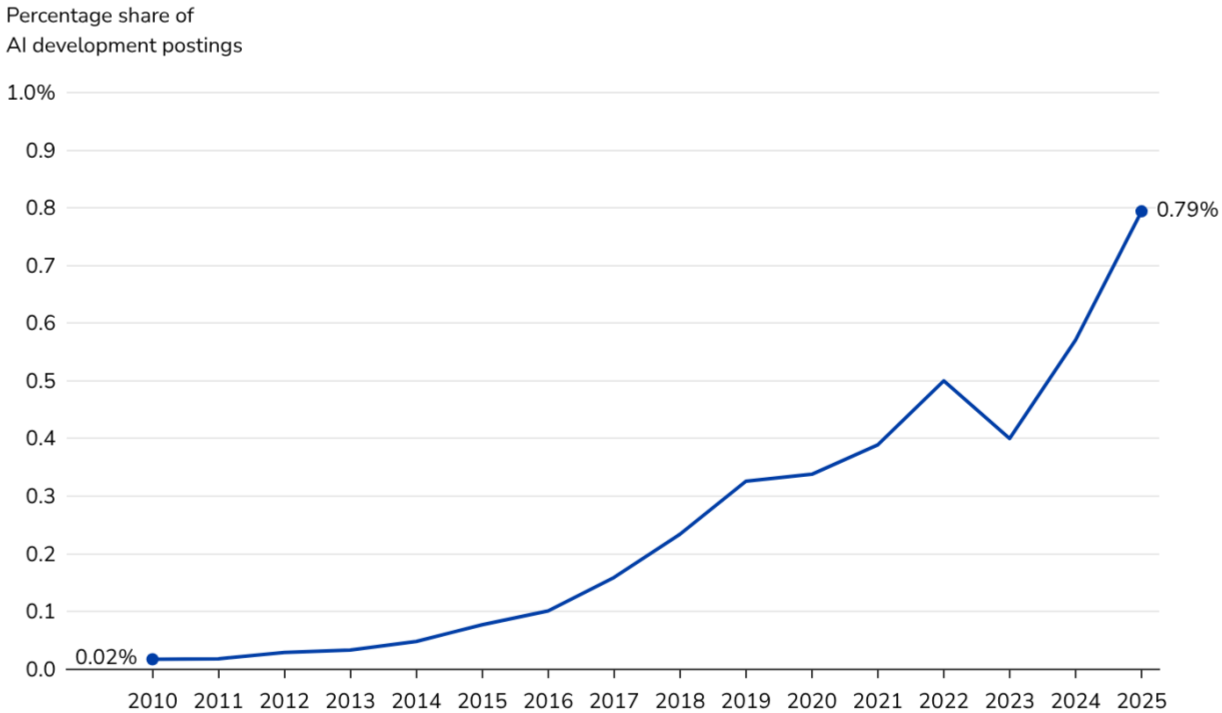
> *"We identified 1.6 million U.S. AI development job postings from January 2010 to February 2026 . . . For 2025, we identified 331,445 AI development job postings, which represented approximately 0.8% percent of job postings in that year."*

We compare this finding to estimates derived from CSET's prior method for measuring the AI workforce.[5] That approach took a broader definition of the AI workforce and identified AI job postings and roles solely on occupation. Applying that method to the same 2010-2026 set of job postings, we found 80 million AI job postings over the period. Our more granular definition and classifier-based approach provides a more precise estimate of AI development-specific role demand.

Since 2010, the share of AI development job postings has increased but remains below 1% of all U.S. job postings, as shown in Figure 2. Note, there was a decline in the total number and share of AI development job postings in 2023, which may relate to a larger temporary decline in demand for technical roles that year, but the share of AI development postings has increased since 2024.[6] This could be related to the increase in postings related to LLMs and requiring generative AI skills.[7]

[*] This source of error is less likely to be a significant factor in a highly technical and online field like AI.

**Figure 2:** Share of AI Development Roles in U.S. Job Postings from 2010 - 2025




Source: CSET analysis of Lightcast U.S. job postings data.

Table 3 displays the 10 most common occupations among the 1.6 million AI development postings. Data scientist, data engineer, software developer/engineer, and ML engineer surface as the most common occupations appearing in our AI job postings. Together, these four occupations comprise nearly 45% of all AI development job postings.

**Table 3:** Top 10 Occupations in AI Development Job Postings, January 2010 - February 2026

| Occupation | Share of AI Development Postings | Number of AI Development Postings |
|---|---|---|
| Data Scientist | 52% | 279,855 |
| Data Engineer | 47% | 168,396 |
| Software Developer / Engineer | 9% | 136,543 |
| Machine Learning Engineer | 94% | 133,124 |
| DevOps Engineer | 16% | 67,621 |
| Research Scientist | 36% | 66,288 |
| Artificial Intelligence Engineer (General) | 66% | 55,397 |
| Platform Engineer | 11% | 38,930 |
| Data Analyst | 10% | 38,672 |
| Natural Language Processing Engineer | 74% | 37,980 |

Source: CSET analysis of Lightcast U.S. job postings data.

We also examined which occupations have the highest share of AI development postings. Interestingly, this surfaces a slightly different list. Notably, while data scientist is the most common occupation in our set of AI development job postings, only 52% of data scientist postings in this period are specific to AI development, as shown in Table 3. Meanwhile, 94% of machine learning ML engineer job postings, 74% of NLP engineer job postings, and 66% of artificial intelligence engineer (general) postings are specific to AI development. We find that cases where postings for these occupations were not classified as AI development tend to be when primary duties relate to the software development life cycle, instead of AI development. For example, a machine learning engineer posting where the primary duties are building software applications with sensor data for geographic mapping or a generative artificial intelligence engineer posting focused on teaching AI development concepts to clients.

This underscores the importance of using the postings' duties and responsibilities along with job titles to accurately identify AI development roles.

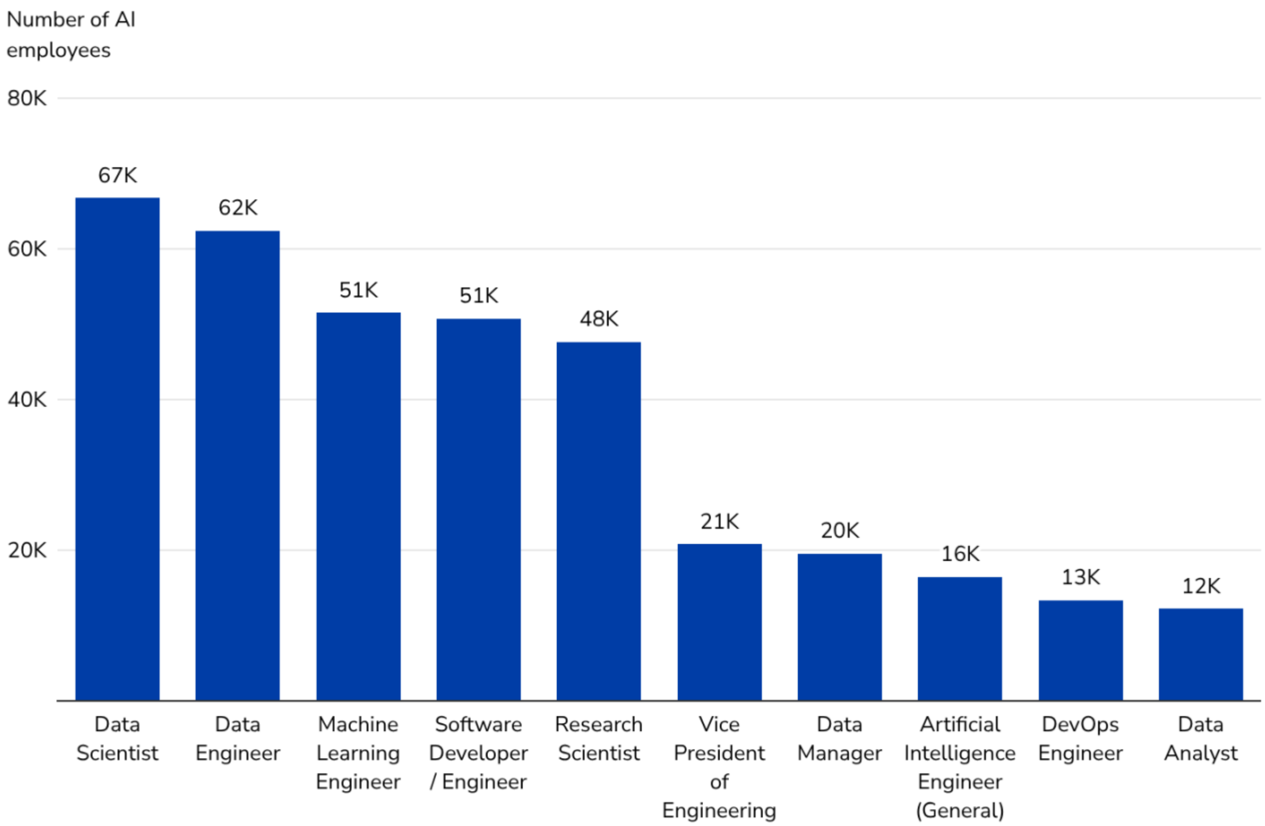
## Results: AI Development Employment

Using our employment methodology, we estimate that there are currently about 519,000 AI employees in the United States. This represents approximately 0.3% of total employed U.S. workforce, indicating that AI development workers constitute a small segment of the total workforce. If we apply CSET's previous definition of AI employment, we estimate nearly 28 million current AI employees in the United States. Like the difference in our new estimate of AI talent demand, this is a result of shifting from an inclusive definition of AI related roles to a narrower focus on employees primarily responsible for AI development.

*"We estimate that there are currently about 519,000 AI employees in the United States. This represents approximately 0.3% of total employed U.S. workforce."*

Using this definition, we found that data scientist and data engineer are the most common occupations among current AI employees, as shown in Figure 3. These are also the most common occupations within our AI development job postings, as seen in Table 3. One notable difference we find, compared to our AI development demand estimates, is that managerial roles such as vice president of engineering and data manager account for a large portion of the current AI development workforce. Future analysis will explore the extent to which this suggests gaps between the current workforce and the desired workforce.

**Figure 3:** Top 10 Occupations among Current U.S. AI Development Employees, March 2026




Source: CSET analysis of Lightcast U.S profiles data.

## Conclusions

The AI workforce is a key component of U.S. AI development and competitiveness. In this report, we present a new method for measuring this workforce. Defining AI development jobs as roles that require specialized KSAs and that directly contribute to the technical development of AI systems, we trained and deployed a machine learning classifier that applies our definition to job postings data.

We found about 1.6 million AI development job postings in the United States between January 2010 - February 2026. We also estimated that there are 519,000 AI development employees as of March 2026. Based on these estimates, AI development talent demand and employment account for less than 1% of the total U.S. labor market, indicating that technical AI development occurs within a niche segment of the workforce. In forthcoming publications, we will further explore the U.S. AI workforce, including analyses on skills and training programs that are required. Through these results, we aim to provide targeted workforce and education policy recommendations. We enable exploration of this data through PATHWISE, our interactive online tool that tracks emerging technology talent in the United States.

## Authors

**Sonali Subbu Rathinam** is a data research analyst at CSET, where her research focuses on the emerging technologies workforce and talent.

**Ronnie Kinoshita** is the deputy director of data science and research at CSET. She works across the organization to align data team projects with analytic needs and organization strategy.

## Acknowledgements

The authors would like to thank Catherine Aiken, Matthias Oschinski, James Dunham, Rebecca Gelles, Jack Karsten, Jacob Feldgoise, Luke Koslosky, Kyle Crichton, Carlo Pizzinelli, Zack Mabel, Sam Manning, Lauren Lassiter and Shelton Fitch for their feedback and assistance.



Document Identifier: doi: 10.51593/20240037

## Endnotes


[1] Luke Koslosky, "Defining the AI Workforce: A New CSET Approach," Center for Security and Emerging Technology, April 24, 2026, https://cset.georgetown.edu/article/defining-the-ai-workforce/.

[2] Diana Gehlhaus and Santiago Mutis, "The U.S. AI Workforce: Understanding the Supply of AI Talent," Center for Security and Emerging Technology, January 2021, https://doi.org/10.51593/20200068.

[3] Alexander Ratner, Stephen Bach, Henry Ehrenberg, et al., "Snorkel: Rapid Training Data Creation with Weak Supervision," *Proceedings of the VLDB Endowment* 11, no. 3 (November 2017): 269–82. https://dl.acm.org/doi/10.14778/3157794.3157797.

[4] Andrew Green and Lucas Lamby, "The Characteristics of the Artificial Intelligence Workforce across OECD Countries," *The Indian Journal of Labour Economics* 68 (February 2025): 541-568, https://link.springer.com/article/10.1007/s41027-024-00549-7; Liudmila Alekseeva, Jośe Azar, Mireia Giné, Sampsa Samila, & Bledi Taska (2021), "The demand for AI skills in the labor market," *Labour Economics* 71, https://doi.org/10.1016/j.labeco.2021.102002; David Autor, Frank Levy, & Richard Murnane, "The skill content of recent technological change: An empirical exploration," *The Quarterly Journal of Economics* 118, no. 4 (2003): 1279–1333, https://doi.org/10.1162/003355303322552801; David Deming, & Lisa Kahn, "Skill requirements across firms and labor markets: Evidence from job postings for professionals," *Journal of Labor Economics* 36, no. S1 (2018): S337–S369, https://doi.org/10.1086/694106; Alexandre Georgieff, & Raphaela Hyee, "Artificial intelligence and employment: New cross-country evidence," *Frontiers in Artificial Intelligence* 5 (2022), https://doi.org/10.3389/frai.2022.832736; Brad Hershbein, & Lisa Kahn, "Do recessions accelerate routine-biased technological change? Evidence from vacancy postings," *American Economic Review* 108, no. 7: 1737–1772, https://doi.org/10.1257/aer.20161570.

[5] Diana Gehlhaus and Santiago Mutis, "The U.S. AI Workforce: Understanding the Supply of AI Talent" Center for Security and Emerging Technology, January 2021, https://doi.org/10.51593/20200068.

[6] N. Maslej et al., The AI Index 2024 Annual Report: AI Index Steering Committee, Institute for Human-Centered AI, Stanford University, Stanford, CA, USA, Apr. 2024, https://arxiv.org/abs/2405.19522

[7] Callam Pickering, "Rapid Growth in GenAI Job Postings: Expectations and Surprising Trends" (Indeed Hiring Lab, November 21, 2024), https://www.hiringlab.org/2024/11/21/growth-in-ai-job-postings-trends-and-surprises/; Alex Johnston and David Immerman, "2024 Global Trends in AI," WEKA, 2024, https://www.weka.io/resources/analyst-report/2024-global-trends-in-ai/?fs=934acb04625235a675b7824ac503ce42&sg=38080.