THE ECONOMIC IMPACT OF THE DATA ANNOTATION INDUSTRY

DECEMBER 2025

OXFORD ECONOMICS

scale

TABLE OF CONTENTS

EXECUTIVE SUMMARY

Data is a critical foundation of the artificial intelligence (AI) industry. High-quality datasets are fundamental to the leading AI models that are reshaping economies and societies. As global demand for AI increases, the data annotation industry has emerged as a cornerstone of the AI value chain and a key enabler of AI growth, but it is also fueling economic activity and supporting new economic opportunities across the United States.

$5.7 bn
Total contribution to US GDP in 2024

$1.2 bn
Tax contribution at the federal, state, and local level in 2024

200,000
earning opportunities in the US in 2024

THE DATA ANNOTATION INDUSTRY IN THE US IS A RAPIDLY GROWING ECONOMIC FORCE

The US market size of the data annotation is estimated to be between $2.7 billion and $5.0 billion in 2024. By 2030, the market is expected to grow to between $10.3 billion and $19.0 billion[1].

We estimate that the data annotation industry's total contribution to the US GDP was $5.7 billion in 2024, which is comparable to established sectors like computer storage device manufacturing ($4.2 billion) and computer terminals equipment manufacturing ($6.5 billion).

$2.1 billion was contributed through the direct channel (this includes the industry's operations). The remaining $3.6 billion was contributed through the indirect and induced channels, which consist of spending through its suppliers and the worker earnings in the wider economy. This economic activity contributed $1.2 billion in tax revenue at the federal, state, and local levels.

The data annotation industry directly supported earning opportunities[2] for nearly 200,000 people and an additional 9,000 full-time jobs. The industry also supported an equivalent of around 25,000 full-time jobs across the indirect and induced channels.[3]

Looking forward, we forecast that the industry's total economic contribution will more than triple to $19.2 billion by 2030,[4] consistent with average annual growth of 25%.

[1] Represented in nominal terms.
[2] This is the total number of earning opportunities directly supported by the data annotation industry. We adjust earning opportunities for multi-homing, that is, contractors working across multiple platforms.
[3] Or 28,000 jobs on a headcount basis.
[4] The economic contribution number is represented in 2024 real dollars.

UNDERSTANDING THE DATA ANNOTATION WORKFORCE

Despite the AI industry's growing importance, relatively little is known about the people who underpin it—the data annotators who create and validate the data inputs that shape AI models. This lack of visibility can lead to misunderstandings about how the industry operates and the individuals who make this work possible.

To address this evidence gap, we conducted a survey to understand the profile, motivations, and aspirations of data annotators in the US.

$19.2 bn
Total contribution to US GDP in 2030

The data annotation workforce is highly educated, relatively young, and often manages personal responsibilities alongside work.

- The vast majority (84%) of respondents held at least a bachelor's degree, which is double the share in the wider US workforce.
- Of the respondents, 65% were under the age of forty-four, which compares to 57% in the general US workforce.
- 94% of respondents are engaged in other professional and academic activities. Annotators also managed family and domestic responsibilities full-time (10%), and others combined their annotator work with ongoing education (7%) or job hunting (16%). Data annotation roles enable these individuals to access earning opportunities on a flexible schedule and leverage their expertise to earn additional income.

Roles in data annotation provide flexible income opportunities that support personal, professional, and academic priorities.

- Income was a key motivation for annotators—94% reported that income drew them to the role. Data annotators used their earnings to strengthen their financial position, cover the rising cost of living, reach financial milestones, and pay off debt.
- Flexibility was an equally important motivating factor for 91% of the respondents. Respondents valued the ability to fit work around their lives, whether that was to balance professional and academic commitments, be more available for their families, or dedicate time for personal enrichment.
- Different groups emphasized certain aspects of flexibility to a larger extent: students for example valued academic and work life balance, female annotators and caregivers focused on being available for their families, while self-employed annotators and those self-reporting that they do other "freelance" work prized time for personal enrichment.

91%
Of respondents cited flexibility as being a motivating factor for working as data annotators

Data annotation is also a pathway to longer-term professional development.

- Many annotators viewed their role as a stepping-stone to future professional growth. Nearly three in four respondents indicated that they hoped to apply the skills acquired through their annotator work to future roles in AI and technology.

- While the most prominent aspiration among annotators was to transition into AI- or tech-related roles (as reported by 50% of respondents), annotators also hoped to seek jobs in other industries (20% of respondents) and expressed ambitions to pursue entrepreneurial ventures (17% of respondents).

EMPOWERING DATA ANNOTATORS AND DRIVING THE FUTURE OF AI

Our survey highlights that data annotation offers a flexible source of income, while also accommodating the diverse professional, personal and family commitments, and aspirations of annotators. In doing so, the data annotation industry fuels current AI development, while helping to empower the human annotators that are so crucial for the ecosystem’s growth.

Understanding the needs of this workforce—particularly the role of flexibility—is essential to sustaining the growth of the AI economy, and ensuring the supply of talent remains widely available. By recognizing the economic and social value created by the data annotation industry, policymakers and industry leaders can support the AI ecosystem and help unleash its full potential.

1. INTRODUCTION

Data annotation represents a critical yet frequently underappreciated component of artificial intelligence (AI) model development, encompassing not only systematic labeling and structuring of raw data but also prompt engineering and refinement that help machine learning systems perform effectively. While the technical sophistication of AI models and the imposing physical footprint of data centers often dominate public attention, the foundational role of high-quality, human-curated annotated data in training and evaluating these systems remains largely unseen.

The data annotation industry has emerged as an economic force within the US AI ecosystem, with market valuations of key players suggesting substantial growth trajectories and economic impact. However, thus far, the economic and technological contribution of the data annotation industry has received relatively limited attention. This includes the industry's workforce, characterized by flexible employment arrangements, often distributed globally, spanning from highly skilled domain experts to entry-level annotators working through freelance platforms.

Understanding the economic dynamics and workforce characteristics of this rapidly evolving industry becomes increasingly important as AI adoption accelerates across the US economy. The data annotation industry operates through diverse business models, creating new income opportunities and pathways for upskilling through flexible work arrangements. Furthermore, the industry's role in determining AI system performance and ensuring safety makes it a critical consideration for AI governance and ethical deployment.

This study aims to enhance the existing body of research into the economic structure and workforce dynamics around the data annotation industry, aiming to empower stakeholders with valuable insights to help support enduring AI development.

The rest of the report is structured as follows:

- Chapter 2 introduces the role of data annotation as a critical enabler of AI technologies and describes the key features of the industry.
- Chapter 3 quantifies the economic impact of the data annotation industry to the US economy and forecasts how it might grow to the end of the decade.
- Chapter 4 explores the profiles, motivations, and aspirations of the annotators that support the data annotation industry.
- Chapter 5 offers concluding remarks.
- The methodological appendices offer more detail on our research inputs and the economic impact modeling.

2. HOW DATA ANNOTATION HELPS POWER TODAY'S AI SYSTEMS

2.1 THE ROLE OF DATA IN AI PRODUCTION

AI is already having a significant impact on the global economy, which is expected to increase as the technology matures and is widely adopted. Economic benefits may arise from applications that increase productivity through process automation and improved decision making, innovation through the creation of new products or services, and/or entirely new organizational structures enabled by these technologies.

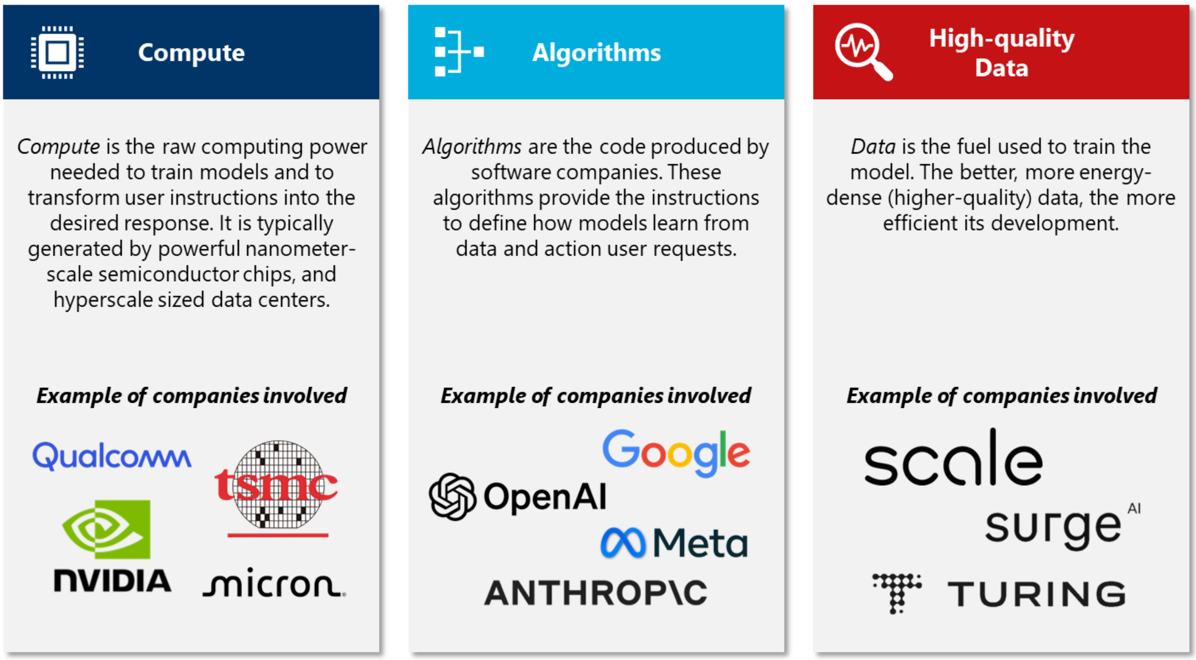
For example, the enormous investment in Generative AI (GenAI) models over the last three years reflects their transformational potential. Indicative of the uncertainty associated with any new technology, estimates around the potential implications for this technology vary widely. However, even in low business adoption cases, the annual productivity boost in the US could be as much as 1.7% by 2032, rising to as much as 3.5% in a high adoption scenario.[5] This corresponds to an uplift in US GDP of $477 billion to $1 trillion within 10 years.[6] Beyond the US, AI is expected to grow global GDP by $7 trillion—or 7%—over the next 10 years.[7]

As AI's economic potential becomes clearer, countries are increasingly competing for global AI leadership—recognizing that superiority in AI will depend not only on innovation and investment, but also on access to large-scale datasets, computing power, and talent. Supporting AI growth would require building a strong AI data industry, and funding for upskilling programs to adapt the workforce.

AI models, including GenAI models and their agentic offspring, depend on a set of critical enablers. We can think of building an AI model, agent or application as relying on three primary inputs: compute, algorithms, and data (illustrated in Fig. 1).

[5] Cognizant, "New work, new world", 2023.
[6] Cognizant, "New work, new world", 2023.
[7] MIT Institute, "A new look at the economics of AI", 2025.

Fig. 1. Key inputs to building an AI model



Each of these three pillars is instrumental in advancing AI models to transform industries, reshape work processes, and enhance efficiencies. The importance of each input grows as AI systems continue to evolve, from simple conversational tools toward more advanced agentic applications capable of performing more complex tasks. Enterprises and governments are also increasingly adopting AI systems that draw on internal data and expertise, enabling more effective business operations and strengthening capabilities. To sustain this momentum, the AI industry will also need to adopt more rigorous methods for evaluating models. Data annotation to train AI models thus plays a critical role in advancing AI.

While the size of the data for training has grown remarkably over the last two decades, [8] **the quality of the data remains an important constraint to model performance.** [9] Major players have scraped and indexed vast portions of publicly available text data, which means that making any more progress via increased quantity is challenging.[10, 11, 12] The point where it is no longer possible to simply add more publicly available data to significantly improve AI systems is also referred to as the "data wall".[13]

[8] Maslej, N., et al., "The AI Index 2025 Annual Report", *AI Index Steering Committee, Institute for Human-Centered AI, Stanford University*, 2025.
[9] George D. and Alexandr W., "Unlocking AI's Future: Alexandr Wang on the Power of Frontier Data", *Andreessen Horowitz.* September 24, 2024.
[10] Aschenbrenner, L, "From GPT-4 to AGI", *Situational Awareness,* 2024.
[11] Villalobos, P., Anson, H. et al., "Will We Run Out of Data? Limits of LLM Scaling Based on Human-Generated Data", *Epoch AI,* 2024.
[12] Wiggers, K., "AI training data has a price tag that only Big Tech can afford." *TechCrunch*, June 1, 2024.
[13] SuperAnnotate, "*AI data wall: Why experts predict AI slowdown and how to break through the plateau*", 2024.

There is a need for more "frontier" data—cutting-edge, novel data sources and advanced annotation techniques—to overcome the data wall and improve data quality. While non-public and enterprise data may present an additional source of "frontier" data, copyright requirements, privacy safeguards, and business considerations, restrict its wider application.

2.2 THE IMPORTANCE OF HUMANS IN CREATING GOOD QUALITY DATA

The reduced ability to expand to new text data sources underscores the important role of humans both as creators of novel information, but also in terms of building better quality datasets, with richer, more accurately labeled data. By embedding context, describing structure, and making the sources that are available richer, data annotators can enhance model performance even where available text has been exhausted. This dimension of progress may be even more important in the context of non-English language models, where there are greater constraints on digitized data (though at the same time, greater performance may be achieved simply by digitizing much more human audio and text data).

While the creation of AI-generated (synthetic) data, combined with the use of Machine Learning (ML) to automate data annotation, is a potential non-human route for expanded data collection, its potential to achieve performance improvements remains questionable. Machines remain limited in their ability to perform certain complex tasks or to understand abstract concepts like sarcasm.[14] On performance tests such as Humanity's Last Exam (HLE),[15] even the most advanced AI systems achieve scores of only about 37%, with many scoring under 15%.[16, 17] Although AI systems' performance in such tests has improved rapidly, with further improvement evident through experimental agents, current evidence suggests that AI models still have substantial room for improvement.

In addition, an important role for humans in the development of these AI models is attention to concerns around human values, such as those around safety. Data annotators must frequently exercise judgement, especially in situations that are highly context specific. Annotators need to occasionally apply their judgment in assessing safety, which may manifest differently across heterogenous geographies and communities. In democratic societies, as these models become embedded in everyday processes at home, at work, in government offices, and in the military, it is hard to imagine that humans will not remain firmly in-the-loop.[18]

---

[14] Stanford University, "*The 2025 AI Index Report*", 2025.

[15] Yue, S., Wang, A., Hendrycks, D., Khoja, A., Ren, R., Hausenloy, J., Zhang, O., Mazeika, M., Hu, J., Zhang, H., Zhang, C. B. C., Shaaban, M., Phan, L., Gatti, A., Han, Z., & Li, N., "*Humanity's Last Exam*", *Center for AI Safety*, 2024.

[16] OrionAI, "*Gemini 3: Benchmarks Guide: Performance Across HLE, GPQA, and More*," 2025.

[17] Taheri, O., et al., "*Humanity's Last Exam: A Multi-Modal Benchmark at the Frontier of Human Knowledge*", 2025.

[18] Zhao, D., Scheuerman, M. K., Chitre, P., Andrews, J. T. A., Panagiotidou, G., Walker, S., Pine, K. H., & Xiang, A., "A Taxonomy of Challenges to Curating Fair Datasets", *Advances in Neural Information Processing Systems 37*, 2024.

DATA ANNOTATION IN THE AI MODEL PIPELINE

Data annotation is used in **data preprocessing**, including text, image and video, audio, and sensor data. Data preprocessing involves the transformation of unstructured and "noisy" raw data into a curated and annotated format. The cleaned and labeled dataset is then used to train leading-edge models. Data annotators manually contextualize data such as:[19]

1. Text: serves as a proxy for context (labeling aspects such as sentiment and intent in the training text) which helps large language models (LLMs) to interpret human language and predict appropriate responses.

2. Image and video: labeling and tracking objects or pixels within images and videos helps computer vision models recognize and classify visual elements.

3. Audio: involves transcribing speech and tagging sounds (e.g., identifying a siren, a laugh, etc.).

4. Sensor and time series: identifying and marking data patterns and anomalies (such as heart rate spikes) enables AI models to detect unusual or outlier events across a variety of settings.

Data annotation is also used in **model evaluation** to fine tune model outputs. The test dataset (using held-out labeled data) is used to measure the performance of the model.

HUMAN PARTICIPATION IN AI DEVELOPMENT EXTENDS BEYOND PREPARING TRAINING DATASETS.

Reinforcement learning from human feedback (RLHF) is a technique used to fine-tune AI and machine learning models.[20] In the context of GenAI model development, data annotators correct and evaluate competing model responses, checking not just for accuracy but also tone, ethical norms, and harm.[21] Supervised learning with human feedback is credited with significant improvements in model performance, with one OpenAI researcher remarking that higher-quality annotations were the "main source of improvements in DALL-E 3", the company's text-to-image model.[22]

As noted, the training of AI models has expanded beyond processing and supervising language to multimodal formats with more potential commercial applications. In pursuit of more sophisticated models that can perform as agents of humans, AI training also involves capturing complex human behavior. For instance, recording human workflows and complex actions like debugging code or switching between apps can augment the problem-solving capabilities of AI models, and train AI agents on how to best organize and plan tasks before starting them.

[19] Kornilov, A., "*Data Annotation Types*", *TrainingData.pro*, (n.d.).
[20] Arize AI, "*OpenAI on RLHF: The Secret Sauce for Aligned LLMs*", May 5, 2023.
[21] Sama., "Machines Still Need Us", (n.d.).
[22] Wiggers, K., "AI training data has a price tag that only Big Tech can afford", *TechCrunch*, June 1, 2024.

2.3 THE DATA ANNOTATION INDUSTRY

As discussed, the data annotation industry is a crucial component of the broader AI and ML industry, involving data preprocessing activities such as categorizing, tagging, and providing context for raw data (such as text, images, video and audio) to train and finetune AI models. Beyond improving model performance, data annotation supports the AI and ML industry by reducing in-house workload for businesses, providing access to advanced annotation tools, and allowing enterprises to scale quickly.[23]

Broadly speaking, we can segment the data annotation industry into two submarkets, though companies might operate in both:

1) **data annotation for GenAI**, which supports the training and evaluation of LLMs and multimodal systems;[24] and

2) **broader data annotation market**, which encompasses a wide range of annotation tasks for applications such as autonomous vehicles (AV), healthcare, robotics, and other machine learning and automation systems.

Human annotators account for the dominant share of the market and are generally more reliable as they can adapt better to niche project requirements and complex datasets. This is especially the case in industries where precision is non-negotiable, such as medical imaging and autonomous vehicles. The remaining share is captured by semi-supervised or automated annotation.

2.3.1 Market composition

The data annotation marketplace is highly competitive and features a diverse mix of players. Many firms offer traditional data annotation services for generic datasets, often leveraging their workforce to manage high volume, standardized tasks. At the same time, a selection of specialized, high-end data annotators, like Scale AI, Surge AI, Appen, Labelbox, and CogitoTech among others, offer bespoke services and are integrated deeply within frontier AI labs and developers. These players offer high-quality data annotation for advanced AI applications.

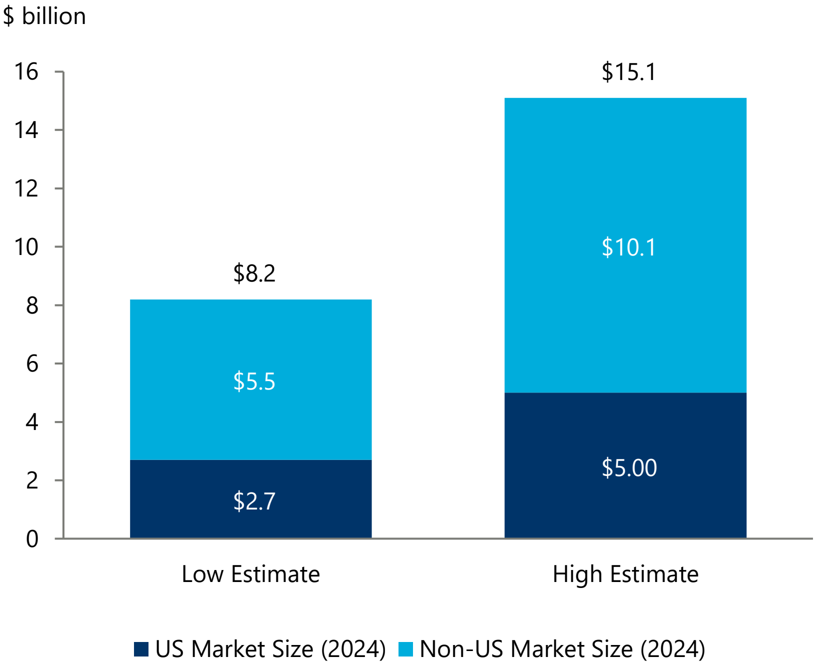
Many of these firms offer end-to-end services, including data collection, generation, curation, annotation, model refinement, and testing. For instance, they might combine machine intelligence with humans-in-the-loop into a semi-automated annotation model that optimizes for accuracy and efficiency.[25] They also offer RLHF services to support continued model refinement and AI red teaming, which involves simulating attacks on AI systems to identify and mitigate vulnerabilities.[26] Furthermore, these data annotation companies support a workforce that often consists of subject matter experts for domain-specific annotation. For instance, the diversity of AI use cases has led to the emergence of niche providers like iMerit for medical imaging and Sapien for legal data annotation.

[23] Thakur, R., "How to choose the right data labeling company in 2025", *Labellerr.* 2025.
[24] There are a limited set of primary customers such as Anthropic, Google, Meta, Microsoft, OpenAI, Stability AI and xAI.
[25] Lasky, J., "Human-in-the-loop (HITL)", *EBSCO,* 2025.
[26] IBM, "What is red teaming for generative AI?", 2024.

2.3.2 Current and projected market size

We have estimated the global market for data annotation and related services to be between $8.2 to $15.1 billion in terms of revenue, with **$2.7 billion to $5.0 billion of that market captured by the US in 2024** (see Appendix 1).

Fig. 2. Market size of data annotation industry, 2024



Source: Oxford Economics

This relatively large share reflects the dominant position of the US in AI research, investment, and infrastructure. In 2024, US private investment in AI hit $109 billion, with China and the UK trailing at $9.3 billion and $4.5 billion, respectively.[27] In the same year, US-based institutions released 40 notable AI models, compared to 15 from China and three from Europe.[28]

The industry is expected to experience rapid growth in the coming decade, with some sources forecasting a compound annual growth rate of 25%.[29] In particular, the North America market is expected to be worth $12 billion by 2032, according to at least one market research firm.[30] Based on these estimates, we project that the US market will grow to between **$10.3 billion and $19.0 billion by 2030 in nominal terms** (see Appendix 1).

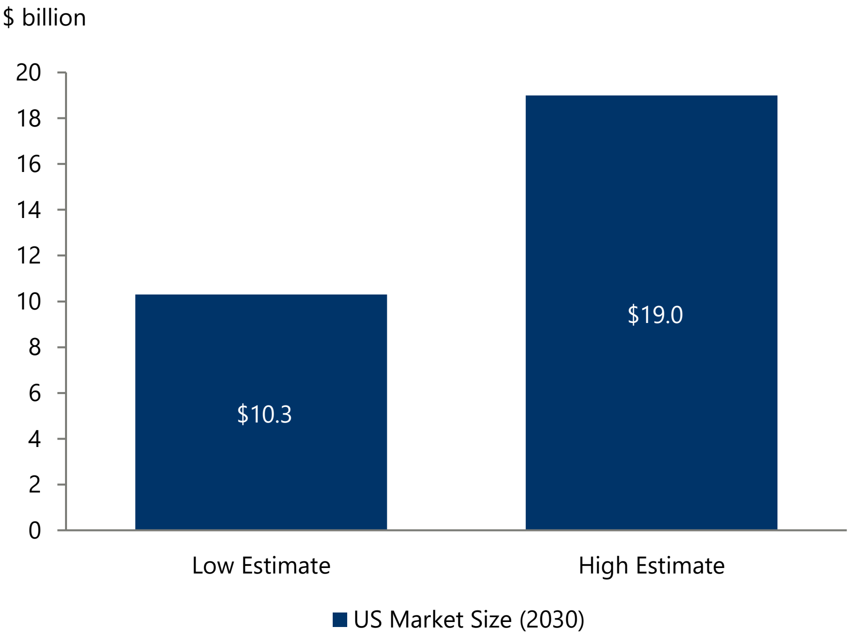
[27] We note, however, that these data leave out China's most significant source of AI investment, government venture capital funds.

[28] Maslej, N., Gil, Y. & Perrault, R., "Artificial Intelligence Index Report 2025: Trends and Insights.", *Stanford Institute for Human-Centered Artificial Intelligence*, 2025.

[29] Global Market Insights, "Data Annotation Tools Market Revenue", *AWS Marketplace*, 2022.

[30] Global Market Insights, "Data Annotation Tools Market Revenue", *AWS Marketplace*, 2022.

Fig. 3. US market size of data annotation industry, 2030[31]



Source: Oxford Economics

Much of this growth will be driven by the increase in the use of AI tools, which is a key component of demand for data annotation services. The information technology sector continues to account for the highest revenue share in the market (approximately 30%), with widespread usage of data annotation services in the development of cloud-based AI applications. For instance, over 5,000 AI-focused startups launched in 2023 used data annotation to develop their models.[24]

Increasing demand for image and video data annotation, particularly in the healthcare, autonomous driving, and entertainment sectors, also informs growth projections. In healthcare, annotated medical images like X-rays, MRIs, and CT scans are essential to train AI models to detect anomalies like tumors, fractures, or infections.

In the automotive sector, annotated data is key to guiding autonomous vehicles that navigate complex, dynamic environments like traffic conditions and weather. Furthermore, the acceleration in video content creation has made the entertainment sector a top use case for AI adoption.[32]

In summary, available evidence suggests increasing demand for data annotation services, as the number of use cases for AI expands (including in areas such as business enterprise and government services). The availability of data, augmented with human intelligence, could become a key differentiator to the advanced deployment of AI models, which makes supporting the data annotation industry ever more important. [33]

---

[31] Market revenues are presented in nominal dollar values.

[32] YouTube, "YouTube sees on average 20 million videos uploaded to the platform every day", YouTube for Press, (n.d.).

[33] Albergotti, R., "Meta's $15 billion investment in Scale AI comes with a hidden perk: data," *Semafor*, June 10, 2025.

2.4 DATA ANNOTATION AS A PATHWAY TO AI UPSKILLING

Data annotation also provides a possible pathway for workers to transition into the rapidly evolving AI economy. As technological advances are expected to force many workers to change careers by 2030, data annotation may offer flexible job opportunities while providing workers the opportunity to better understand the feedback loops underpinning the development of AI products.[34] Data annotators help to produce the "ground truth" data, while prompt engineers help shape model behavior. The skills developed through this work—attention to detail, contextual reasoning, and applying ethical judgment—and the understanding of AI production, can potentially help workers transition into other roles in the AI sector.[35] These include emerging occupations such as:

- **Prompt engineers**, who design and refine text-based prompts to guide large language models and optimize outputs for specific applications. [36]
- **AI compliance and ethics officers**, who ensure that AI systems are developed and deployed responsibly and in accordance with laws and ethical standards, addressing challenges such as bias, privacy, and transparency.
- **Annotation policy developers and analysts**, who establish the frameworks and guidelines that ensure consistency and quality in annotation processes, particularly in natural language processing and policy-related domains.

By equipping workers with the skills to transition into an economy shaped by AI, data annotation not only strengthens AI development but may also support broader workforce economic resilience in the face of large-scale technological disruptions.

[34] Illanes, R., et al., "Retraining and reskilling workers in the age of automation", *McKinsey Global Institute*, 2018.
[35] In the US in 2024 there were approximately 750,000 postings in 2024 that required AI skills, including 66,000 that mentioned generative AI, 20,000 mentioning LLMs, and 6,300 citing prompt engineering. See PWC, "The Fearless Future: 2025 Global AI Jobs Barometer - US Analysis", Lightcast and HAI, The Stanford AI Index Report, WEF, The Future of Jobs Report 2025
[36] Kelly, J. "The Hot, New High-Paying Career Is An AI Prompt Engineer.", *Forbes*, 2024.

3. THE ECONOMIC IMPACT OF THE DATA ANNOTATION INDUSTRY IN THE US

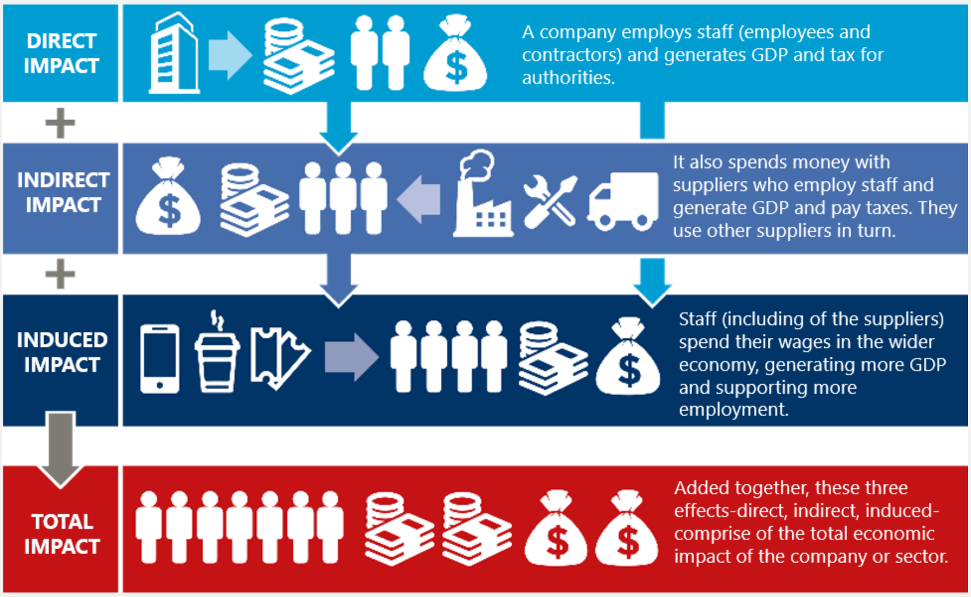
Data annotation companies have a substantial economic footprint both within and outside the US. In this section, we report our estimates of their direct, indirect, and induced economic impacts in the US in 2024, as well as our forecasts for 2030.

METHODOLOGICAL INTRODUCTION: ECONOMIC IMPACT ANALYSIS

Our economic model was constructed using the IMPLAN software[37], an industry-standard tool used for input-output modeling. Input-output models capture the flow of value throughout the economy, following the inter-industry linkages of supply chains. Results reflect contributions associated with data annotation firms, their supply chains, and employee wage spending. All results are presented in 2024 dollars.

In calculating our results, we captured the following three "channels" of economic activity: (1) the direct impacts of operational expenditure by data annotation companies, (2) the indirect supply chain impacts stimulated by the procurement spending of data annotation companies, and (3) induced impacts associated with annotators and workers employed along the supply chain who spend their wages in the US economy (Fig. 4).

Fig. 4. Economic impact by channel



[37] IMPLAN is an industry standard software for modelling supply chain linkages within the US economy. See https://implan.com/

These impacts are quantified across the following three metrics:

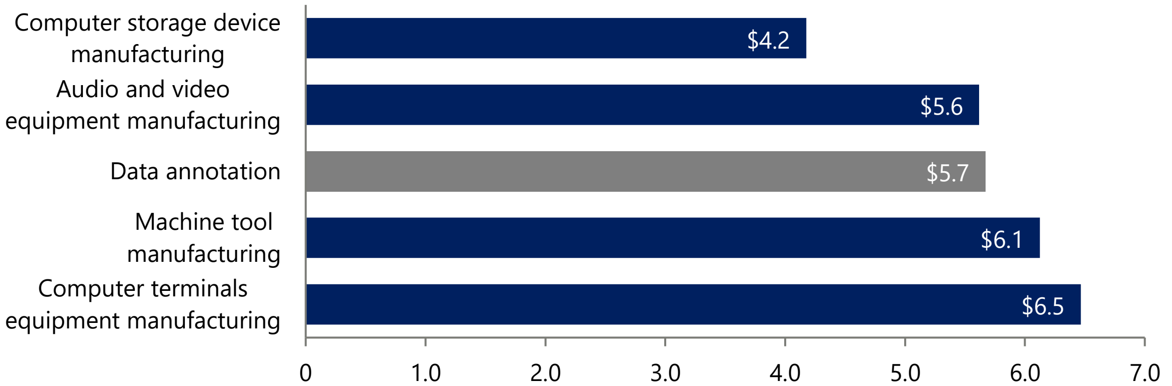
- **Gross Domestic Product (GDP).** This measures the economic value added of each entity in the supply chain and thus reflects the sum of economic activity occurring as a result of the activity being modeled.
- **Earning opportunities**. This is the earning opportunities generated by the data annotation industry in 2024 on a headcount basis, using data from Scale AI combined with our estimates for the 2024 US data annotation industry market size.
- **Tax revenues.** This includes taxes from the economic activity at the local, state, and federal levels.

Additional details of the data, assumptions, and estimates that underpin the economic impact model can be found in Appendix 1.

3.1 THE ECONOMIC FOOTPRINT OF THE DATA ANNOTATION INDUSTRY

Combining all three channels of impact—direct, indirect, and induced—we estimate that the total impact of the data annotation industry on the US economy amounted to approximately **$5.7 billion** in 2024. Its economic footprint is comparable to the total value added of established industries like computer storage device manufacturing ($4.2 billion), audio and video equipment manufacturing ($5.6 billion), machine tool manufacturing ($6.1 billion), and computer terminals equipment manufacturing ($6.5 billion) (See Fig. 5)

Fig. 5. GDP contribution of data annotation industry compared to other industries[38]



Source: Oxford Economics, IMPLAN

Of the $5.7 billion total GDP contribution, **$2.1 billion** was through the direct channel. This reflects the value-added generated by the data annotation industry, including wages paid, its capital income, and taxes paid to the government. An additional **$1.6 billion** was generated through the indirect channel, which represents the data annotation industry's supply chain. The remaining **$2.0 billion** was generated through the induced channel, which includes wages spent by employees and freelance workers in the data annotation industry and those in its supply chain.

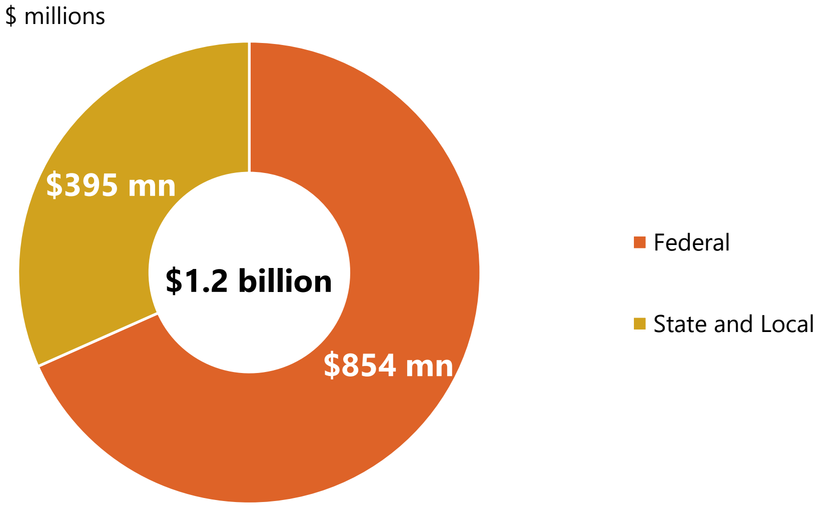
[38] The GDP comparators are sourced from IMPLAN and adjusted to 2024-dollar years.

The GDP multiplier of the data annotation industry is 2.7. This means that for every $100 spent by the industry an additional $170 is generated in economic activity elsewhere in the US economy.

We estimate that the data annotation industry directly supported more than 200,000 earning opportunities[39] and an additional 9,000 full-time jobs in the US. Beyond these direct roles, the industry also sustained an additional 25,000 full-time jobs through supply-chain purchases (indirect effects) and workers' spending in local communities (induced effects).[40]

The economic activity of data annotation companies contributed to tax revenues at the federal, state, and local levels. These include income and payroll taxes, corporate income taxes, and taxes paid as part of the production process, such as excise taxes and import duties.[41] We estimate that the data annotation industry supported almost $1.2 billion in total tax revenues in 2024, out of which $854 million (68%) were federal tax revenues and $395 million (32%) were state and local tax revenues (Fig. 6).

Fig. 6. Total tax impact of the data annotation industry, by level of government, 2024



Source: Oxford Economics, IMPLAN

[39] This is the total number of earning opportunities directly supported by the data annotation industry. We adjust earning opportunities for multi-homing.

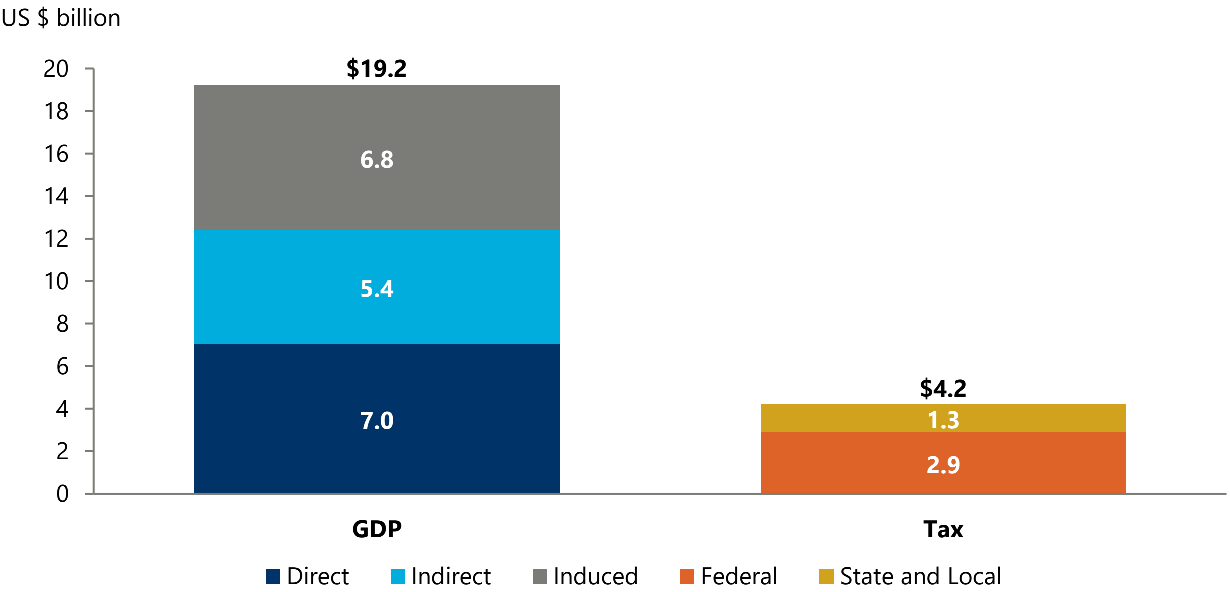
[40] Or 28,000 jobs on a headcount basis.

[41] Specifically, the IMPLAN software includes the following federal tax categories: social insurance (Social Security and Medicare), personal income, corporate income, excise taxes, customs duties, and other taxes on production and imports. These categories are all included in our results. In addition, IMPLAN software includes the following state and local tax categories: social insurance, personal income, taxes on corporate profits and dividends, personal and business property taxes, various sales and excise taxes, motor vehicle taxes, severance tax, and other personal and business taxes.

3.2 THE ECONOMIC IMPACT OF THE DATA ANNOTATION INDUSTRY IN 2030

We project the size of the US data annotation market to range between $9.2 billion and $16.9 billion in 2030.[42] Based on this forecast, we estimate that the GDP impact of US data annotation firms in 2030 will total $19.2 billion (an increase of 237% from 2024), of which $7.0 billion represents the direct GDP impact.[43] We expect the tax impact to amount to approximately $4.2 billion (an increase of 250%)[44] Fig. 7 below illustrates the estimated GDP, tax, and employment impacts in 2030, broken down by impact channel.

Fig. 7. GDP and tax impacts of the data annotation industry, 2030[45]



Source: Oxford Economics, IMPLAN
Totals may not sum due to rounding.

[42] We project revenue in 2030 will range between $9.2 billion and $16.9 billion in 2024 real dollars. When expressed in nominal 2030 dollars (including forecast inflation), this is $10.3 billion to $19.0 billion.
[43] GDP and tax impact forecast for 2030 are in 2024-dollar year.
[44] Economic impact modeling output estimates for the US are based on the midpoint ($13.0 billion) of the lower ($9.2 billion) and upper bound ($16.9 billion) of the revenue estimates in 2024 real dollars.
[45] Dollar values have been adjusted to 2024 prices.

4. ANNOTATORS IN THE AI DATA INDUSTRY

Despite the critical and growing role data annotation plays in the development of AI models—and the importance of human input in this ecosystem—the individuals performing this work remain relatively under-studied.

Platforms such as Outlier, operated by industry leader, Scale AI, provide individuals from across a range of backgrounds and professions the ability to sign up for an account and offer their expertise in conducting data annotation work. Scale and its competitors frequently match professionals to the annotation tasks requested by its AI model-building customers, before collecting and providing the labeled high-quality data and prompt engineering as a product. The resulting flexible work opportunities grant professionals greater schedule flexibility and the freedom to engage in project-based roles rather than traditional full-time positions.

Overall, data annotation appears to be positioned at a unique intersection of the platform economy and the specialized skill demands of AI/ML workflows, appealing to technically proficient individuals who require flexible work arrangements. This chapter discusses the findings of a survey we conducted to better understand the profile, motivations, and career goals of the data annotation workforce.[46] The analysis that follows is based on a sample of 914 US-based survey respondents, who are data annotators on Scale's Outlier AI platform.

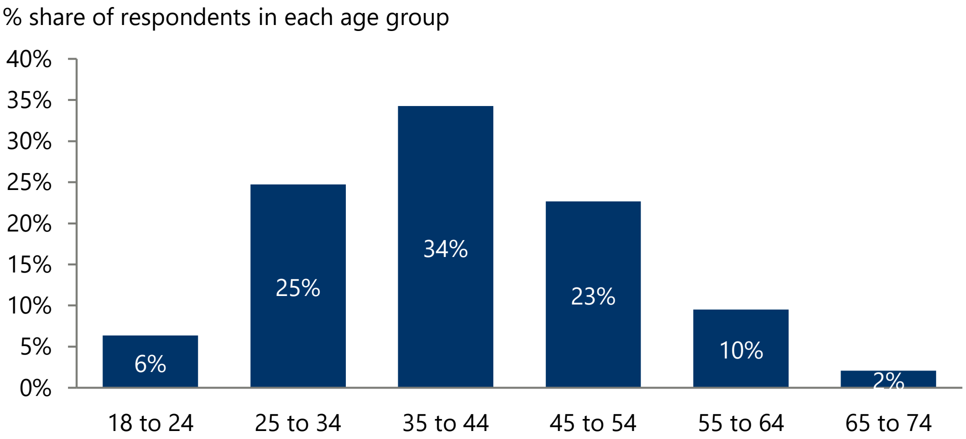
4.1 WHO ARE DATA ANNOTATORS?

4.1.1 Data annotators reflect a culturally diverse and middle-aged workforce

The surveyed data annotation workforce has a median age between 35 and 44 years old. This age profile of data annotators is similar to that of the US workforce, where the median age is approximately 42 years,[47] and individuals under the age of 44 make up 57% of the US workforce,[48] compared to the survey sample's 65% (Fig. 8).

[46] More detail on our survey methodology is in the appendices.
[47] Richardson, N., "The Age of Work", part two: The country's youngest workforce", *ADP Research Institute*, 2025.
[48] U.S. Bureau of Labor Statistics, "Civilian labor force, by age, sex, race, and ethnicity.", *United States Department of Labor*, 2024.

Fig. 8. Age profile of data annotators



Sample: n = 914
What is your age? (Single answer)
Total may not sum due to rounding.
Source: Oxford Economics

Survey respondents were predominantly women, with 58% identifying as female and 39% as male. Respondents speak a wide range of languages spoken at home, with nearly 30 languages cited. Of these, English topped the list, with 84% of respondents reporting it as their primary language. 35% of respondents speak more than one language besides their primary language, such as Spanish, Russian, and Chinese. In terms of racial and ethnic demographics, around two-thirds of respondents identified as White, 14% as Asian, 13% as Black or African American, and 9% as Hispanic or Latino.[49]

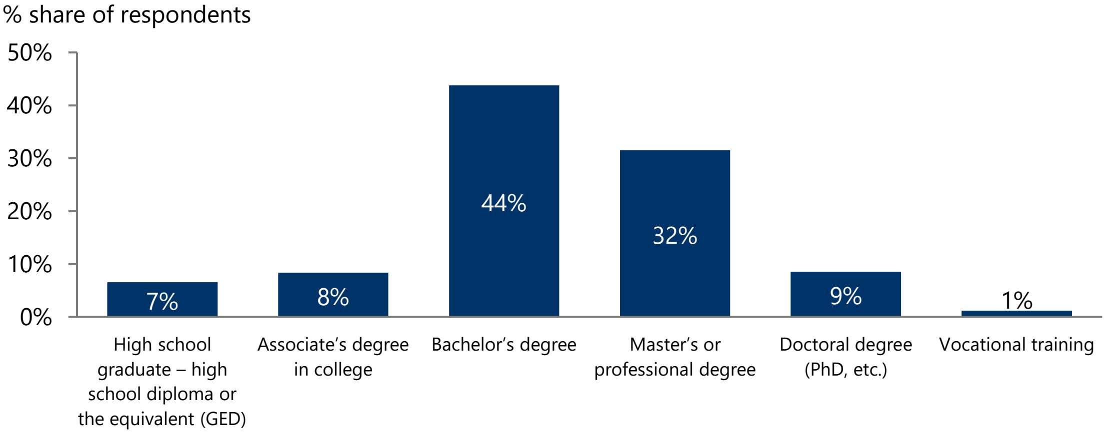
4.1.2 AI annotators are highly educated and have an extensive skillset

AI annotators are highly educated. Among the respondents, 84% held at least a bachelor's degree, which is double the share among the wider US workforce (42%) (Fig. 9).[50] Over four in ten respondents held at least a master's or a doctoral degree, whereas only 17% of the US workforce held a master's or a doctoral degree. This reflects the often technically challenging nature of data annotation work, especially for recent, more advanced GenAI models, which require subject matter expertise when providing reinforcement feedback.

[49] Some respondents identified with more than one race.

[50] U.S. Bureau of Labor Statistics., "Educational attainment for workers 25 years and older by detailed occupation", *United States Department of Labor*, 2022.

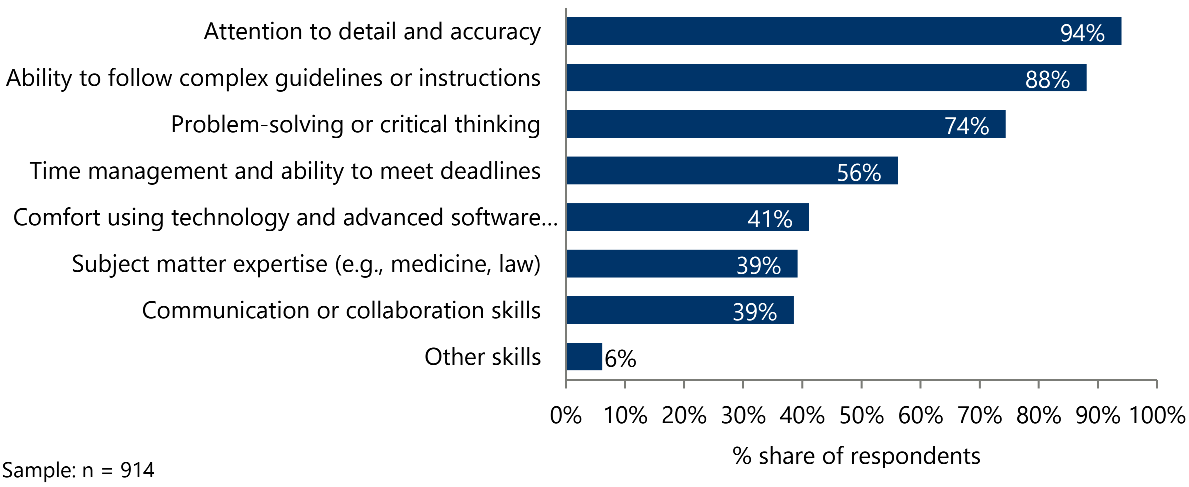
Fig. 9. Highest level of education of data annotators



Sample: n = 911
What is the highest level of education you have completed? (Single answer)
Source: Oxford Economics

There was consensus among respondents regarding the non-domain specific essential skills needed to succeed in data annotation. Over nine in 10 (94%) respondents agreed that "attention to detail and accuracy" was important, along with "the ability to follow complex guidelines or instructions" (88%), problem solving (74%), and time management skills (56%), among others (Fig. 10). Adaptability, flexibility, and the ability to deal with ambiguity are also key attributes for performing annotator work effectively. In open-ended responses, respondents emphasized that skills such as the "ability to adapt to an ever-changing AI environment" as well as the "ability to discern errors and deal with ambiguous instructions" were essential for thriving in the role. Others pointed to the importance of "quick understanding of complex instructions", "writing ability", and "the ability to read between the lines", underscoring the dynamic skillsets needed.

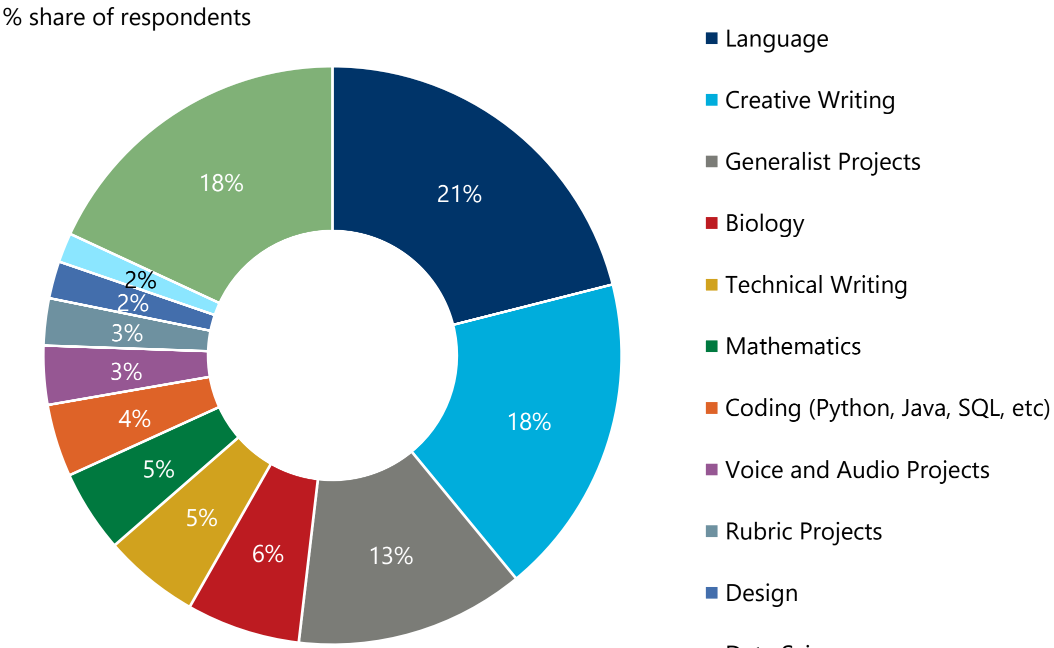
Fig. 10. Skills required for data annotator work



Sample: n = 914
In your opinion, which of the following skills are most important for performing your contributor work effectively? (Multiple answer)
Source: Oxford Economics

We asked annotators what domain of AI model training they contributed the most to. More than a fifth of the annotators reported primarily working in the "Language" domain (Fig. 11), followed by "Creative Writing" (18%) and "Generalist Projects" (13%). Some respondents specified areas such as "AI response quality check" and "evaluating prompt responses from AI." The range of types of contribution were varied; while some respondents contributed to more technical domains (such as biology, mathematics, coding, and data science), others mainly contributed to more niche areas of data annotation—namely rubric creation for reviewing models, or voice and audio projects involved in the creation of speech datasets for new model training.

Fig. 11. Domain of data annotators



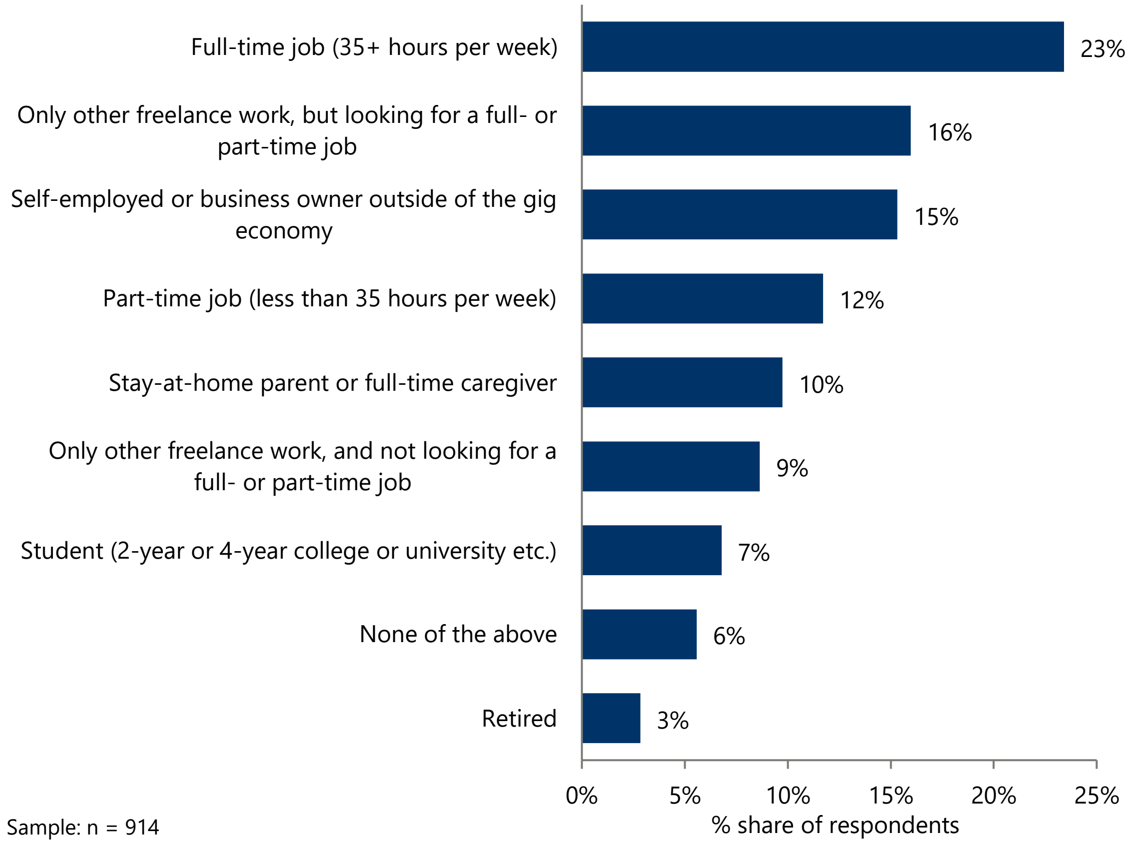
Sample: n = 912
Which domain is the one where you have contributed the most? (Single answer)
Source: Oxford Economics

4.1.3 94% of data annotators are engaged in other professional or educational activities outside their annotation work

Respondents are active members of the US labor force, and many engage in data annotation as supplemental work alongside other jobs—out of all respondents, only 6% are not engaged in any other professional or academic activity. Across respondents, 23% had full time jobs,[51] 12% were part-time workers, and 15% were self-employed (Fig. 12). 25% self-identified as only doing other freelance work, of which more than six in 10 reported actively looking for a full- or part-time job. One in 10 respondents were stay-home parents or full-time caregivers, while 7% are students.

[51] The Bureau of Labor Statistics defines a full-time worker as those working 35 hours or more per week. BLS, "Labor Force Statistics from the Current Population Survey."

Fig. 12. Data annotators engaged in other professional activities



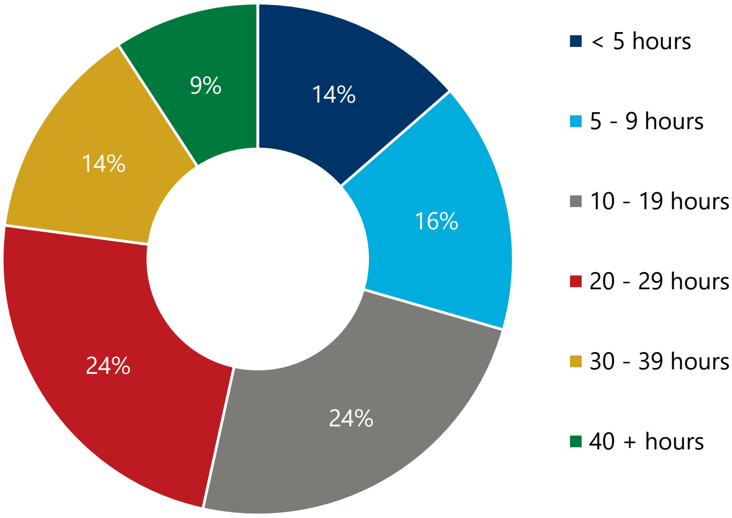
Sample: n = 914
What is the primary professional or educational activity that you engage in outside of your work as a contributor (Single answer)
Source: Oxford Economics

Among those who were employed or self-employed, many worked in education and training (19%), arts, design, entertainment and media (16%), business and finance (10%), and computer and mathematical industries (9%). Most of these individuals were also employed in relatively senior positions: almost half (48%) were in their mid-careers, whereas nearly a third (32%) held senior management or leadership positions.

The time annotators devote to data annotation is shaped by the other roles and responsibilities they balance in their lives. Across all respondents, three in 10 spend less than 9 hours on data annotation work on average (Fig. 13). More than half spend less than 20 hours per week and more than three quarters less than 30. Annotators with full-time jobs tended to spend less time as annotators, with nearly 75% of them dedicating fewer than 19 hours weekly. In contrast, those that may have more flexibility in their schedules are able to commit more hours to their annotation work. For example, more than 10% of freelancers and the group consisting of stay-at-home parents or full-time caregivers committed to working over 40 hours per week.

Fig. 13. Average hours per week spent on data annotation

% share of all respondents

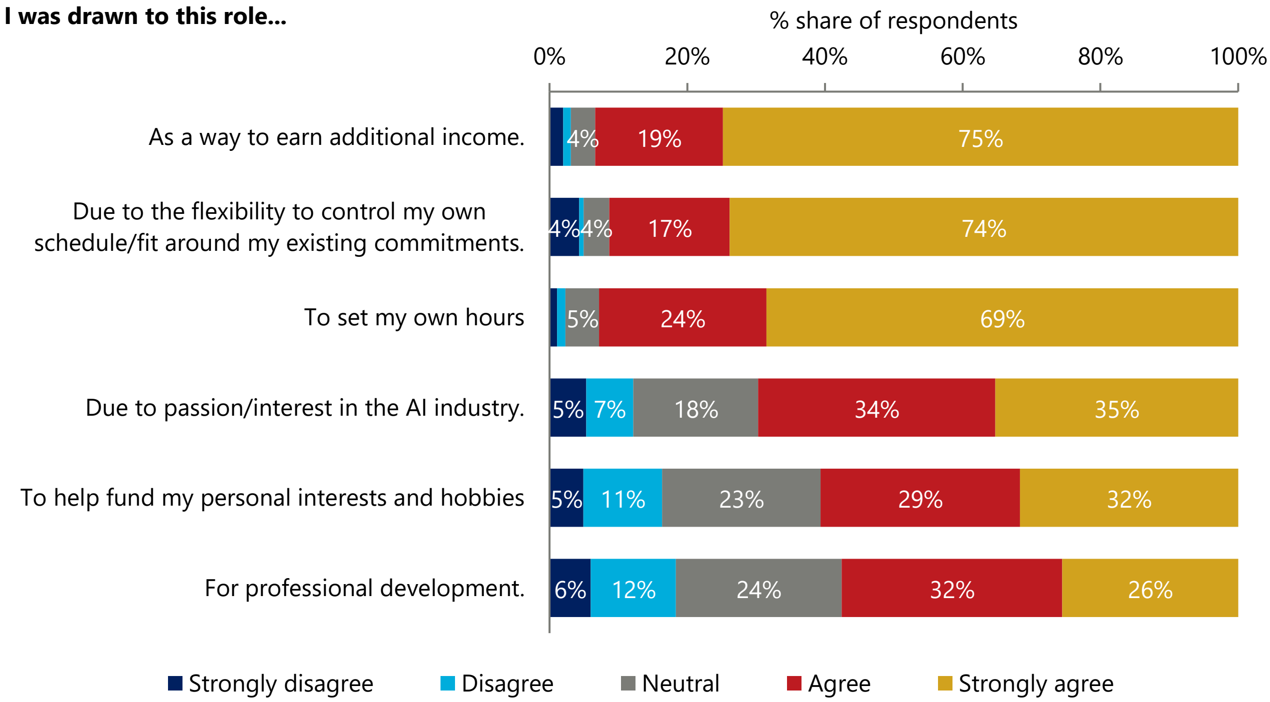


Sample: n = 913
How many hours do you usually spend as a contributor each week on average? (Include onboarding, research, etc.) (Single answer)
Source: Oxford Economics

4.2 MOTIVATIONS FOR PERFORMING DATA ANNOTATION WORK

The survey respondents reported a wide variety of motivations for engaging in data annotation, which span financial, practical, personal, and professional incentives. Respondents agreed that income (94%), the ability to set one's own hours (93%), and flexibility (91%) were the top motivations for working as data annotators (Fig. 14).

Fig. 14. Motivations for working as data annotators



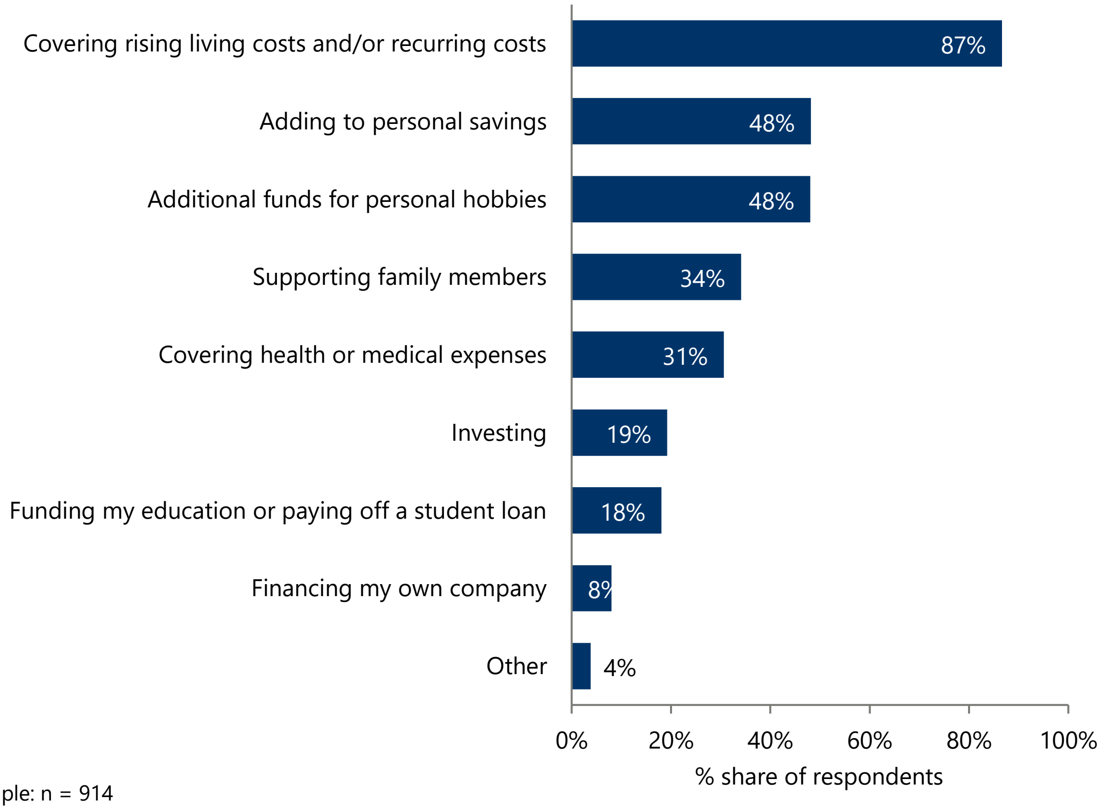
Sample: n = 906 (Additional income), 910 (Flexibility), 902 (Set own hours), 904 (Passion for AI), 899 (Personal interests), 900 (Professional Development).
To what extent do you agree with the following statements regarding your role as a contributor at Outlier AI? (Single answer)
Source: Oxford Economics

4.2.1 Income earned is a key motivator for annotator work

Annotator work is often used to cover rising living costs. The most common use of annotator income was to cover rising living and recurring costs, with 87% of respondents reporting its value for this purpose. Nearly half of the respondents used their annotator income to add to personal savings, 34% reported using it to support their family members, while three in 10 used it for health and medical expenses—highlighting the role of this income in addressing both practical and personal livelihood needs[52] (Fig. 15).

[52] These findings are consistent with recent studies looking at other examples of alternative work arrangements. Fos et al (2025) found that flexible work can serve to smooth extreme income shocks, with unemployed workers with access to Uber relying less on household debt and less likely to rely on unemployment insurance. Vyacheslav Fos, Naser Hamdi, Ankit Kalda, and Jordan Nickerson. (2025). "Gig labor: Trading safety nets for steering wheels," *Journal of Financial Economics*, vol. 163, issue C. See also

Fig. 15. Uses of income from data annotation work



Sample: n = 914
Which, if any, of the following do you use your income from this work for? (Multiple answer)
Source: Oxford Economics

There is also evidence suggesting that stay-home parents and full-time caregivers used their earnings to support their families (58%)—compared to 34% of all respondents. For these respondents, their annotator work played a role in supporting their livelihoods, as well as ensuring their families' wellbeing.

58%
Stay-home parents and/or full-time caregivers used their earnings to support their families

Annotator incomes also support individuals in pursuing their personal goals. Beyond helping to meet rising living costs, many data annotators reported using their earnings to pursue diverse personal goals, such as achieving financial independence (by investing and saving), funding personal hobbies (48%), and funding academic pursuits. A common goal for many annotators was building wealth; nearly half (48%) of all respondents reported using their annotator income to boost their savings, while 19% allocated it to investments. This is especially so for

Jones and Manhique (2024) who found some evidence that digital platforms can help labor markets in low-income settings with widespread informal work to adjust to economic shocks. Jones, S., & Manhique, I. (2024). "Digital labour platforms as shock absorbers: Evidence from the COVID-19 pandemic in Mozambique." *Journal of African Economies*, 34(1), 116-141.

students—Fig. 1556% of students used their annotator income to boost their savings, and 27% invested their annotator income.

In the open-ended responses, several respondents also added that they used their annotator income to "save for early retirement", and to "pay off debt(s)"—suggesting that this role also supports them in meeting their individual financial obligations and goals. In addition, a few respondents specified that they used their earnings to support other miscellaneous expenses, such as "large expenses like car repairs and home maintenance".

Certain groups were especially focused on using their earnings to achieve specific objectives. For students, annotator income was often directed towards funding their (current or future) education, with over half (55%) of them relying on it to cover tuition or to pay off student loans—more than three times the share across all respondents. Meanwhile, self-employed business owners frequently used their annotator earnings to support their own companies (24%), using this income to finance operations, expand their offerings, or sustain their entrepreneurial ventures until their businesses were more established. These targeted uses of income demonstrate how annotators leveraged their work to not only support their present needs, but also to create opportunities for future growth.

Annotators will turn to other income-earning opportunities when data annotation work is not available. Working as a data annotator empowered individuals with a flexible financial resource to cover their immediate costs and build a more stable financial future. In the absence of such annotator work, 93% of respondents reported that they would seek other income-generating opportunities to fill their gap in earnings, while nearly four in 10 (39%) of them would cut their personal expenses. Others would deplete their savings (16%) or take on additional debts (13%). Younger annotators would be more likely to reduce their expenditures (43% of annotators aged 18 to 24 compared to 36% of annotators aged 45 and above) and become more dependent on family support (21% of annotators aged 18 to 24, relative to 15% of all respondents).

4.2.2 Annotator work offers flexible work arrangements

The flexibility associated with the work is another key motivation for annotators, who juggle multiple roles and responsibilities (see Section 4.2, Fig. 14). As discussed, 91% of respondents agreed that they were drawn to the work because of the flexibility to manage their own schedule and accommodate existing commitments, with 74% strongly agreeing. The key aspects of this flexibility are reflected in respondents' preferences. The most highly valued aspect (74% of respondents) was the **ability to balance the work with other professional or educational endeavors** (Appendix 3, Fig. 22Fig. 22). Half of respondents emphasized the importance of being able to work during hours that do not conflict with their full- or part-time jobs. This flexibility allowed annotators to earn an income without sacrificing other career or academic pursuits (Fig. 16). 36% of respondents also value annotation work as a flexible source of supplementary income (Fig. 16).

The **ability to be present for the family** is also important, with nearly a third (32%) of all respondents valuing this benefit. Many annotators, especially those with caregiving responsibilities, also valued the option to be available for loved ones, with 31% and 11% emphasizing the ability to care for children and other dependents, respectively (Fig. 16).

Having time for personal enrichment and health were also key aspects, with 30% of respondents valuing the time to work on their own projects and 23% appreciating the opportunity to pursue hobbies. While less common, 13% of respondents noted the importance of having time for healthcare needs, further underscoring the holistic nature of flexibility in this work[53] (Fig. 16).

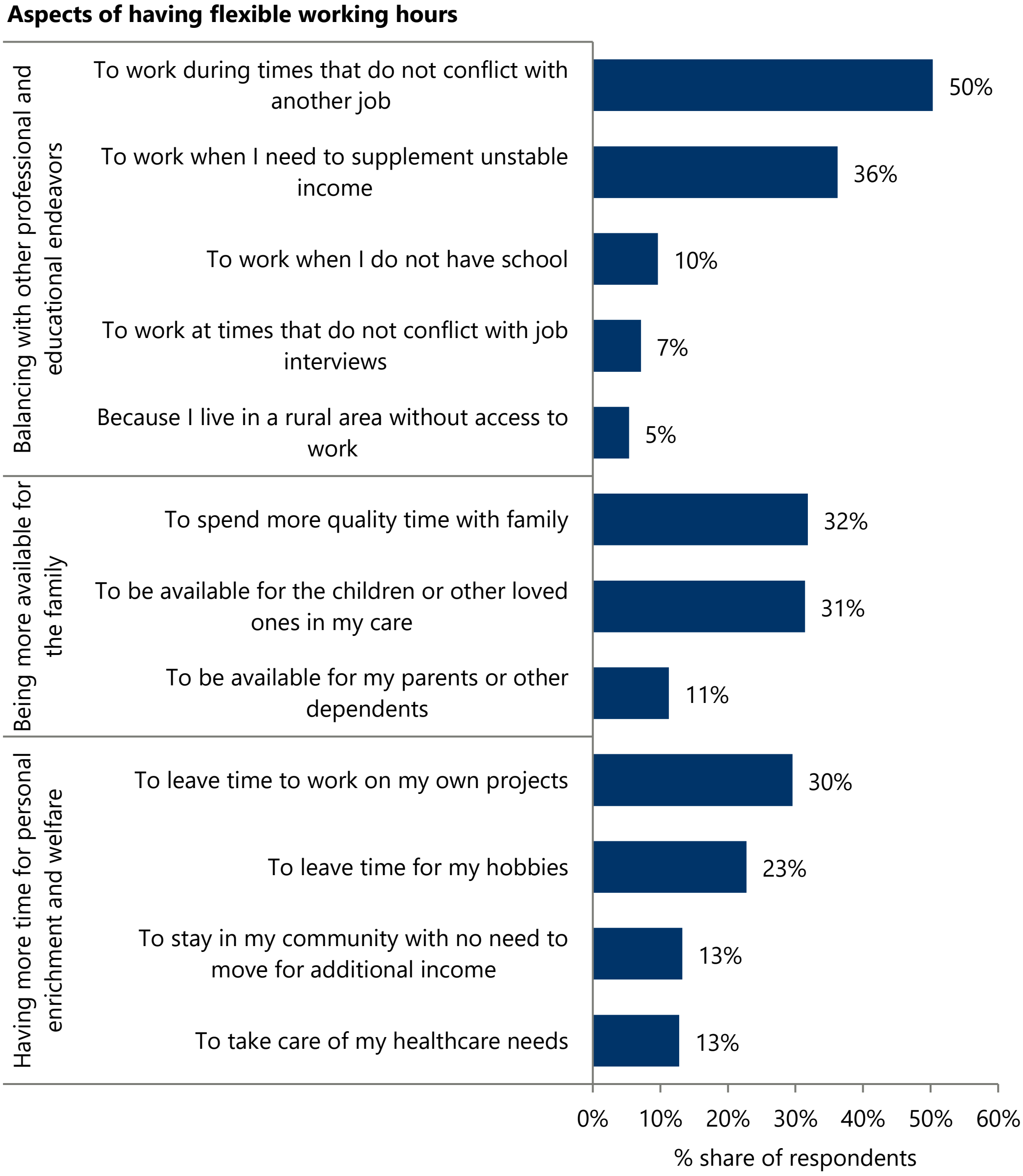
Different groups prioritize these three broad aspects of flexibility in varying ways. Those that tend to prioritize the flexibility to balance with professional or educational commitments included male respondents (83%), students (100%), and those with existing jobs (97% for full-time workers, 93% for part-time workers) (Appendix 3, Fig. 20, and Appendix 3, Fig. 22). For these groups, data annotation offered a way to supplement income or gain valuable experience without interfering with their academic commitments or full-time work. The flexibility to work around school hours or existing job schedules seemed particularly crucial for these segments, as they were navigating a phase where career paths, education, and early professional experiences were key drivers of their future success.

In contrast, annotators in the 35 to 44 age group (68%), women (62%), and those in caregiving roles (79%) placed a higher value on being available for family, compared to the wider respondent sample (Appendix 3, Fig. 20 and Appendix 3, Fig. 21). This sentiment is most strongly felt among stay-home parents and full-time caregivers (96%) (Appendix 3, Fig. 22). These annotators shouldered more domestic responsibilities, making the ability to schedule work around family needs an essential aspect of their involvement in data annotation.

Meanwhile, self-employed business owners (60%), freelance workers not looking for a job (72%), and freelance workers looking for a job (79%) valued the flexibility to invest time in personal growth and welfare (Appendix 3, Fig. 22). For these individuals, data annotation offered the flexibility to pursue side projects, hobbies, or self-improvement activities without the constraints of a traditional work schedule.

[53] In the literature around alternative work arrangements there are estimates of the value of flexibility and independence. Katsnelson and Oberholzer-Gee (2021), examining food delivery drivers, found that, on average, forcing workers out of their preferred shift (i.e., reducing flexibility) was equivalent to cutting their weekly earnings by 5.3%. See Anderson, M., McClain, C., Faverio, M., & Gelles-Watnick, R., "The state of gig work in 2021" *Pew Research Center*, 2021 and Katsnelson and Oberholzer-Gee, "Being the Boss: Gig Workers' Value of Flexible Work", *Harvard Business School Working Paper*, No. 21-124, May 2021.

Fig. 16. Important aspects of having flexible working arrangements



Sample: n = 831 (Respondents who agree/strongly agreed that flexibility was a motivation)
What are the most important aspects of having flexible working hours? (Multiple answer)
Source: Oxford Economics

In essence, the differing priorities among these groups reflect the diverse ways in which flexibility serves as both a practical and strategic tool for managing competing demands and commitments. Students and working professionals tend to be more focused on developing their skills and experience through their studies and jobs, while respondents with families pursue a balance between work and family life. On the other hand, self-employed workers might seek pursuits that align with their broader personal and/or professional goals. This rich array of priorities illustrates how flexibility in data annotation can cater to a wide spectrum of needs, in a way that more rigid employment structures often cannot accommodate.

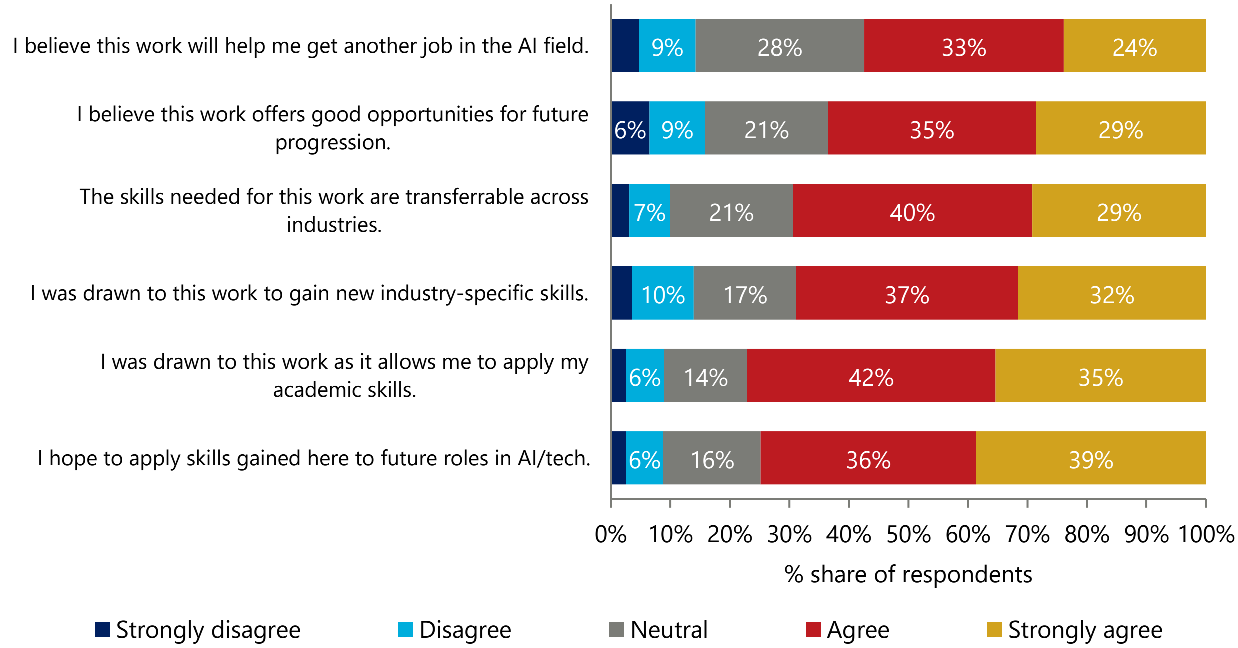
Moreover, respondents also reported having great appreciation for the autonomy that comes with flexible working hours, mentioning that they could tailor their schedules to fit personal needs and responsibilities. 93% of all respondents said that they were drawn to the role due to the ability to set their own hours (see Section 4.2, Fig. 14). In open-ended responses, responses cited included having "the freedom to choose when to work", such as during peak productivity times or when they had spare time.

4.2.3 Respondents see annotator roles as providing a pathway for professional development

Data annotators are highly motivated by the professional growth opportunities that this role offers, with many regarding the role as a launchpad to future careers in AI and technology. Three-quarters (75%) of respondents expressed hope that the skills gained from data annotation would be valuable in their future roles within AI and technology, with over half of these respondents (39%) strongly agreeing with this sentiment. (Fig. 17) Similarly, the ability to apply their academic knowledge in a practical setting resonated with 77% of annotators, with more than a third (35%) strongly agreeing that the role allowed them to integrate what they have learned into real-world applications.

This aligns with the broader belief that the job provides strong prospects for future career advancement—64% of respondents agreed that the role offered opportunities for progression, with nearly three in 10 (29%) in strong agreement. Among annotators, 69% also regarded the role as a chance to acquire valuable, industry-specific skills to enhance their employability and open pathways into the technology sector (Fig. 17).

Fig. 17. Professional growth opportunities gained through data annotation



Sample: n = 899 (Get another job in AI), 909 (Good opportunities for progression), 905 (Transferrable skills), 905 (Industry-specific skills), 899 (Apply acacdemic skills), 906 (Apply skills in future roles in AI)
To what extent do you agree with the following statements about your job as a contributor? (Single answer)
Source: Oxford Economics

The appeal of professional development was also a key driver for many respondents, with 58% of annotators drawn to the job specifically for its potential to aid their career growth (see Section 4.2, Fig. 14). Moreover, the skills required for data annotation were seen as transferable, with 69% of respondents agreeing that these competencies could be applied across multiple industries, making the role a valuable experience for those looking to diversify their professional portfolios (Fig. 17).

Overall, working as a data annotator was not just seen as a source of income, but as an important part of building a future in adjacent emerging fields, equipping annotators with relevant and marketable skills for the long term.

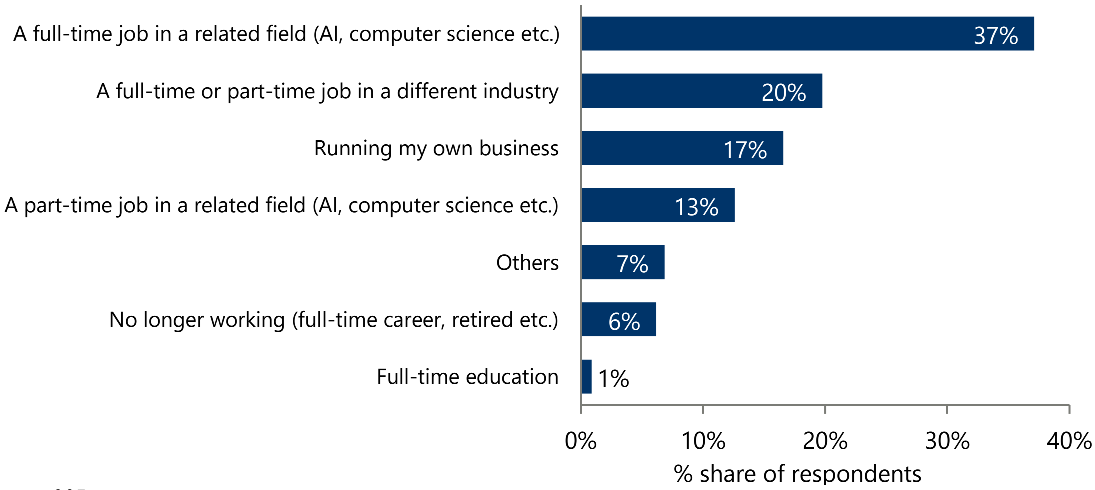
However, for many data annotators, the motivation to take on this role went beyond career progression. Data annotators are fueled by a deep passion for AI and a genuine interest in the technology sector; nearly 70% of respondents expressed that they were drawn to the job because of their passion for the AI industry, with nearly half (35%) strongly agreeing that this was a major factor in their decision (see Section 4.2, Fig. 14).

4.2.4 Future aspirations of annotators

Data annotators exhibit strong ambitions for their future in AI and related tech fields, with many aligning their career goals with the ongoing technological shift. Half of the respondents aspired to work in AI, computer science, or software development—with nearly four in 10 seeking to work full-time (37%), while 13% aimed for part-time roles in these areas. Additionally, 17% expressed entrepreneurial aspirations, seeking to build their own businesses and 20% of annotators planned to pursue careers in different industries (Fig. 18).

The alignment between their present work and future goals suggest that some annotators view their involvement in AI as part of a long-term journey toward growth and innovation, reinforcing their commitment to staying connected to the technology industry's evolution.

Fig. 18. Data annotators' career aspirations



Sample: n = 905
Looking ahead to the next 5-10 years, which best describes where you see yourself? (Single answer)
Source: Oxford Economics

5. CONCLUSION

The data annotation industry drives economic impact while supporting technological progress and providing flexibility in the labor market. It helps power the growing AI ecosystem by supplying reliable, high-quality datasets that are indispensable for model development, while at the same time providing its workforce with a pragmatic income opportunity that aligns with their diverse life circumstances and long-term ambitions.

Data annotation is more than an income stream for annotators. Its flexible nature allows annotators to accommodate their professional, family, and educational commitments. Flexible roles in the data annotation industry support annotators in pursuing a variety of personal goals, from financial independence to entrepreneurial ventures to academic pursuits. For many, this work is viewed as a potential springboard to transition into more advanced roles.

The industry's influence extends well beyond its workers. By facilitating a $5.7 billion contribution to GDP and driving $1.2 billion of tax revenues, the data annotation industry has established itself as a rapidly growing force in the US economy. Additionally, the data annotation industry directly provides more than 200,000 earning opportunities for people in the US economy as well as around 9,000 full-time jobs. The industry plays a critical role in the AI value chain, ensuring that innovation and advancements in AI rests on human judgement, precision, and adaptability. As demand for AI accelerates, data annotation will continue to create opportunities for a highly educated and motivated workforce and will influence how well society navigates the broader transitions brought about by AI adoption.

Recognizing the dual importance of the data annotation industry—both as an enabler of AI progress and as a flexible platform catering to the needs of a diverse workforce—will be important for policymakers, businesses, and educators. Supporting the conditions that keep data annotation work accessible and rewarding will ensure that the industry can grow sustainably, while also strengthening the pipeline of talent that will help to drive the next generation of AI development.

6. APPENDIX 1: ECONOMIC IMPACT ANALYSIS

This technical annex explains the approach and assumptions used to quantify the economic impact of the data annotation market in the US economy in 2024 and forecasts this impact in 2030.

A. METHODOLOGY: INPUT-OUTPUT MODELING

The economic impact analysis is conducted using an Input-Output model. An Input-Output model provides a detailed snapshot of an economy for a single year, capturing the flow of goods and services between various sectors. It describes how the output of one industry is an input for another, effectively mapping the relationships and interdependencies among industries within a region or country.

The model is typically presented in the form of a matrix, where each row shows the distribution of an industry's output to other industries and to final demand (e.g., households, government, exports), and each column shows the inputs that an industry uses to produce its output. This framework allows analysts to trace the flow of spending through an economy, thereby quantifying the effects on supply chains, consumer spending, economic leakages, and government revenues.

This report uses the IMPLAN economic impact software, which is an Input-Output modelling system used to build models at various levels of geography, including national, state, county, and congressional district. It allows for adjustable assumptions of supply-chain connections and leakages from input data and improves accuracy of assumptions for missing data.

IMPLAN data contains 528 sectors representing all private industries in the United States (e.g., from grain farming and oil and gas extraction to software publisher and automotive repair and maintenance) based on the North American Industry Classification System (NAICS) codes. Employment, employee compensation, industry expenditures, commodity demands, relationships between industries, and more are collected to form IMPLAN's database.

B. MODELING INPUTS: INDUSTRY SIZE

The IMPLAN software typically incorporates output, procurement, and/or compensation of workers as key inputs to estimate the economic impact of an organization or industry. In this analysis, we use the revenue of the entire data annotation market as the primary input for the economic impact analysis. Secondary data sources were used to obtain industry revenue estimates.

We have taken two approaches to estimate the global revenue related to data annotation. First, we reviewed a set of publicly available market research reports providing estimates of global market revenues.[54] The mean estimate was **$15.1 billion** in 2024. In our second approach, we compiled a list

[54] "Data Labeling Solution And Services Market Size, Share & Trends", *Research and Markets*, 2024., "Data Labeling Solution and Services Market", *Fact.MR*, 2024., "Data Labeling Solution And Services Market Size", *Grand View Research*, 2024., "Data Labeling

of companies active in the space and aggregated publicly available revenue data for these companies. This is a conservative estimate, as some significant players do not report revenues attributable to these services. Based on this research, our estimate for data annotation market size in 2024 was around **$8.2 billion.** According to our survey of the same set of publicly available reports, the United States accounted for approximately a third (33%) of the global market in terms of revenue. Thus, our estimate for the US data annotation market size ranged from **$2.7 billion to $5.0 billion**. Our economic impact modeling conservatively bases the output estimates for the US on the midpoint **($3.8 billion)**.

We also referred to existing market research reports to gather forecast growth rates for the data annotation market, to project the size of the US market by 2030. The industry is consensually expected to experience rapid growth in the coming decade, with some estimates forecasting a compound annual growth rate (CAGR) of 25%[55]. Applying the 25% CAGR to our 2024 estimates provides a range of the US market size of **$10.4 billion to $19.0**[56] **billion by 2030 (in 2030-dollar terms)**. We used the midpoint of the range above **($14.9 billion)** to estimate economic impact in 2030. We inflation-adjusted the 2030 figure to 2024 dollars **($13.0 billion)** before running the IMPLAN model.

C. THREE CHANNELS OF IMPACT

A standard economic impact assessment identifies three channels of impact that stem from the activity of an industry—the direct, indirect, and induced impact.

Direct impacts refer to the economic effects generated by the activities of data annotation firms. This includes the wages and salaries paid to their workforce, along with the direct taxes these firms pay to the government.

Indirect (supply chain) impacts refer to the economic activity supported through the procurement of goods and services by data annotation firms. This includes the purchases made by suppliers who provide technology, software, office equipment, and other necessary inputs. Additionally, it encompasses the subsequent spending by those suppliers within their own supply chains, creating a ripple effect through the economy. This channel also includes expenditures related to office setup and maintenance, such as costs for leasing or renovating workspace, purchasing hardware and furniture, and investing in necessary infrastructure to support data annotating operations.

Induced (workers' spending) impacts reflect the economic activity generated by the spending of wages and salaries earned by direct employees, as well as by employees working at companies within the data annotation firm's supply chain.

The three channels of impact are presented in terms of the following three metrics.

- **Gross Domestic Product (GDP).** This is the value of the output produced by data annotation companies and their annotators, minus their intermediate expenditure on inputs (goods and

Market Size and Share Analysis – Growth Trends & Forecasts", *Mordor Intelligence*, 2025., "Generative AI in Data Labeling Solution and Services Market", Dimension Market Research, 2024., Zoting, S., "Data Labeling Solution and Services Market Size and Forecast 2025 to 2034", *Precedence Research*, July 8, 2025.

[55] Global Market Insights, "Data Annotation Tools Market Revenue", *AWS Marketplace*, 2022.

[56] The US market size for 2030 is reported in nominal 2030 dollars.

services) that are used in their operations. Aggregated across all economic operators in the economy, this forms GDP (when added with production taxes and subsidies), which is the most widely recognized measure of total economic output.

- **Earning opportunities**. The report also estimates the number of people benefiting from total earning opportunities generated by the data annotation industry in 2024, using data from Scale AI combined with our estimates for the 2024 US data annotation industry market size.
- **Tax revenues.** Activities in the data annotation industry generate significant tax benefits that can be calculated. We estimate federal, state, and local tax contributions using IMPLAN data. The types of tax impacts include social insurance taxes, taxes on production and imports (TOPI), corporate profit taxes, and personal taxes.

Our results are presented on a gross basis. They therefore ignore any displacement of activity from other uses of the resources used in data annotation.

We present our findings in 2024 prices. When adjusting prices to real terms, we use OE proprietary forecasts for economy-wide deflators.

7. APPENDIX 2: SURVEY DESIGN

SURVEY OBJECTIVES AND STRUCTURE

Oxford Economics conducted a survey of Scale's data annotators (who use Scale's wholly owned Outlier AI platform) in October 2025. Through this survey, we aimed to profile the data annotation workforce, explore the annotators' perceptions and experiences of their roles, and examine the economic and personal value they derive.

To generate the results outlined in this report, Oxford Economics, in consultation with Scale AI, conducted a literature review and prepared a series of survey questions to better understand the data annotator workforce. The survey was structured to allow for both qualitative and quantitative insights into the nature of this workforce. Annotators were prompted to reflect on the characteristics of their roles, such as the independence to control their own schedule, and the ability to earn an income on their own terms.

In addition to insights into the role that flexibility and financial benefits play in annotators' choice to join this workforce, the survey also explored the skills required to contribute to AI oversight, the potential wider applications of these skills, as well as the workers' future aspirations.

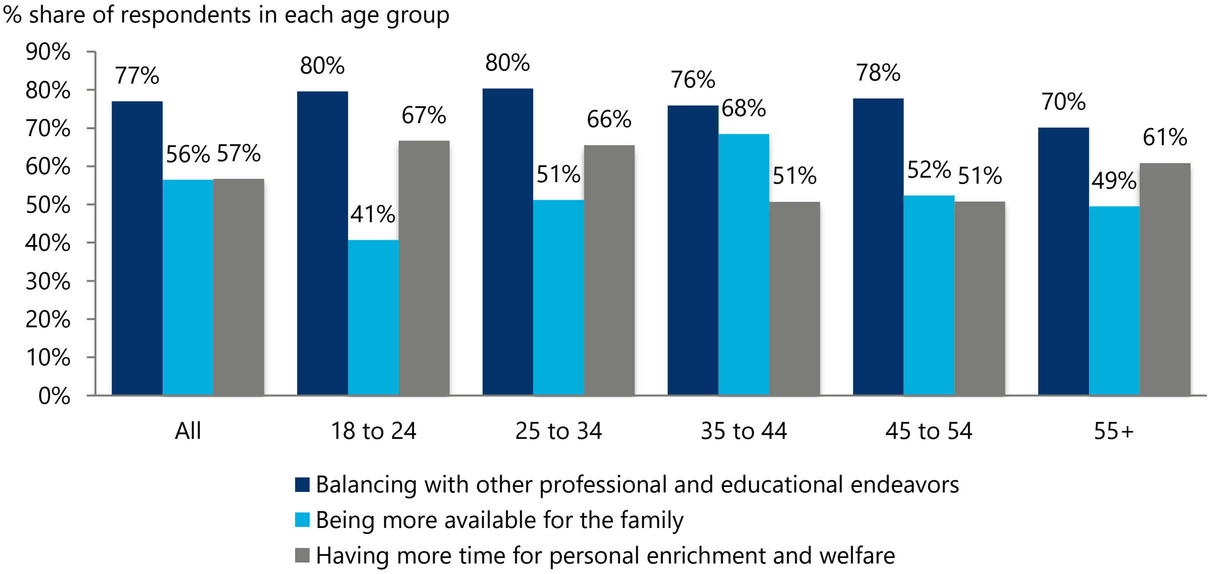
SURVEY ADMINISTRATION

The survey was distributed via email to a randomly selected set of Outlier AI annotators based in the United States. It was administered online through the SurveyMonkey platform, and participation was voluntary and anonymous. The survey consisted of approximately 30 questions, which respondents on average completed in around 9 to 10 minutes.

Following data quality checks, the responses were filtered to a sample of 914 data annotators for analysis. The checks led to the removal of responses that were self-contradictory, straight-line responses, or responses where less than 95% of questions were answered.

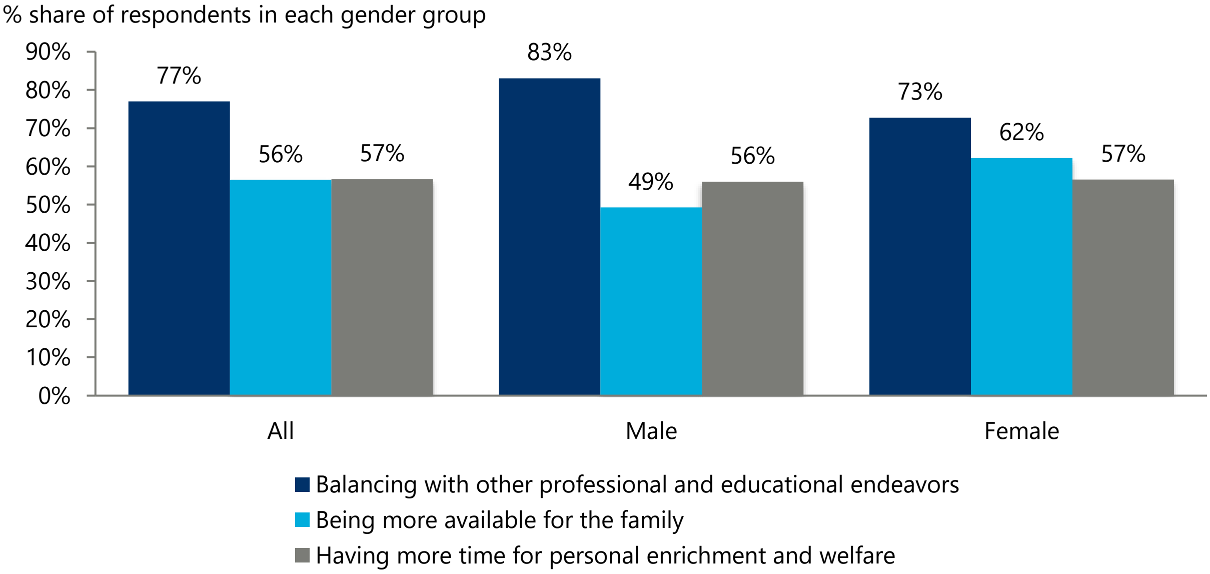
8. APPENDIX 3: ADDITIONAL SURVEY FINDINGS

Fig. 19. Important aspects of having flexible working arrangements, by age



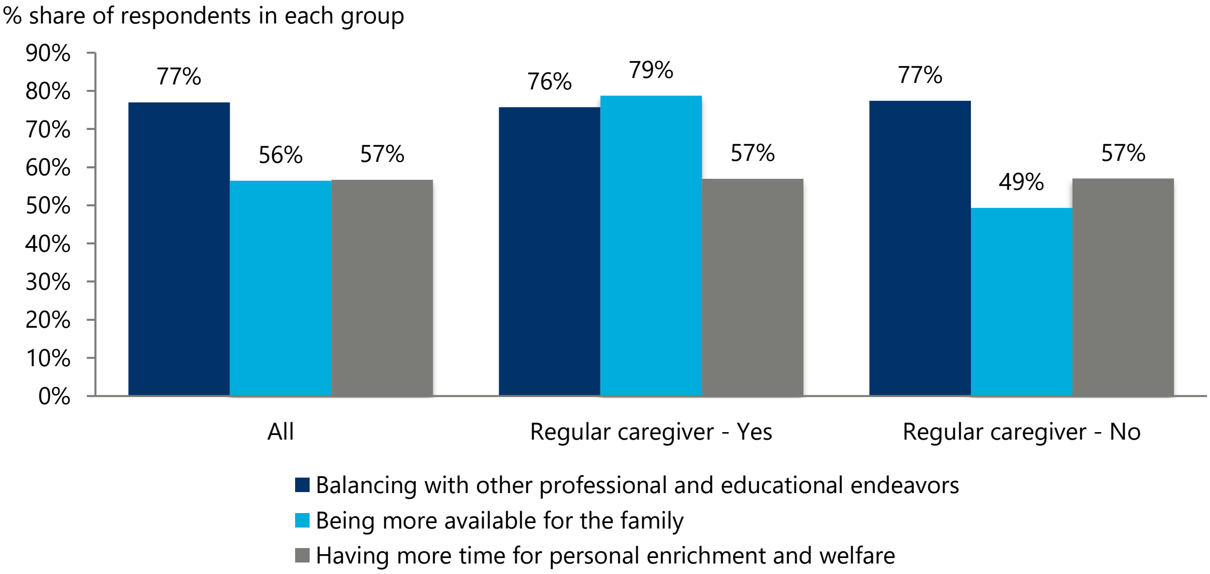
Sample: n = 831 (All), 54 (18 to 24), 209 (25 to 34), 282 (35 to 44), 189 (45 to 54), 97 (55+)
What is your age? (Single answer)
What are the most important aspects of having flexible working hours? (Multiple answer)
Source: Oxford Economics

Fig. 20. Important aspects of having flexible working arrangements, by gender



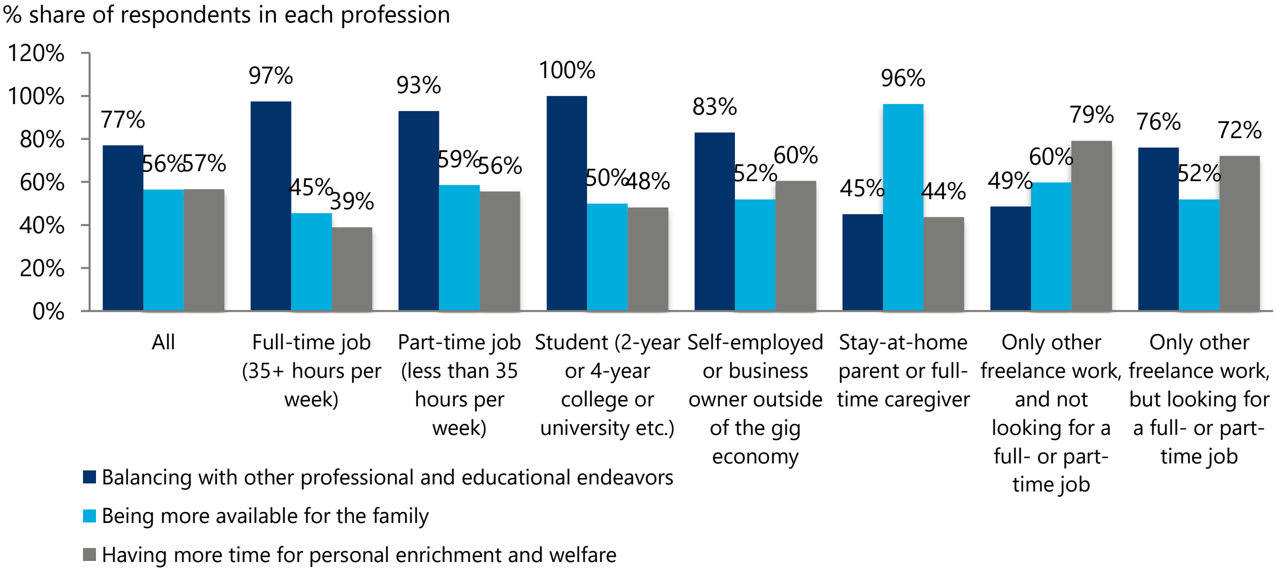
Sample: n = 831 (All), 325 (Male), 481 (Female)
What is your gender? (Single answer)
What are the most important aspects of having flexible working hours? (Multiple answer)
Source: Oxford Economics

Fig. 21. Important aspects of having flexible working arrangements, for regular caregivers



Sample: n = 831 (All), 202 (Regular caregiver - Yes), 598 (Regular caregiver - No)
What is your gender? (Single answer)
What are the most important aspects of having flexible working hours? (Multiple answer)
Source: Oxford Economics

Fig. 22. Important aspects of having flexible working arrangements, by profession



Sample: n = 831 (All), 198 (Full-time job), 99 (Part-time job). 56 (Student), 129 (Self-employed), 80 (Stay-at-home parent or full-time caregiver), 72 (Freelance and not looking for job), 129 (Freelance but looking for job)
What is the primary professional or educational activities do you engage in outside of your work as a contributor? (Single answer)
What are the most important aspects of having flexible working hours? (Multiple answer)
Source: Oxford Economics

ABOUT OXFORD ECONOMICS

Oxford Economics was founded in 1981 as a commercial venture with Oxford University's business college to provide economic forecasting and modelling to UK companies and financial institutions expanding abroad. Since then, we have become one of the world's foremost independent global advisory firms, providing reports, forecasts and analytical tools on more than 200 countries, 100 industries, and 8,000 cities and regions. Our best-in-class global economic and industry models and analytical tools give us an unparalleled ability to forecast external market trends and assess their economic, social and business impact.

Headquartered in Oxford, England, with regional centers in New York, London, Frankfurt, and Singapore, Oxford Economics has offices across the globe in Abu Dhabi, Belfast, Chicago, Dubai, Dublin, Hong Kong, Los Angeles, Mexico City, Milan, Paarl, Paris, Philadelphia, Sydney, Tokyo, and Toronto. We employ 450 staff, including more than 300 professional economists, industry experts, and business editors—one of the largest teams of macroeconomists and thought leadership specialists. Our global team is highly skilled in a full range of research techniques and thought leadership capabilities from econometric modeling, scenario framing, and economic impact analysis to market surveys, case studies, expert panels, and web analytics.

Oxford Economics is a key adviser to corporate, financial and government decision-makers and thought leaders. Our worldwide client base now comprises over 2,000 international organizations, including leading multinational companies and financial institutions; key government bodies and trade associations; and top universities, consultancies, and think tanks.

December 2025

Scale AI, Inc. provided the funding for this independent study. The findings of this research reflect the analysis of Oxford Economics. This report is confidential to Scale AI, Inc. and may not be published or distributed without their prior written permission.

The modeling and results presented here are based on information provided by third parties, upon which Oxford Economics has relied in producing its report and forecasts in good faith. Any subsequent revision or update of those data will affect the assessments and projections shown.

To discuss the report further please contact:

Vasilis Douzenis
Laurence Wilse-Samson

Oxford Economics
5 Hanover Sq, 8th Floor
New York, NY 10004
Tel: +1 646-786-1879

OXFORD ECONOMICS

Global headquarters
Oxford Economics Ltd
Abbey House
121 St Aldates
Oxford, OX1 1HB
UK
Tel: +44 (0)1865 268900

London
4 Millbank
London, SW1P 3JA
UK
Tel: +44 (0)203 910 8000

Frankfurt
Marienstr. 15
60329 Frankfurt am Main
Germany
Tel: +49 69 96 758 658

New York
5 Hanover Square, 8th Floor
New York, NY 10004
USA
Tel: +1 (646) 786 1879

Singapore
6 Battery Road
#38-05
Singapore 049909
Tel: +65 6850 0110

Europe, Middle East and Africa

Oxford
London
Belfast
Dublin
Frankfurt
Paris
Milan
Stockholm
Cape Town
Dubai

Americas

New York
Philadelphia
Boston
Chicago
Los Angeles
Toronto
Mexico City

Asia Pacific

Singapore
Hong Kong
Tokyo
Sydney

Email:
mailbox@oxfordeconomics.com

Website:
www.oxfordeconomics.com

Further contact details:
www.oxfordeconomics.com/about-us/worldwide-offices