The Business of AI Startups

James Bessen, Boston University
Stephen Michael Impink, HEC Paris
Lydia Reichensperger, Boston University
Robert Seamans, New York University

July 25, 2023

Abstract: Increased computing power and advanced machine learning (ML) techniques have accelerated the development and use of "artificial intelligence" (AI). Despite the increased commercial focus on AI, we have limited information about how AI products are developed by resource constrained startups. We collect survey data from startups producing commercial AI products to provide a glimpse in how these product impact labor across industries. This paper chronicles the last five years of AI startup survey data, documenting information about their business models, product, suppliers, partners, and customers. Our descriptive findings add to conversations on whether AI products enhance or replace human capabilities; which algorithms, data protections, and frameworks high-tech startups use to develop their AI products; and which geographies, customer-types, and industries are sales targets.

JEL codes: O33, J21, L10
Keywords: artificial intelligence, entrepreneurship, automation, labor, technology

1. Introduction

Breakthroughs in machine learning have led to the acceleration of commercial "artificial intelligence" (AI) development for a wide range of uses. The potential of these new technologies has prompted numerous assessments that these products will drastically alter the economy (Acemoglu and Restrepo, 2018; Frey and Osborne, 2017; Furman and Seamans, 2019; Felton et al., 2021). According to some commentators, we are entering a "Race Against the Machine" (Brynjolfsson and McAfee, 2011), a "jobless future" (Ford, 2015), a "Fourth Industrial Revolution" (Schwab 2016) and the "Second Machine Age" (Brynjolfsson and McAfee, 2014). Results from these studies differ in their estimation of AI's impact and their tone[1], yet there is consensus that AI development and using big data to train AI not a passing trend.

Much of this new literature on the economic consequences of AI has been based on estimated capabilities of the new technology, without accounting for possible difficulties in commercialization and technology adoption. For instance, the Frey and Osborne study (2017) is based on the assessments of a small panel of machine learning experts who were asked which of 70 occupations were "completely automatable" in 2013. Even if these assessments are correct, just because a job is automatable does not mean that it will be automated. Moreover, both theoretical models and historical evidence show that even when most of the labor of an industry is automated, employment can grow rapidly (Bessen, 2019).

If AI is likely to drastically change the economy in a decade or so, then we should see evidence of that at the actual businesses where AI is being applied today. That is, if 47% of jobs will soon be at risk from AI, then surely jobs will be at risk already at those firms using AI

[1] On the more pessimistic side estimates that 47% of jobs in the US will be "at risk of automation" in the next decade or so (Frey and Osborne 2017). On the less-pessimistic side, Bessen (2019) and Furman and Seamans (2019) discuss the potential for AI to augment labor's capabilities.

applications. A look at the cutting-edge applications of the technology provides a window into likely outcomes over the next decade or so.

To glimpse the impact of AI development, our paper reports on a global survey of startups developing and licensing commercial AI products. The survey questionnaire covers a wide range of topics, including questions about startups' markets and customers, their AI technologies and frameworks used in product development, their use and protection of training data, their product users and use cases, and their customers' demographics and industries. These findings paint a vivid picture of how startups pull together the needed resources to develop and monetize their AI applications. And that picture is at odds with alarmist predictions about the impact of AI. Rather than drastically reducing jobs, current data commercial AI applications are much more often about enhancing human capabilities than about reducing labor costs. While some occupations are experiencing job losses, others are experiencing job creation. The actual effect of AI on jobs varies by occupation (Felton et al., 2021), which our data reflects.

The survey provides a window into other topics in addition to how AI effects labor, including how AI startups address barriers to entering this nascent industry, how they protect customer data and respond to new data protection regulations such as the EU General Data Protection Regime (GDPR), and the extent to which they develop their own machine learning algorithms or rely on third party providers. While the survey establishes that there is a vibrant startup community of AI application developers and many new startups enter each year, it is possible that investments by more established firms may create barriers to entry in some markets. Respondents answer questions on data, hardware, software and firm size, which have previously been linked to entry barriers.

Though startups are constrained by the type of data they collect, in many cases lacking the diversity or scale of data collected by larger firms, 80 percent of survey respondents report using their customers' data to train AI. Customers proprietary data correlated with increased future performance, suggesting that the inability for startups to access barriers to scale from having less imitable data-resources. In the last several years, access to cheaper computing power (i.e., IT hardware) has made it easier to enter the industry and develop their products, as introductory offers from cloud provider and venture capital "spray and pray" investing have reduced the cost of IT (Ewens et al., 2019). Moreover, access to standardized tools and other resources from cloud suppliers may lead to time saving, enabling startup to focus on differentiating their product instead of building core app infrastructure (Impink, 2023). These more standardized tools may benefit performance without necessarily inhibiting innovation (Miric et al., 2022). Due to the availability of cloud IT and basic development tools, labor (Rock 2022) and proprietary data (Bessen et al, 2022) may be a key source of firm differentiation.

2. Prior research on how automation impacts labor and occupations

In addition to basic facts about the business of AI, this study provides evidence that bears on several major questions that have been raised in the literature, including the impact of AI on jobs and occupations, and the role of data as a critical resource for AI-enabled startups. The first question, noted above, concerns the impact of AI on jobs. Frey and Osborne (2017) base their estimate on a technical evaluation of "automatability," the technical feasibility of automation. The literature highlights several economic and technical reasons this might not be a sufficient metric for understanding the impact of AI on jobs. One reason is that automating some or even most tasks an occupation performs does not necessarily eliminate the occupation. Historically, most automation has been partial, such as the introduction of the assembly line in manufacturing (Rosenberg, 1963). Even though automation increased productivity, workers may still be cogs in the broader process (Chandler, 1962). Bessen (2016) finds that of 270 detailed occupations listed in the 1950 Census, only one can be described as having been eliminated due to automation, namely the job of elevator operator. Furthermore, partially automating a job can increase employment in that occupation or industry as well as decrease employment (Acemoglu and Restrepo, 2018). This is because automation tends to decrease prices, driving greater demand. When demand is elastic enough, greater demand will offset the labor-saving effect of automation. For example, during the 19th century, the textile industry was heavily automated, yet employment rose (Bessen, 2019). More recently, the automatic teller machine (ATM) automated some of the work of bank tellers, yet their employment grew as well (Bessen, 2016). Also, the effect of AI is not just to automate tasks which replace humans, but also to enhance human capabilities at both performing new tasks and old tasks more effectively.

In the recent empirical literature, (Arntz et al., 2017) modify the basic estimates of Frey and Osborne to account for the partial nature of automation. However, they do not consider the adoption of the technology or the labor demand effects that might cause automation to increase employment. The McKinsey Global Institute (Manyika, Chui et al., 2017; Manyika, Lund et al., 2017) attempts to take a range of hypothetical adoption rates into account in estimating impacts, and they also consider hypothetical labor demand effects. Accenture (2018) also estimates how the impact of AI on employment will vary depending on the role of different skills, and also predict where there will be deficits and surpluses of these skills on a geographic basis.

Several papers have looked at the impact of robots on employment, with differing results (Acemoglu and Restrepo 2017, Graetz and Michaels 2015, Dauth et al. 2018). Graetz and Michaels

(2015) use robot shipment data at the country, industry and year level from the International Federation of Robotics (IFR) and find that robots added an estimated 0.4 percentage points of annual GDP growth between 1993 and 2007 on average for the 17 countries in their sample (accounting for about one-tenth of GDP growth during this time period) while having a slightly positive but statistically noisy effect on employment.

In contrast, Acemoglu and Restrepo (2017) use the IFR data to examine the impact of the increase in industrial robot usage on regional U.S. labor markets between 1990 and 2007. These authors find that industrial robot adoption in the United States was negatively correlated with employment and wages during this time period—according to their estimates, each additional robot reduced employment by six workers and that one new robot per thousand workers reduced wages by 0.5 percent. The authors note that the effects are most pronounced in manufacturing, particularly in routine manual and blue-collar occupations, and for workers without a college degree.

Dauth et al. (2017) combines German labor market data with IFR robot shipment data and finds that while each additional industrial robot leads to the loss of two manufacturing jobs, enough new jobs are created in the service industry to offset and in some cases over-compensate for the negative employment effect in manufacturing. This finding echoes a broader study by Autor and Salomons (2018) which explores the general impact of productivity growth on employment, including both own industry and cross industry effects, and finds that productivity growth does not reduce employment in general, thanks in large part to positive spillovers into related sectors.

A second question we address is which occupations will be affected. Frey and Osborne (2017) argue that lower wage occupations will experience greater job losses than higher wage occupations. The McKinsey Global Institute projects, somewhat differently, that high wage occupations will grow while mid-wage occupations shrink. Kaplan (2015) posits that "automation is blind to the color of your collar" and many professions will be devastated. Susskind and Susskind (2015) argue that new technology will lead to the decline of the professions. Our survey provides some evidence about which occupations are growing and which are losing jobs in response to AI. Other recent papers that take the task- or ability-based approach include Brynjolfsson, Mitchell and Rock (2018) and Felten, Raj and Seamans (2021). Moreover, a large literature has explored the differential impact of technology on different groups of workers. This includes the literature on skill-biased technical change (Katz, 1999) and on job polarization (Goos et al., 2009; Autor, Katz, and Kearney 2008).

Next, a third question our study addresses are the extent of entry barriers into AI markets. Investment in AI is currently dominated by large firms, especially a few large tech firms with slack cloud compute resources – Amazon, Google, and Microsoft. Moreover, these firms also control generative AI models (BARD, ChatGPT, Dalle) and AI development frameworks (TensorFlow, PyTorch), and they have tons of customers that provide data to use their services and products. In particular, some observers, such as Stucke and Grunes (2016), argue that the combination of data and network effects creates substantial entry barriers in online markets. Others point to scale barriers stemming the scale, IT infrastructure, and data-resource wielded by the largest technology firms (Benzina, 2019; Khan 2017). For example, because Google has an enormous amount of search data, it might be hard for a new competitor to compete in the search engine market.

Others contend that data, by itself, is not likely to pose an entry barrier (Lambrecht and Tucker 2015, Sokol and Comerford 2016). It may be that Google's advantage comes not from the amount of data, but from the results of the product-focused experiments. Varian, Google's chief economist, and co-authors (2010) report that Google conducted 6,000 experiments on its search engine in 2008. Moreover, while large amounts of data are typically needed for machine learning, there may be diminishing returns to the amount of data beyond a certain point. For instance, Bajari, Amazon's chief economist, and co-authors (2019) find that increasing the number of online products that Amazon tracks does not significantly improve machine learning prediction accuracy after a certain point, implying that data quantity provides only a low barrier to entry. Amazon and other large technology companies benefit from the breadth and recency (Chiou and Tucker, 2017) of the data captured. For example, startups are more likely to sell a subset of the broader product offering of a larger company and to engage with small to medium size customers. In any case, our survey provides basic information about startups' access to data and circumstantial evidence on relative entry barriers in different industries.

Lastly, the fourth question this study addresses is how startups use data protections. Data protection and data privacy has become a flashpoint in the media, due in part to high profile data breaches such as at Equifax in 2017 and in part to high profile exposure of Facebook user's personal data to Cambridge Analytica in 2016 and 2017. Partially in response to these events, government regulators have instituted tighter rules on data protection. This has most notably manifested in Europe in the form of General Data Protection Regime (GDPR). There has been some concern that the increased data protection required under GDPR may constitute a barrier to entry for startups, and early research suggests that GDPR may in part be contributing to a slowdown in VC investment in European based startups (Jia et al., 2021). Additional evidence suggests that GDPR more negatively effects smaller web technologies firms than larger ones. For instance, Google gains significant share after GDPR was passed at the detriment of smaller firms (Peukert et al., 2022).

Our survey adds to this nascent literature by providing basic information about the data protections that AI startups have in place and allow us to compare the use of data protections across small vs large startups as well as European vs non-European based startups.

3. Survey and Sample

Sample selection

The main portion of our sample comes from the Crunchbase database of firms. Each year we select startups tagged as AI firms or described with the term ‘artificial intelligence’ or machine learning’. At the time of selection, these firms are listed as in-business and operational. We focus on English-language countries, where Crunchbase has greater coverage, and omit China. The list of startups selected are primarily based in developed countries, which the vast majority of firms are headquartered in the USA, Canada, UK, or the EU. In an attempt to have a more homogenous sample, we only include responses from startups that are less than ten years old and less than 500 employees. This enables us to focus on high growth startups, while weeding out mom and pop ventures and different business models.

In the first round of the survey, we also included startups from three additional sources: alumni of the Creative Destruction Lab, a startup incubator, who were identified as working with AI; a list of machine learning companies obtained from Philipp Hartmann and Joachim Henkel (Hartmann and Henkel, 2018); also, O’Reilly Media ran a notice of the survey in its AI newsletter, providing a link to the online survey. Once a startup is added to our sample list, we send them a survey every year. Moreover, we reach out to multiple email addresses at the firm. For instance, we would email the individual who responded from the firm in the prior year and any other email aliases listed on Crunchbase (i.e., ‘ceo@aistartup.com’ or ‘leadershipteam@aistartup.com’).

Over the last five years, we have reached out to 5,254 startups. For each individual survey, our response rate is around 8%. In total, firms responding at least once our response rate is 22%.

Survey rounds

Survey 1 was administered online using Qualtrics from in the summer of 2018. Surveys 2-4 occurred the first few months of 2020, 2021, and 2023. Survey 5 occurred in the early spring of 2023. They survey consists of around 36 questions in a typical year. These questions (excluding name and email questions) are pre-tested on academics and on roughly a half dozen firms with interviews. In survey 2 we added questions on GDPR; in survey 3 we added questions on ethical AI development; and in survey 4 we added question on the governance of AI. Survey 5 asks the same questions as survey 4, with the goal of building a panel of responses throughout time. After each round of the survey, we receive feedback from startups that helps us adapt the survey in the next round. However, the core questions on firm demographics, customers, and core product development that we report on have remained identical in all five survey rounds.

Across all surveys, we have received 1,178 responses from 917 unique startups. Figure 1 shows the number of responses by survey round. More than 160 startups have responded to the survey more than one time in the last five years; two startups have responded all five times.[2] Figure 2 reports then number of times that each startup has responded to the survey. This will be particularly important in future projects as we build out a time-series of responses to determine how those responses change over time.

[2] Expect your gold star to be delivered via mail, if you responded all five times.

Figure 1 – Responding startups (count) by survey round

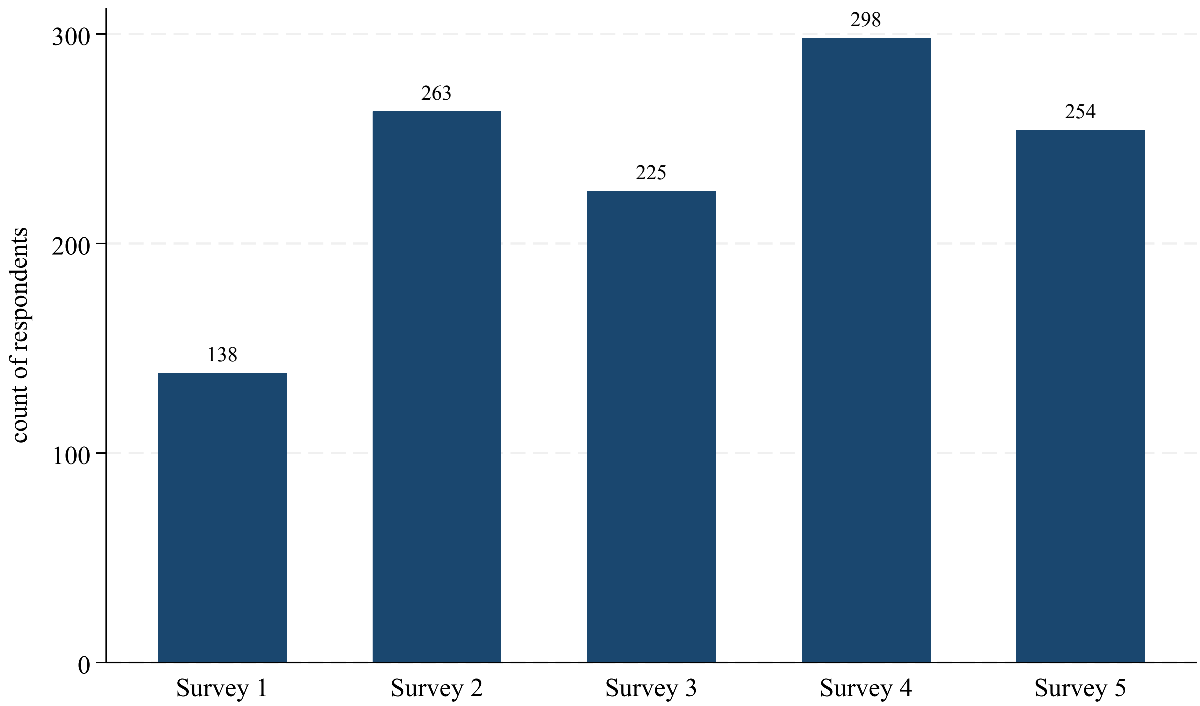
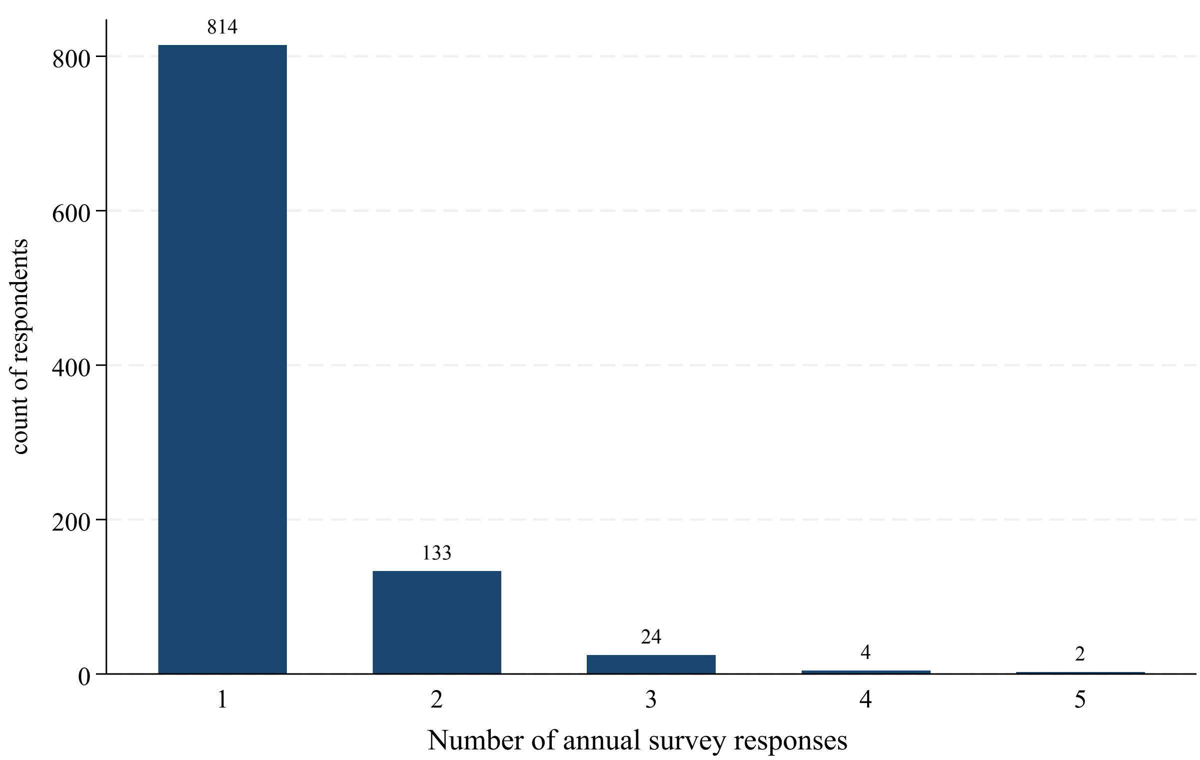


Figure 2 – Startups responding in multiple survey rounds



Characteristics of respondents

We show the representativeness of subsets of our data in several papers (Bessen et al., 2022, Bessen et al, 2023). Generally, we find firms based in California are less likely to respond. In certain subsets of the data, we find that firms that are less than two years old are less likely to respond, potentially suggestion that more nascent ventures may not have full time employees.

Table 3 reports the basic summary statistics for responding firms in all surveys. The startups in our sample are quite small with around half of the firms having less than 10 employees (Figure 4) and existing for three or fewer years (Figure 5). More than half of the responding startups are located in North America, for multiple reasons. First, there is a higher concentration of startups in California, New York, Massachusetts, and Toronto. However, we are no blind to the fact that Crunchbase has coverage issues in the developing and non-English language countries and the survey was sponsored and run by two US universities. Upcoming survey will be sponsored by HEC Paris in addition to NYU Stern and Boston University to increase response rates withing Europe. Figure 6 shows the count of survey responses by region.

Table 3 – Responding startup demographics (all survey)

	Mean	SD	Min	Max
Age	4.55	2.25	1	10
Emp. (Count)	26.78	40.97	1	500
Small (< 11 emp.)	0.52	0.5	0	1
Healthcare Ind.	0.12	0.32	0	1
Financial Ind.	0.06	0.23	0	1
US	0.44	0.5	0	1
UK	0.05	0.21	0	1
France	0.03	0.17	0	1
Germany	0.04	0.19	0	1
Canada	0.05	0.21	0	1
Founder MBA	0.2	0.4	0	1
Founder Big Tech	0.04	0.19	0	1
Founder Female	0.08	0.27	0	1
Responses	1178			

Figure 4 – Age distribution

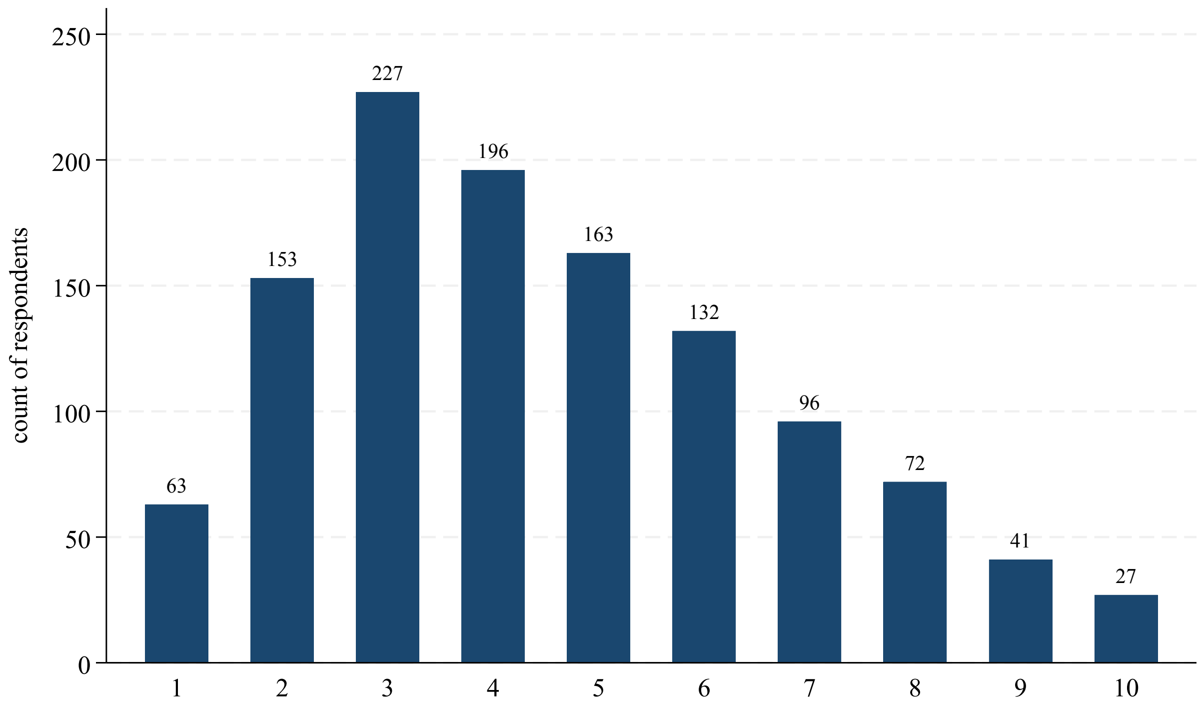


Figure 5 – Employee count distribution

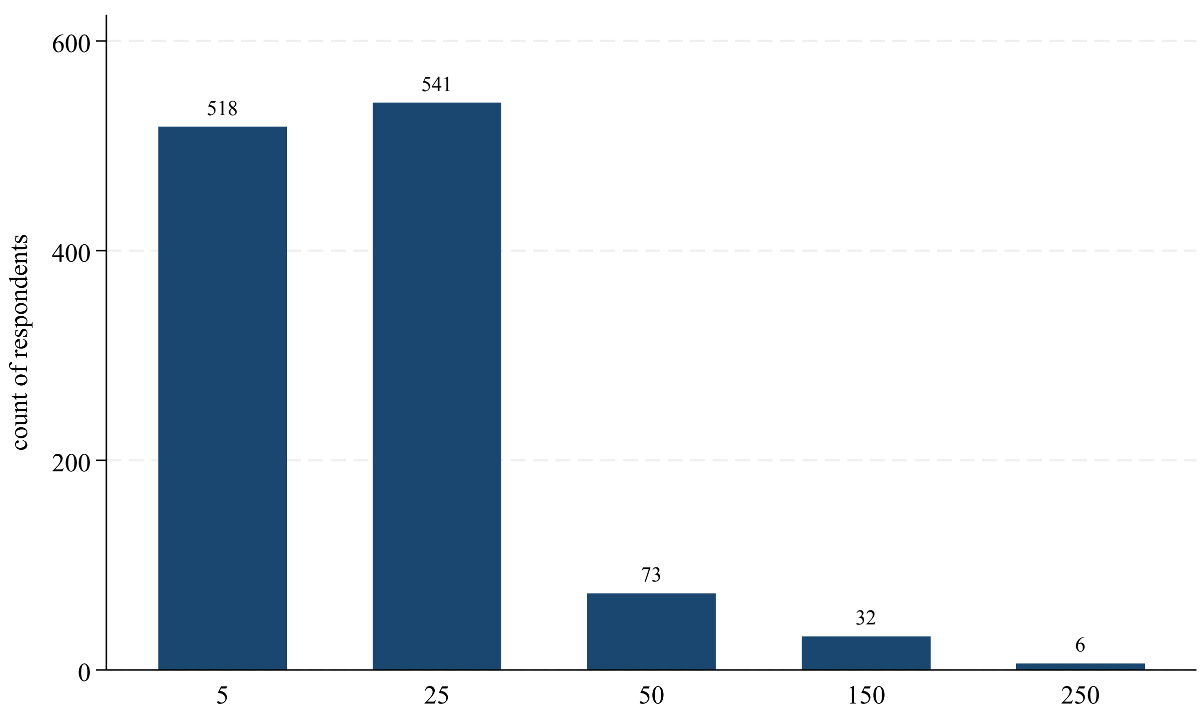
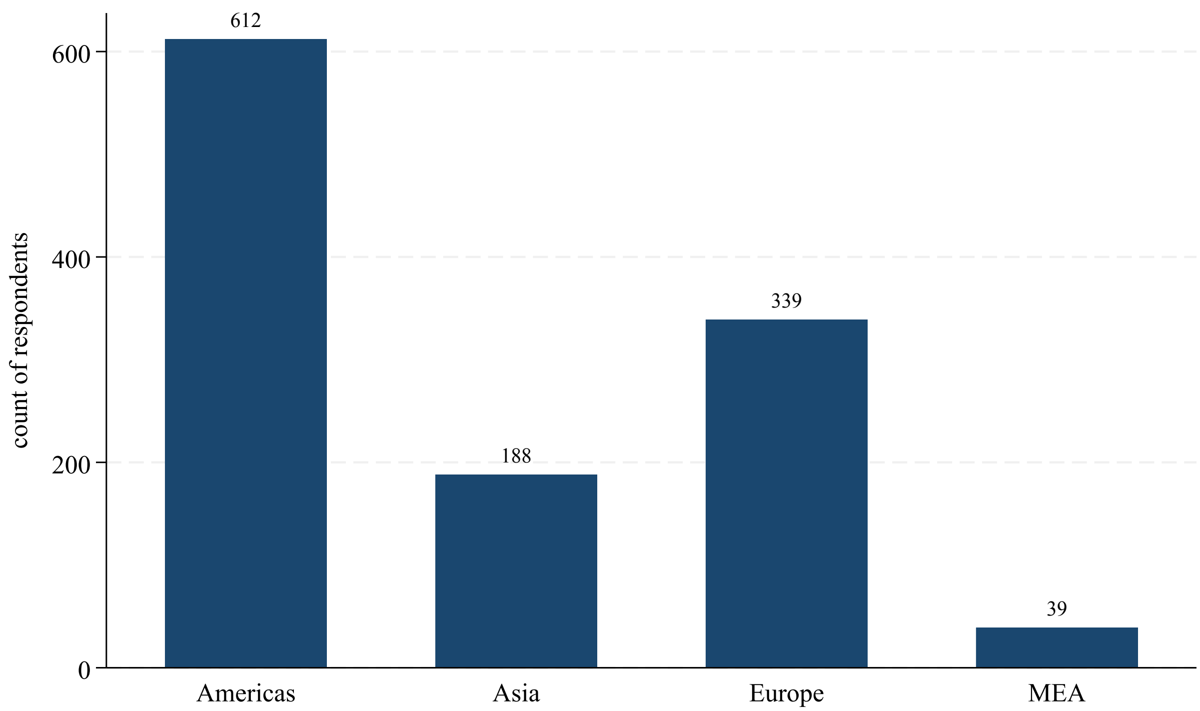


Figure 6 – Region



Survey respondents disproportionately sell to mid-sized firms. Figure 7 shows the share of customers in each size class from the survey. Almost half of the startups sell to firms in the middle category with 51-1000 employees, but these firms make up only 26% of the market as measured by employment size from the US Census Longitudinal Business Database (LBD). The responding startups also sell to both smaller and larger customers, but proportionately less than would be expected given the distribution of firms in developed countries. Most of the firms are shipping product (71%), but 20% are in beta testing, and 9% are pre-beta (Figure 9).

Figure 7 – Customers' mean employment size

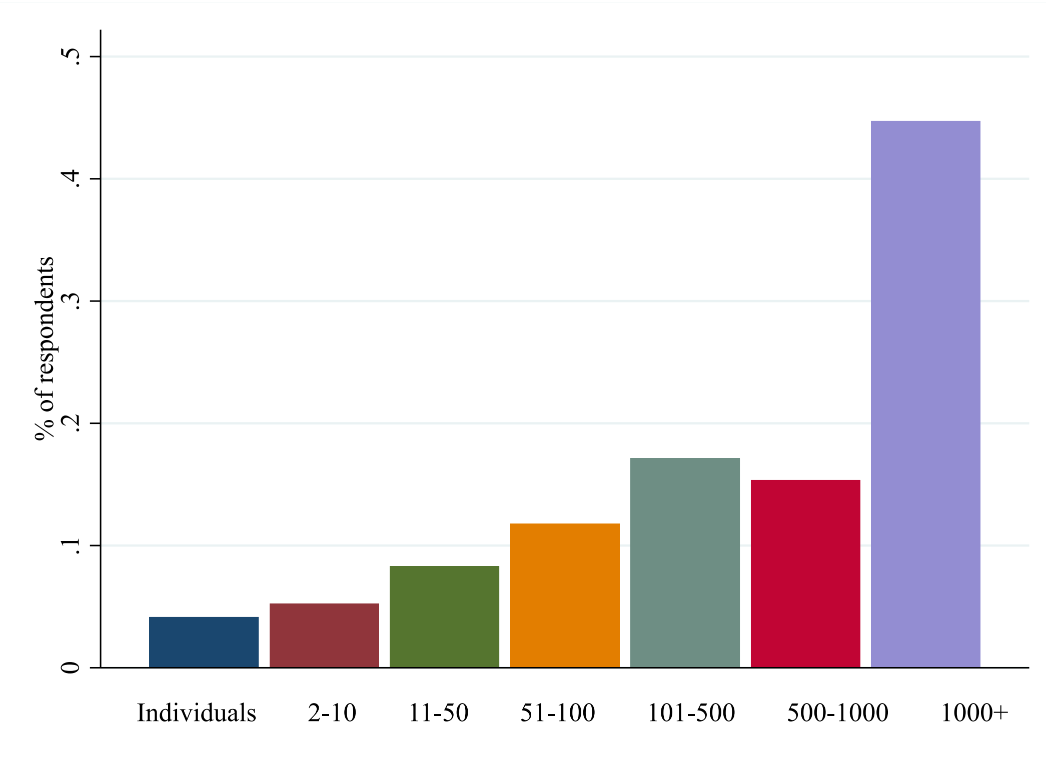


Figure 8 – Product stage by region

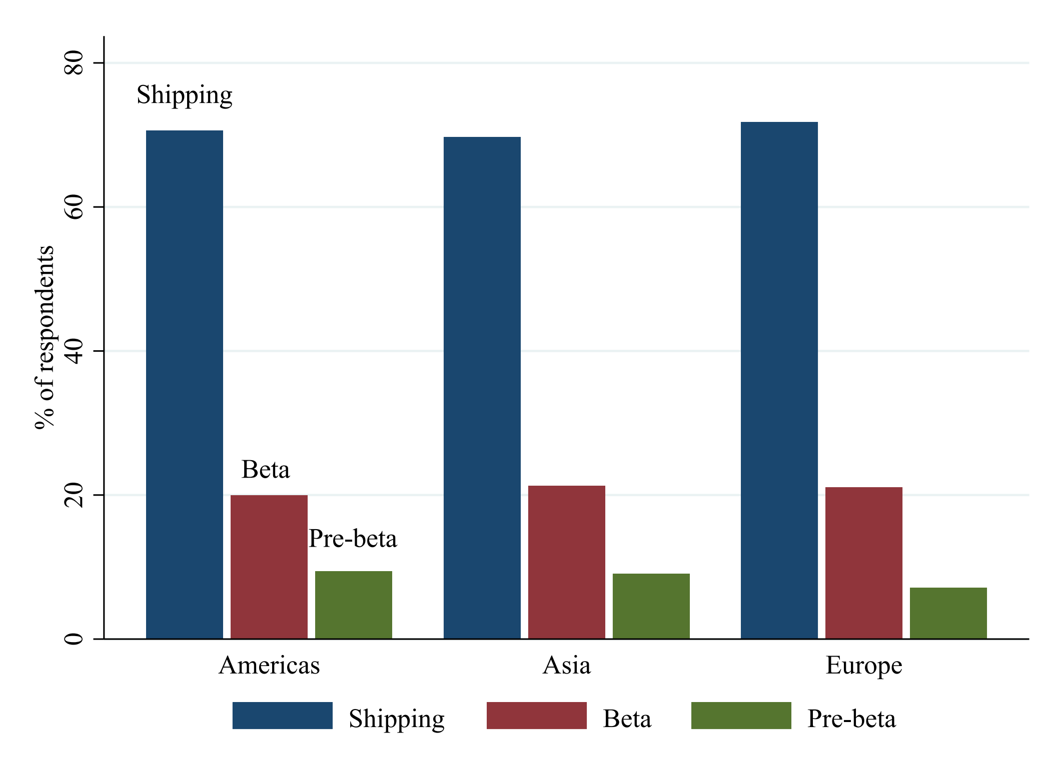


Table 4 reports summary statistics by survey round. Startups in the nascent industry are become a bit older, from 3.7 years old on average in 2018 to 5.3 years on in 2023. The mix of responses by geography shows a decline in the percent of respondents from Canada and an increase in the percent of responses in Europe. Our initial list of startups included data from Creative Destruction Labs (University of Toronto) in Canada, which may explain of this change. However, we also had specific data from the TUM accelerator program in Germany in the first round of the survey but show increases in Germany instead of declines.

The number of respondents with an MBA is also declining over time, suggesting earlier AI startups had a higher proportion of management education. This potentially suggests entry barriers were higher initially, required greater credentialling to kickstart the venture. Lastly, about 10% of respondents are female, and this remains stable over time.

Table 9 – Responding startup demographics (by survey)

	Mean				
	Survey 1	Survey 2	Survey 3	Survey 4	Survey 5
Age	3.69	4.18	4.14	5	5.26
Emp. (Count)	36.32	26.1	23.88	26.92	24.69
Small (< 11 emp.)	0.46	0.49	0.61	0.51	0.54
Healthcare Ind.	0.09	0.14	0.15	0.1	0.09
Financial Ind.	0.02	0.08	0.07	0.05	0.05
US	0.46	0.46	0.41	0.46	0.42
UK	0.08	0.06	0.1	0.01	0.01
France	0.03	0.02	0.02	0.03	0.04
Germany	0.02	0.02	0.05	0.05	0.04
Canada	0.12	0.04	0.06	0.03	0.03
Founder MBA	0.25	0.27	0.21	0.17	0.12
Founder Big Tech	0.04	0.05	0.06	0.04	0.01
Founder Female	0.01	0.09	0.07	0.1	0.1
Responses	138	263	225	298	254

4. Findings

Heterogenous impact by industry

The survey respondents sell to customers in all major industry sectors as well as to individual consumers (Figure 10). In this sense, machine learning is a general-purpose technology that can be used in a variety of applications across a variety of industries (Cockburn et al 2018).

This broad distribution across industries suggests that the startup environment is healthy and many opportunities for entry exist. However, some industries may have relatively higher entry barriers than others. If there are no entry barriers, then both large and small firms invest based on the size of economic potential in an industry. This means that without entry barriers, the distribution of AI development spending should look the same for smaller firms as it does for larger ones. If, on the other hand, startups face significant entry barriers in an industry, then large firms would spend proportionately more than smaller firms in industries with entry barriers, holding the anticipated economic potential across industries constant. Comparing startup funding in different industries relative to total investment in AI in those industries helps identify those industries with possible entry barriers.

Although our survey only includes startups and not large established firms, the McKinsey Global Institute conducted a survey of top managers at large and small firms (Bughin et al. 2017). By comparing estimates from the two studies for different industries, we can identify differences in relative funding that might signal entry barriers. The McKinsey study reports the current adoption of AI for select industries, presented as a weighted mean by firm size (Bughin et al. 2017: Exhibit 4). Using this as a proxy for total AI investment in each industry for 2018, Figure 11 shows the share of startup participating in each industry from our survey compared with the share of total investment from the McKinsey survey.

The chart shows that the share of investment going to startups and to all firms is roughly the same in many industries, but in three industries total investment—likely reflecting large firm investment—is disproportionately large: consumer packaged goods, transportation and logistics, and high tech and telecommunications. On the other hand, it appears that startup investments in AI applications serving retail and finance are disproportionately larger, likely reflecting venture capital priority areas.

Figure 10 – Industry

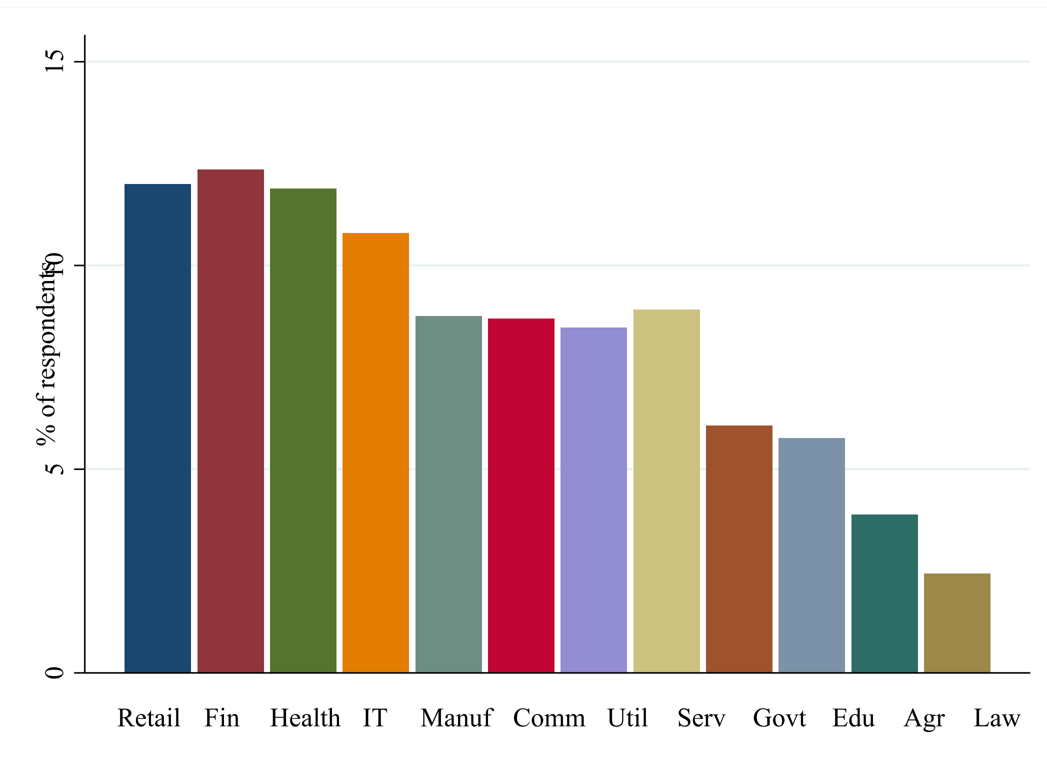
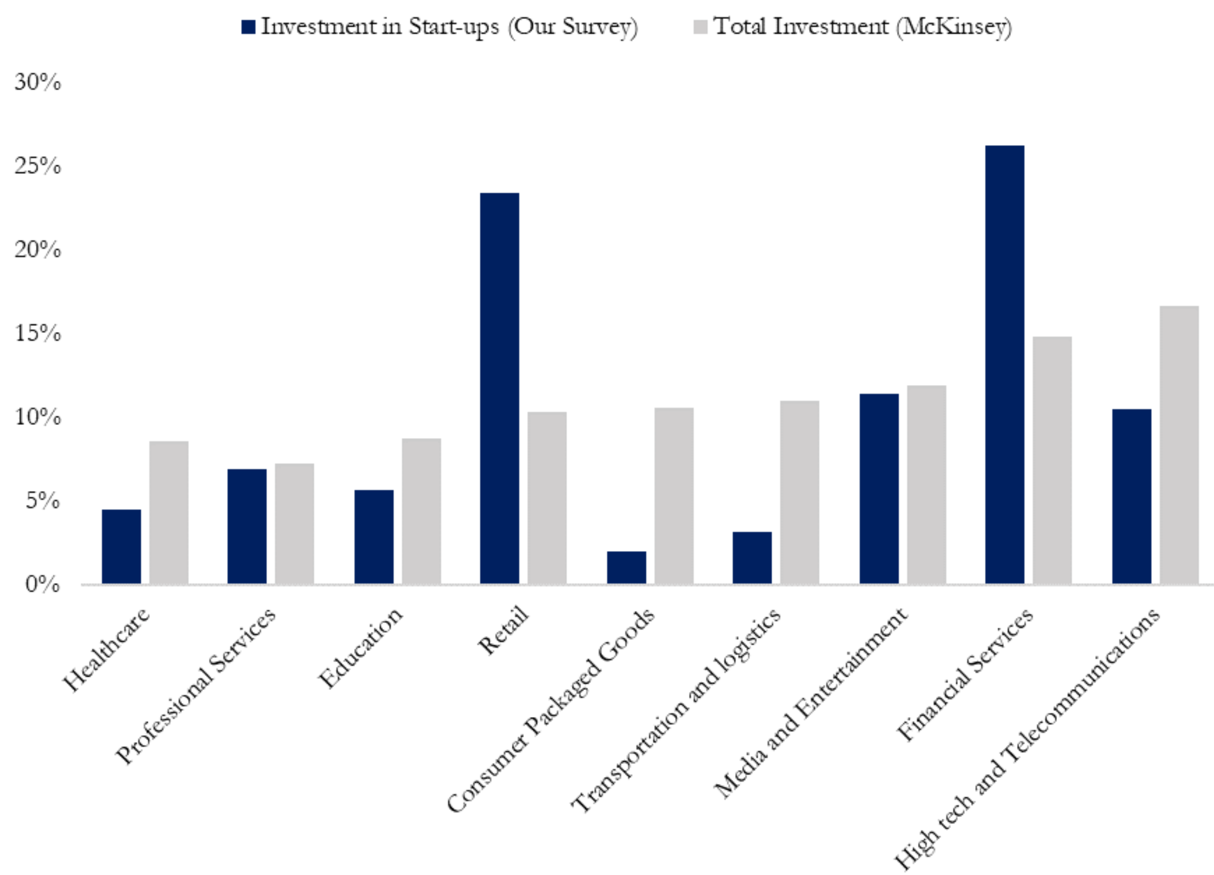


Figure 11 – Industry Share of Funds: Survey respondents vs. McKinsey projections (2018 only)



Do AI products augment productivity or replace workers?

Finally, to understand the nature of the marketing appeal—the "unique selling proposition"—the survey asked firms to rate various benefits that their products provided to customers. Each benefit is rated on a 5-point scale from "strongly disagree" to "strongly agree" (or "not applicable"). Figure 12 shows the share of responses ranked "strongly agree."

The three most frequent benefits reported are capabilities to make predictions or decisions, to manage and understand data, and to create new and improved products and services. These answers provide a gauge for understanding how much AI enhances human capabilities and how much, instead, it tends to replace them. Of the survey responses, 53% strongly agree that their products automate routine tasks and 48% strongly agree that their products reduce labor costs. That is, AI appears to be about enhancing human capabilities, not just replacing them. AI may very well eliminate certain types of occupations or asymmetrically affect industries (something we explore further below), but because facets of the technology appear to augment human capabilities, it might well be associated with increased employment and wages for at least some occupations.

Moreover, the replacement of workers appears to be significantly concentrated in the certain sectors of the economy. We find that, for firms that sell to customers in agriculture, manufacturing, utilities and transportation industries, 74% strongly agree that their products benefit customers by automating routine tasks and/or reducing labor costs. For firms that do not sell to those industries, only 46% strongly agree. This difference across industry groups is highly significant (t-test probability value of 0.000).

Figure 13 reports the share of firms reporting whether each occupational group at their customer firms is a user or not. White collar occupations are listed most frequently, including professionals (58%), managers (49%), and sales and marketing occupations (35%). Occupations that include clerical, administrative, manual or service work are less likely to be listed. In this regard, the distribution of users of AI across occupations is very similar to the distribution of computer users across occupations (Bessen 2016).

Figure 12 – Strongly agree (%): AI product augments labor or reduced labor costs

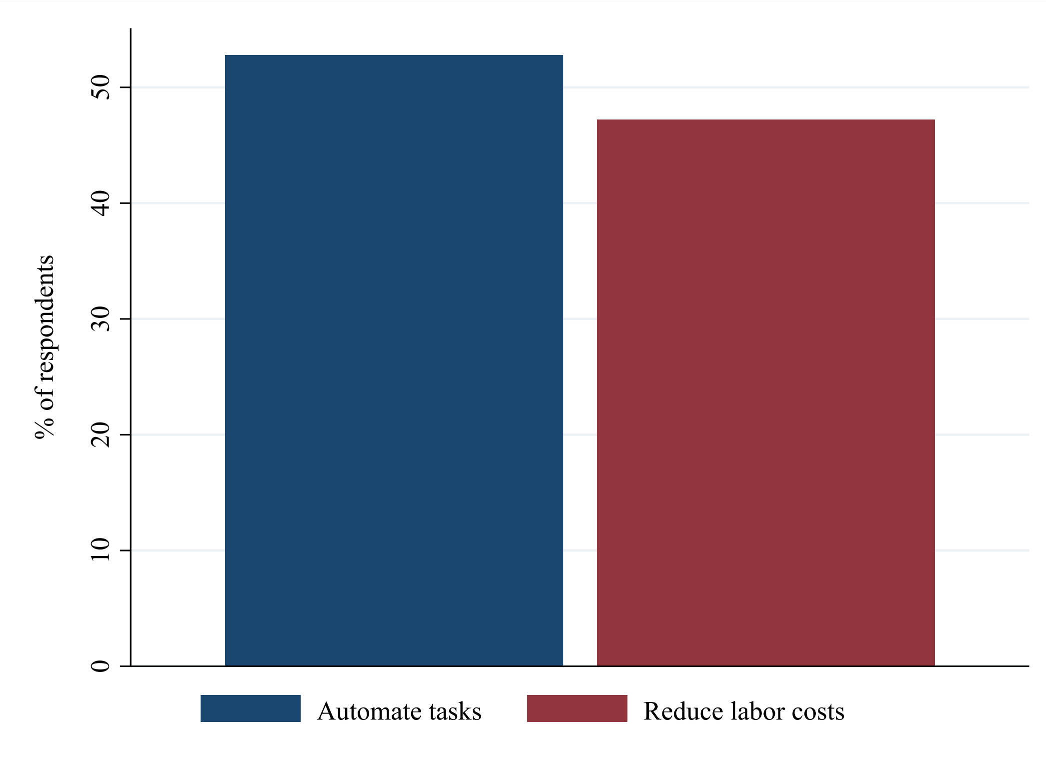
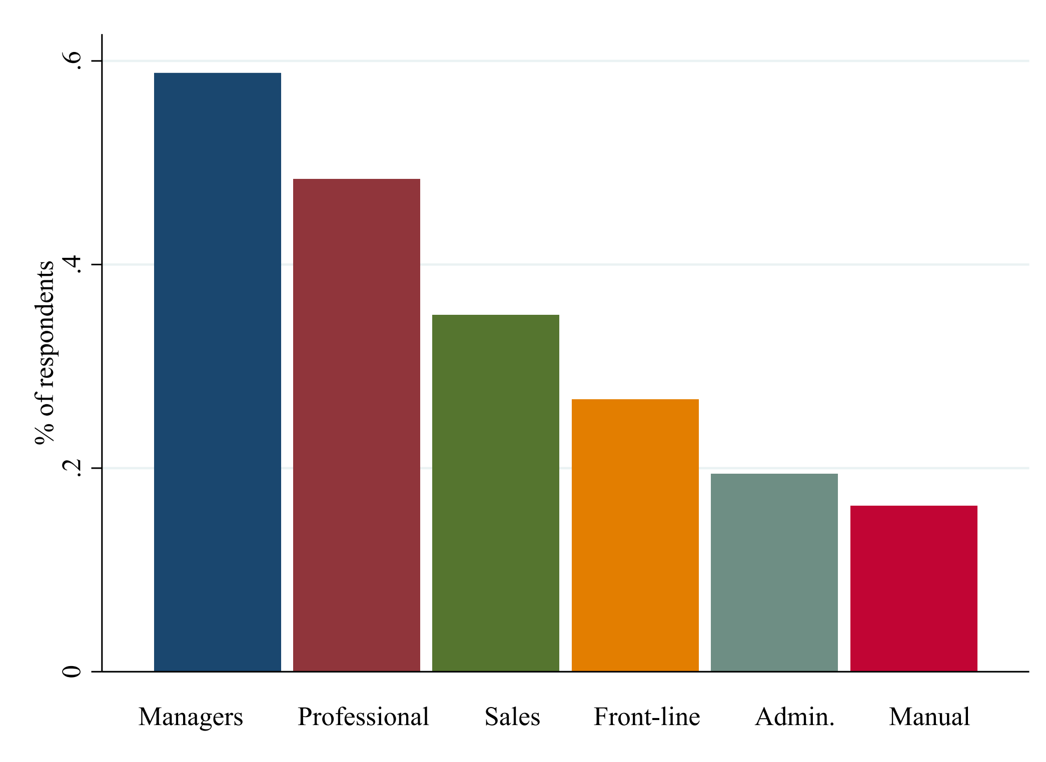


Figure 13 – Types of workers impacted by the startups AI product



AI products appear to require a level of skill to the extent that the occupations involved have skill requirements. But does the use of these products require STEM skills or specialized training? The survey responses suggest that in most cases specialized computer skills or specific training demands are modest. Only 10% of firms require users to have expert coding or data skills. 59% require general familiarity with computers and the remainder require no special skills at all.

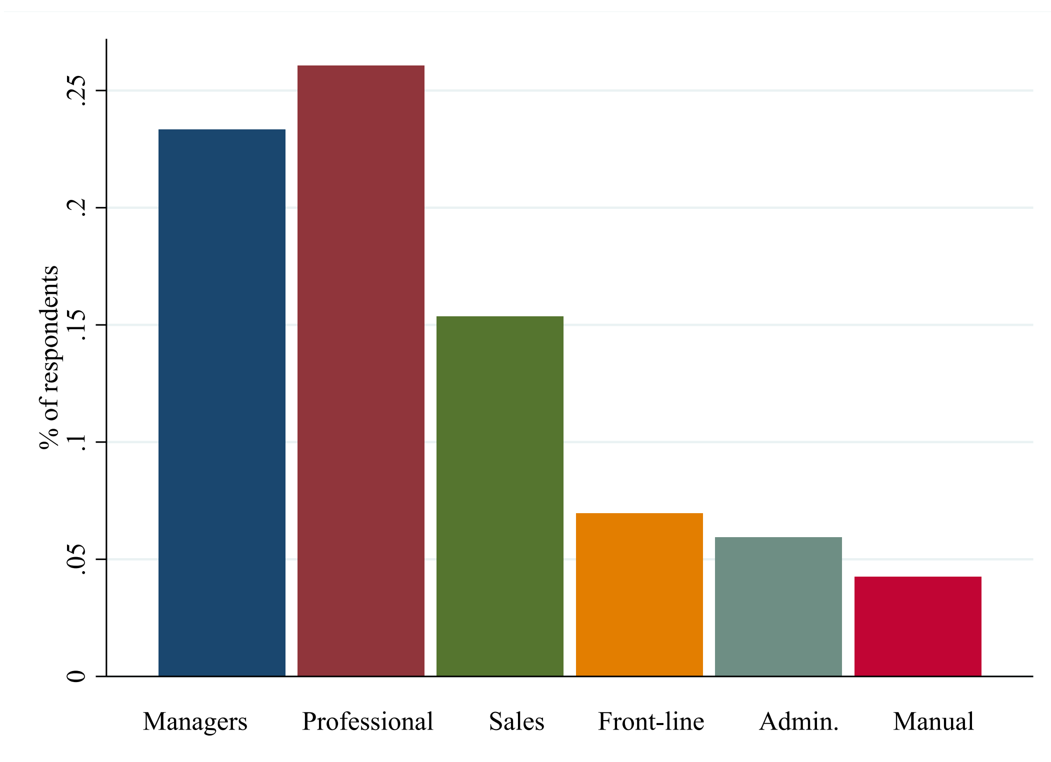
In terms of the training provided to users, 86% of firms report this training takes one week or less. 12% report that 2 to 4 weeks of training is required and 2% report up to 3 months of training is required. Fully 44% of the firms provide no training to their customers. Of those firms that do provide training, 70% provide onsite training classes, 32% provide offsite training classes, and 47% provide online courses.

Augment productivity or replace workers?

The analysis of customer benefits above suggests that most firms are oriented to enhancing customer capabilities rather than reducing customer labor costs, especially in the broad service sector. This suggests that AI might have some job-creating potential, especially for the professional, managerial, sales and marketing occupations that tend to be the users of these products. On the other hand, AI also reduces labor costs for some customers. Moreover, these effects might differ across occupations because use of AI differs dramatically across occupations. Figure 14 reports the share of firms that expect their products to create jobs or eliminate jobs in different occupational groups at their customers, listed in the same order as in Figure 13.

AI is not all about destroying jobs. These findings suggest that occupations using AI may benefit form augmentation of their capabilities (i.e., increased productivity). At the same time, many jobs will be eliminated, especially in the three occupational groups that use AI relatively less. In many cases, jobs will be created in some occupations and jobs will be destroyed in other occupations at the same affected firms. Of the firms that responded to this question, almost half (46%) reported that their products both create jobs and destroy jobs at customer firms. 26% report only creating jobs and 28% report only eliminating jobs. It is possible, of course, that survey respondents, perhaps sensitive to publicity about job losses, shaded their answers to reflect better job outcomes than is actually the case. Nevertheless, Figure 13 reveals dramatic relative differences between occupational groups. Professionals, managers, and quantitative sales workers may benefit from AI more than administrative, frontline, and manual workers.

Figure 13 – Types of workers augmented by AI products (versus being replaced)



Technology, frameworks and relationships with Big Tech

In building their products, responding startups use many different types of algorithms, AI development frameworks, and cloud tools to develop their products. The most commonly used technology is natural language understanding and text analysis (63% of firms), followed by natural language classification and decision management (both at 56%). These are followed by visual recognition, including image, face and video (45% of firms) and sentiment/emotion analysis (43% of firms). Other technologies are used by smaller percentages of firms.

Overall, many more firms develop their own software for the most commonly used technologies rather than purchase them from an external vendor. Only in two areas do firms rely more on outside vendors: speech recognition with 19% using external products while 13% develop their own, and natural language translation, with 17% using external software and 13% develop their own.

Regarding the type of algorithm used (Figure 14), 65% of reporting firms use neural networks including recurrent, convolutional, and generative adversarial neural networks. The next most common methods are clustering algorithms (46% of firms) and Bayesian or other methods of probabilistic inference (41% of firms). Other methods are used less frequently. Most startups use TensorFlow or PyTorch (Figure 15). However, the selection of the AI development framework depends on the fit of that framework with the algorithm, training data, and product. For instance, Figure 16 shows that startups using neural networks or ensemble learning, commonly discussed as more sophisticated algorithms, are more likely to use PyTorch and Google ML than TensorFlow and Keras.

Figure 14 – Commonly used algorithms

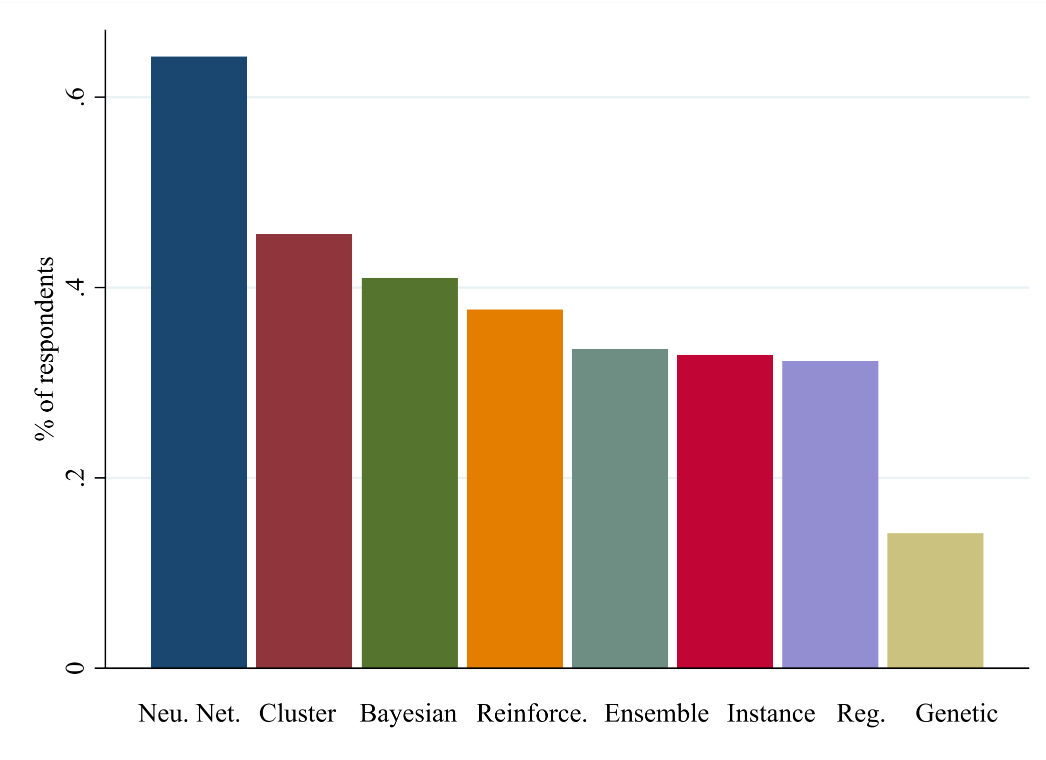


Figure 15 – Commonly used AI frameworks

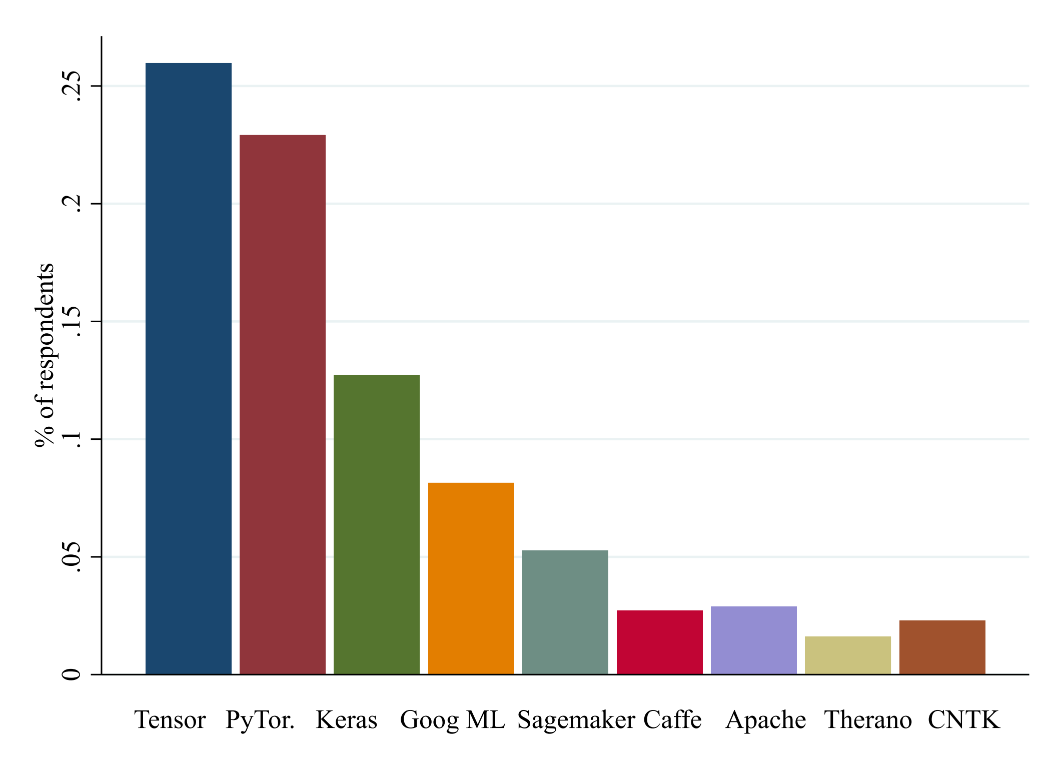
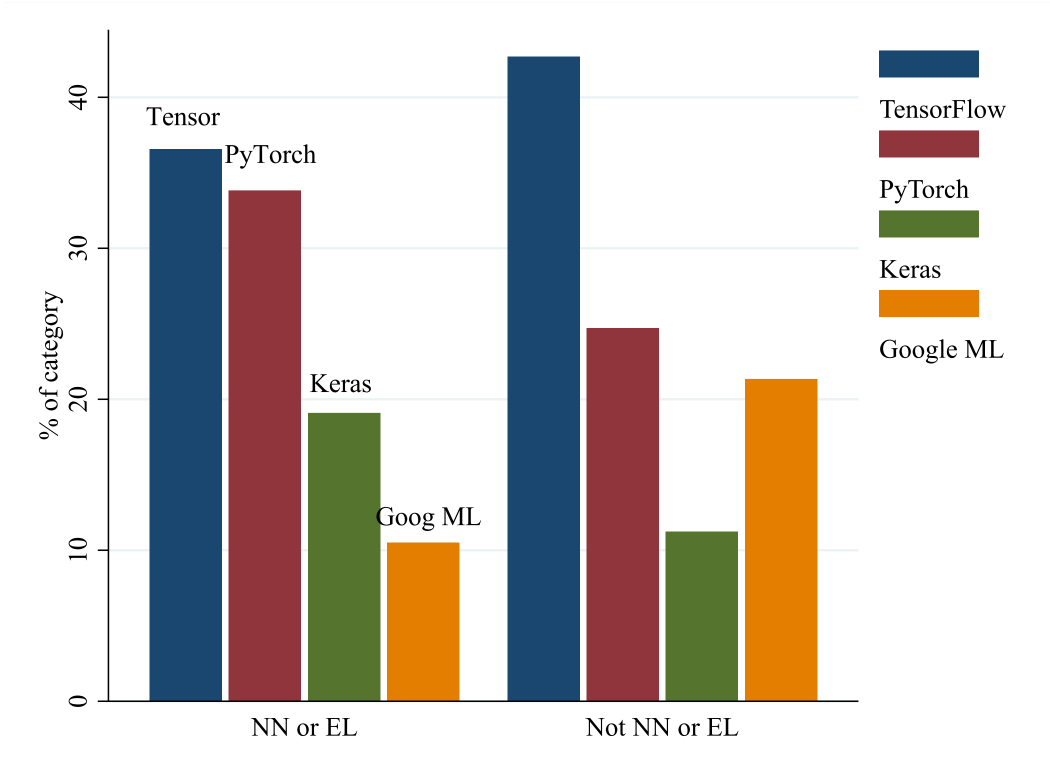


Figure 16 – Commonly used AI frameworks by most popular algorithms



Almost all the startups report that their products are cloud-based (97%) although 33% of the firms additionally provide software on premises. Only 3% of the firms provide software on premises only. Most of the firms (68%) provide a commercial application using AI. Some use the AI in their own products and services (43%) and 12% provide developer tools for AI applications (these are not mutually exclusive categories).

Thus, while startups use a wide variety of technologies to perform a variety of functions, a typical firm uses Amazon AWS cloud platform services, neural networks, and either TensorFlow or PyTorch to develop their AI product

Figure 17 – On Premise versus cloud IT by region

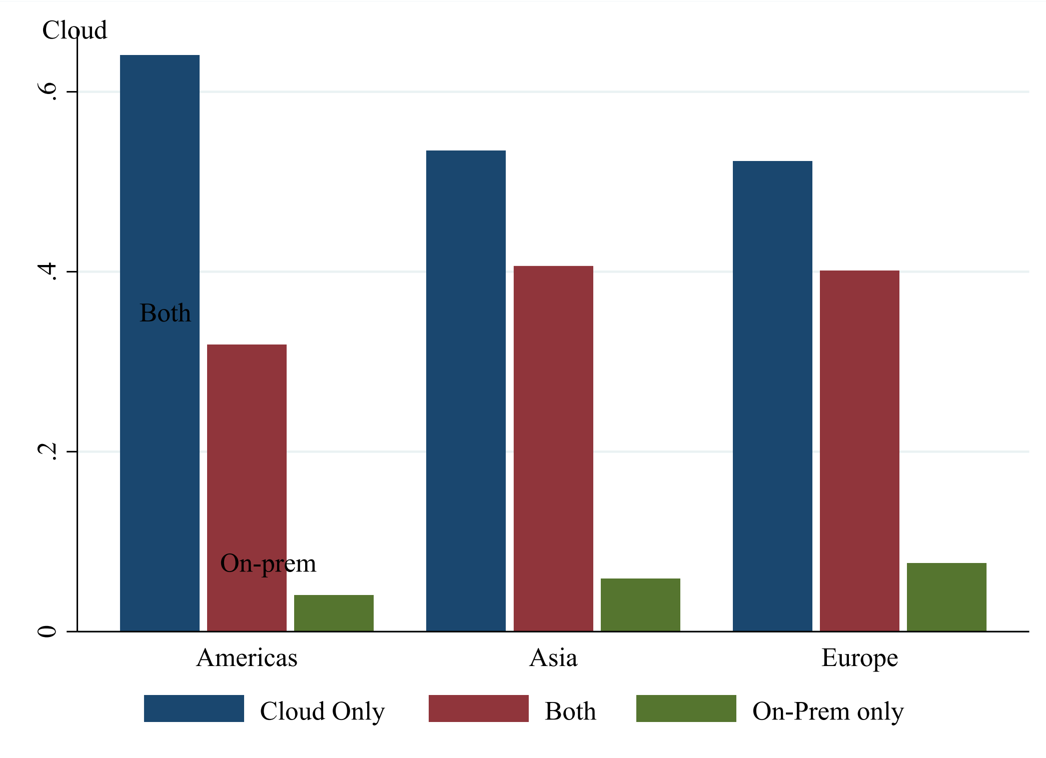


Figure 18 – Cloud IT provider by region

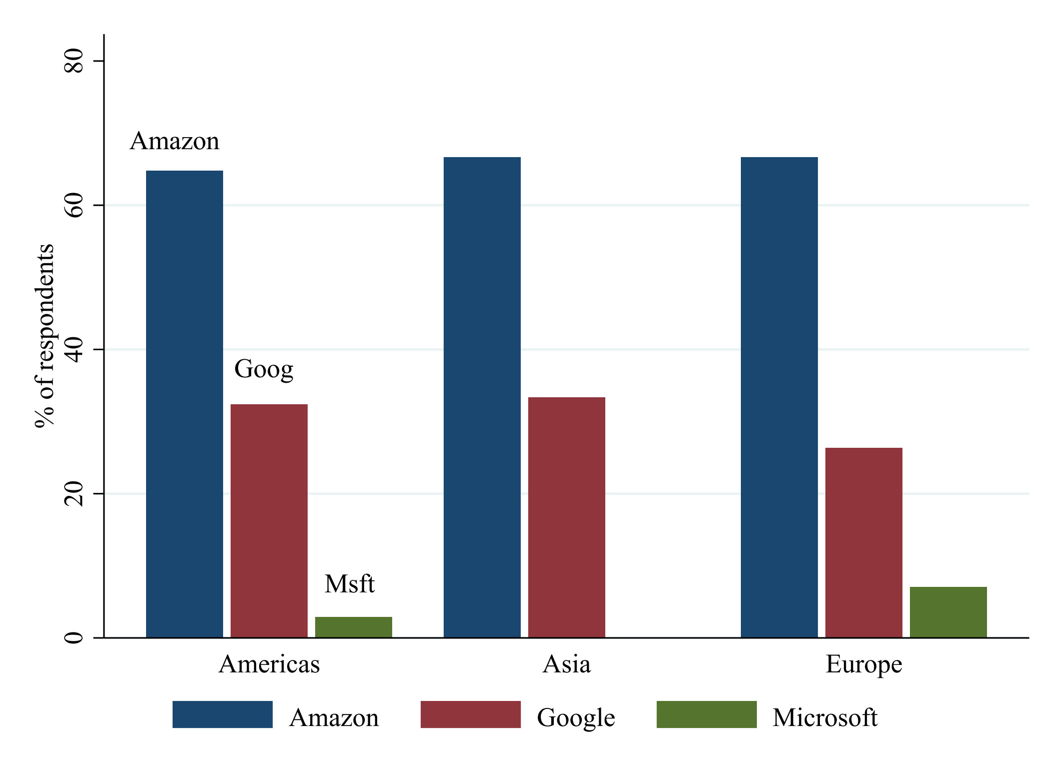


Figure 19 – Cloud IT provider by firm size

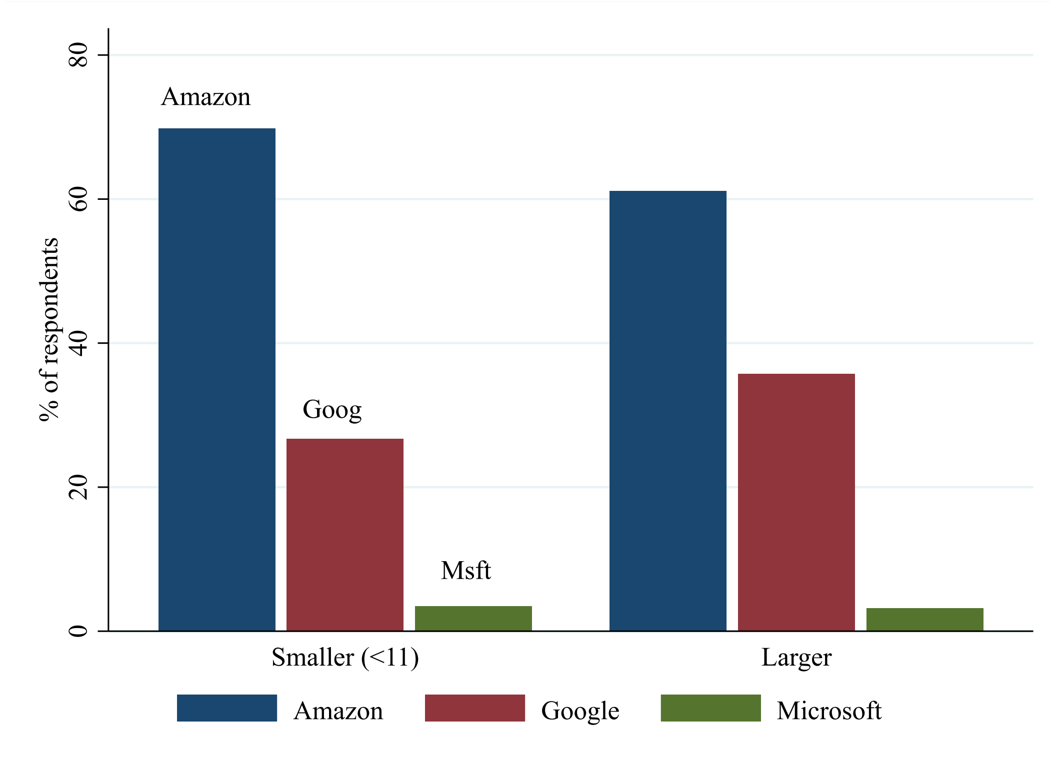
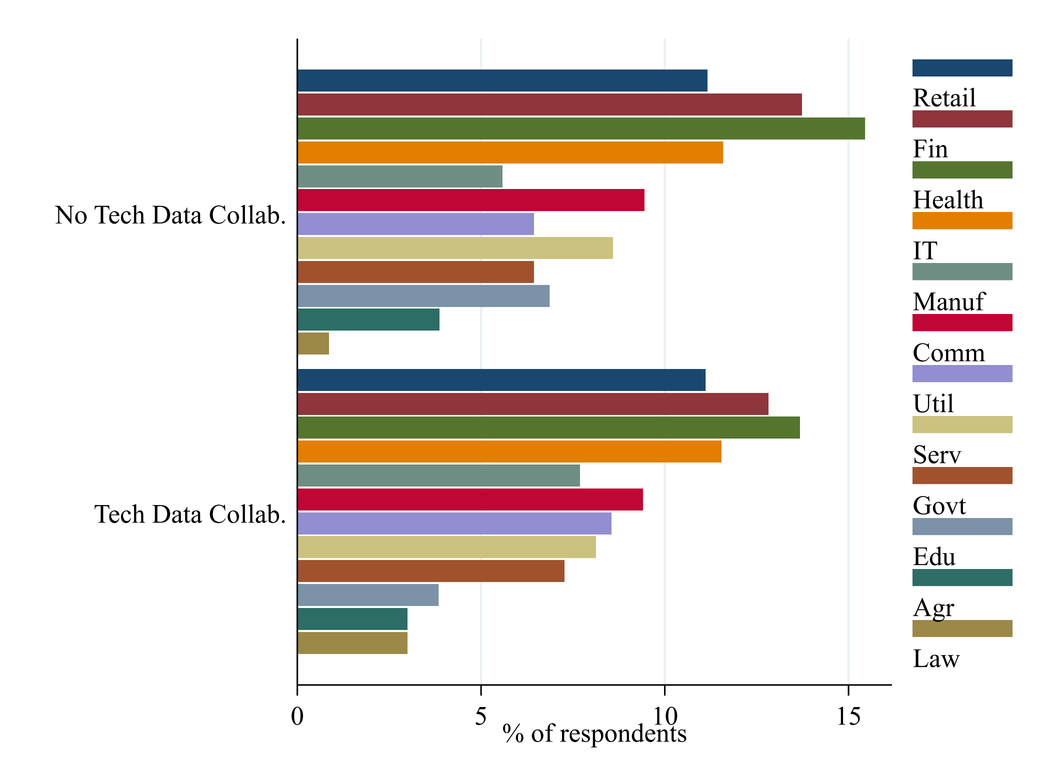


Figure 20 – Startup data collaboration with Big Tech by industry



Data and data protections

Startups also use a wide variety of data. 57% of the firms use unstructured text, 44% use transaction data, 38% use images, 37% use administrative data or other structured records. A smaller share of firms use audio, video, or other types of data. These data are mainly used to train algorithms. Consequently, the algorithms are re-trained as more data accumulates. Roughly a quarter of firms report refreshing their models daily, weekly, or monthly each. 13% of firms report having models that are not refreshed with new data.

Startup firms generally use other people's data. The most common source of data is from customers. 80% of the startup firms report using customer data, including data about their customer's customers and users as well as other data. 63% use data from third parties, including government data, data scraped from the Internet, and public benchmarking data. 51% of the firms report using their own proprietary data. Most of this use is in combination with data from other sources; only 6% of firms rely only on their own data.

To protect their access to data, startup firms who use customer data retain secondary reuse rights 52% of the time. To control the use of proprietary data between the firm and its customers, 83% of the firms use legal contracts that specify data uses. Additionally, firms use a variety of technical means to protect and control data access, including de-identification, encryption, passwords, access logs, and application program interfaces.

Only 22% of firms report that the EU's General Data Protection Regulation (GDPR) has impacted sales and marketing to non-EU countries. That figure is 27% for firms headquartered in Europe, excluding the United Kingdom. Given that the GDPR went into effect during our survey period, these figures might change as firms have more experience with the regulation. The survey respondents were asked to select all types of data protection used. As reported in Figure 22 , across all types of data protection, startup firms with customers in the EU report using data protections more intensively than startup firms without customers in the EU. This could reflect the impact of the GDPR and also different customer sensitivities.

There are some differences in data protection across firm size. Startup firms with ten or fewer employees represent close to 40% of the total survey respondents. These small startup firms are less likely to use legal contracts, de-identification, data encryption and password protected access, as reported in Figure 8. However small startup firms are equally likely to use logged access and application program interface as large startups.

Figure 21 – Types of job augmented by AI products (versus being replaced)

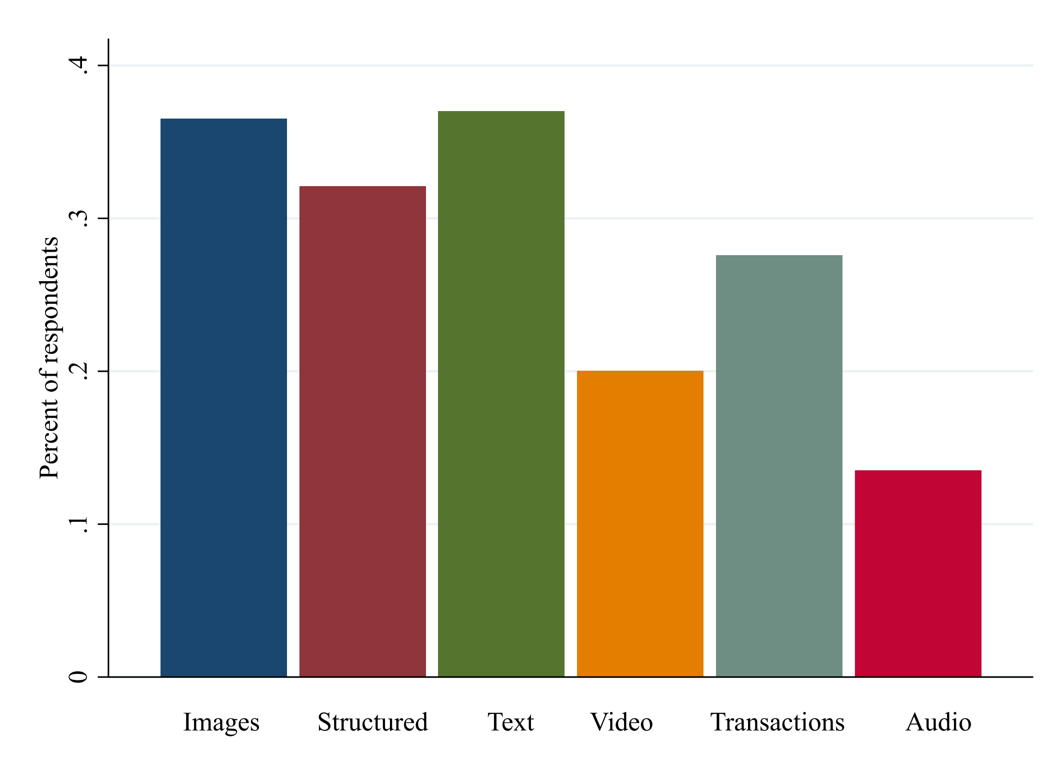
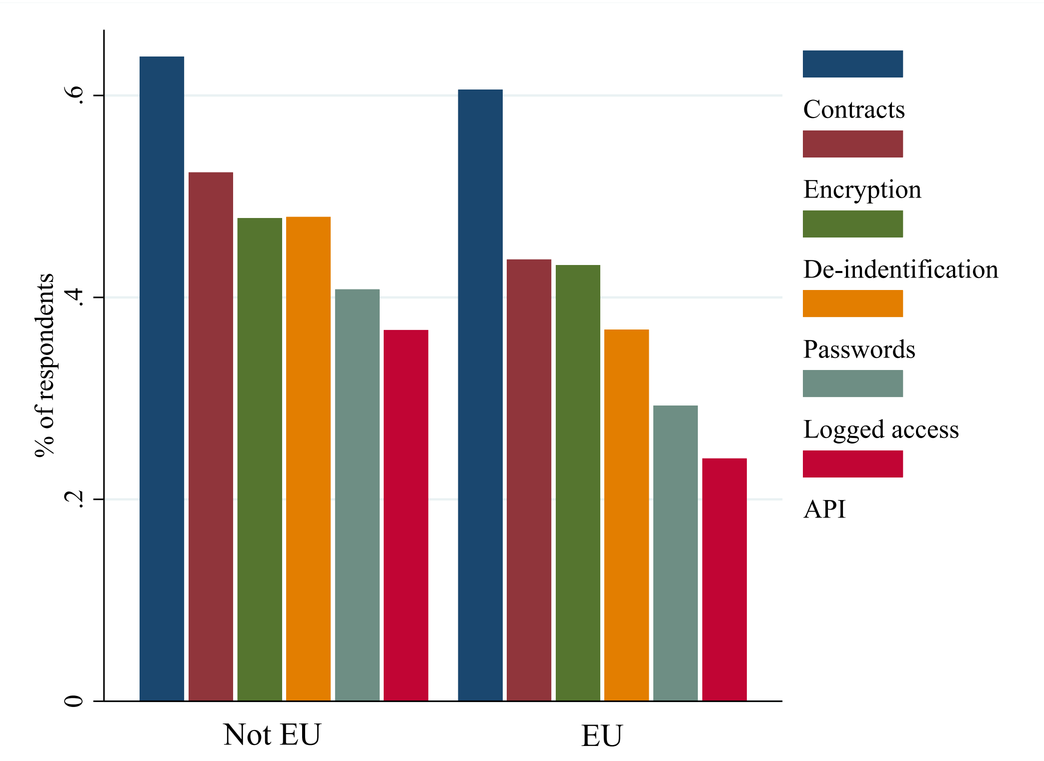


Figure 22 – Types of job augmented by AI products (versus being replaced)



The attached appendix includes data for the survey question on ethics. Bessen et al. (2023) uses responses from the 2021 and 2022 to examine the relationship between AI ethics policies, pro-ethics actions, and venture funding. This appendix also includes data from the 2023 survey.

5. Discussion

AI and Jobs

The evidence above in Figure 6 suggests that the benefits of AI are more about enhancing human capabilities than about replacing humans. This suggests that the technology often makes human workers more valuable rather than less, possibly raising labor demand for professionals, managers and sales and marketing personnel. The evidence in Figure 10 suggests that this is exactly what is happening. AI is associated with increasing labor demand for some occupations, decreasing demand for others. For the customers of these startups, AI appears to be less about eliminating labor overall and more about shifting work from some occupations to others.

This shift is illustrated by an account given by one of the firms interviewed in the survey pre-test. The firm's product automated the retrieval of contact information for prospective sales contacts from text files. This eliminated clerical work for the people who had manually searched for contact information, but because it was able to generate more sales prospects quickly, it also created jobs for sales and marketing personnel. This pattern of shifting work appears quite similar to the pattern observed for computer use where professional and managerial occupations appeared to grow while clerical and administrative occupations in the same industries shrank (Bessen 2016).

Entry Barriers

Commentators have identified several factors that might make it difficult for startups or for their customers to compete using AI. Yet the survey shows that these factors do not pose entry barriers for many startups in many markets. 80% of startups use customer data and 63% use data available from third parties, including publicly available data. While data might pose a barrier to entry in some markets, like search, where large amounts of diverse data are needed, there are clearly many markets where it does not.

Hardware. Almost all startups provide their products over the cloud, and evidence shows that small firms are able to access and utilize cloud computing effectively (Jin and McElheran 2017). Consequently, access to large computing hardware does not seem to be an issue for startups.

Most of the startups in our sample develop their own software for most applications, suggesting that skilled developers and basic software tools are available to them. Anecdotally, many software tools are available for free under Open Source licenses, such as TensorFlow.

In many cases, large firms may have an advantage because of economies of scale or network effects. In particular, if the fixed costs of developing a new AI application are large, then big firms can develop their own, but medium and small sized firms may need to rely on commercial application developers. These developers effectively spread the fixed costs over multiple customers. The survey finds that, indeed, startups sell disproportionately to mid-sized firms, suggesting that a) these mid-sized customers can compete using AI purchased from startups, and b) the startups can provide a more affordable solution than if the customers developed their own AI applications.

Data protection regulation has been described as a barrier that raises the cost of entry for startups. We do find that data protections are higher for startups based in Europe, and large firms appear more likely than small firms to have data protections in place. But on the other hand small firms and non-European firms do have a range of data protections in place as well.

This evidence of the entry of startups into some markets is encouraging, however, it is also possible that competition, entry, and innovation could all further be improved. Moreover, Figure 5 suggests that there are some industries that might have a substantial entry barriers. In particular, in consumer packaged goods, transportation and logistics, and high tech and telecommunications, large firms appear to invest disproportionately more than startups in AI development. The disparity in transportation and logistics might be related to the apparently large investments needed to credibly enter the market for self-driving cars or the competitive power of the largest firms in those industries. The disparity in high tech might reflect the importance of network effects in markets for online commerce and advertising or the disproportionate ownership of IP by the larger firms.

6. Conclusion

Frey and Osborne (2017) predict that 47% of US jobs will be at risk of automation in the next decade or so. If so many jobs were really at risk, we would see evidence of substantial job loss associated with emerging AI applications today. This survey makes clear that is not the case. The commercial AI applications offered by startups today are more about enhancing human capabilities than they are about replacing humans. Only half of the firms strongly agreed that labor cost reduction was a benefit to customers. And survey respondents replied that their customers are using AI to create jobs in certain occupations about as often as they use it to eliminate jobs. As a group, their applications more often prompt customers to increase employment in managerial, professional and sales and marketing roles—contrary to some predictions—and to decrease employment in service, clerical and manual jobs.

It is quite possible that AI might have different employment effects in the more distant future. It is also possible that large firms might have a different impact on employment than the startups in this survey. Nevertheless, the evidence tempers concern about mass unemployment or disemployment of professionals. AI appears to be associated with a shift in work from some occupation to others, meaning some people will lose jobs and other jobs will be created. However, the new jobs may require major new skills, requiring workers to make major investments and perhaps to endure difficult transitions. Thus, it is important to consider whether current labor market policies are adequate to address these potential demands (Furman and Seamans, 2019). The changing nature of demand for occupations and skills might well be highly disruptive even if there is not mass unemployment.

Finally, the survey documents the broad scope of AI applications being offered by startups, across a variety of industries. While the business for AI startups appears robust overall, there is also some evidence that startups may face barriers to entry in some major industries. The extent of these barriers, their nature, and possible policy remedies is a topic for future research. The survey also suggests that startups use a variety of data protection mechanisms, and that startups with EU customers are even more likely to engage in data protection. However, it is too early to assess how GDPR is affecting AI startups, and this is another topic for future research.

References

Accenture (2018). “It’s Learning, Just Not as We Know It.” https://www.accenture.com/us-en/insights/future-workforce/transforming-learning.

Acemoglu, D., & Restrepo, P. (2018). Artificial intelligence, automation, and work. In The economics of artificial intelligence: An agenda (pp. 197-236). University of Chicago Press.

Arntz, M., et al. (2017). "Revisiting the Risk of Automation." *Economics Letters* 159: 157-160.

Autor, D., & Salomons, A. (2018). Is automation labor-displacing? Productivity growth, employment, and the labor share (No. w24871). National Bureau of Economic Research.

Autor, D. H., Katz, L. F., & Kearney, M. S. (2008). Trends in US wage inequality: Revising the revisionists. *The Review of Economics and Statistics*, 90(2), 300-323.

Bajari, P., Chernozhukov, V., Hortaçsu, A., & Suzuki, J. (2019). The impact of big data on firm performance: An empirical investigation. In *AEA Papers and Proceedings* (Vol. 109, pp. 33-37).

Benzina, K. (2019). Cloud Infrastructure-as-a-Service as an Essential Facility: Market Structure, Competition, and the Need for Industry and Regulatory Solutions. *Berkeley Tech*. LJ, 34, 119.

Bessen, J. (2016). "How Computer Automation Affects Occupations: Technology, Jobs, and Skills." Boston University School of Law Working Paper No. 15-49.

Bessen, J. (2019). Automation and jobs: When technology boosts employment. *Economic Policy*, 34(100), 589-626.

Bessen, J. E., Impink, S. M., & Seamans, R. (2023). The Role of Ethical Principles in AI Startups. Available at SSRN 4378280.

Bessen, J., Impink, S. M., Reichensperger, L., & Seamans, R. (2022). The role of data for AI startup growth. *Research Policy*, 51(5), 104513.

Brynjolfsson, E., & McAfee, A. (2014). The second machine age: Work, progress, and prosperity in a time of brilliant technologies. WW Norton & Company.

Brynjolfsson, E., Mitchell, T., & Rock, D. (2018). What can machines learn and what does it mean for occupations and the economy?. In *AEA Papers and Proceedings* (Vol. 108, pp. 43-47).

Chandler, A. D. (1962). Strategy and structure: Chapters in the history of the industrial empire. Cambridge Mass.

Chiou, L., & Tucker, C. (2017). Search engines and data retention: Implications for privacy and antitrust (No. w23815). National Bureau of Economic Research.

Dauth, W., Findeisen, S., Suedekum, J., & Woessner, N. (2018). Adjusting to robots: Worker-level evidence. Opportunity and Inclusive Growth Institute Working Papers, 13.

Ewens, M., Nanda, R., & Rhodes-Kropf, M. (2018). Cost of experimentation and the evolution of venture capital. *Journal of Financial Economics*, 128(3), 422-442.

Felten, E., Raj, M., & Seamans, R. (2021). Occupational, industry, and geographic exposure to artificial intelligence: A novel dataset and its potential uses. *Strategic Management Journal*, 42(12), 2195-2217.

Ford, M. (2015). The rise of the robots: Technology and the threat of mass unemployment. International *Journal of HRD Practice Policy and Research*, 111.

Frey, C. B. and M. A. Osborne (2017). "The future of employment: How susceptible are jobs to computerisation?" *Technological Forecasting and Social Change,* 114: 254-280.

Furman, J., & Seamans, R. (2019). AI and the Economy. *Innovation Policy and the Economy*, 19(1), 161-191.

Goos, M., Manning, Salomans (2009). "Job Polarization in Europe." *American Economic Review,* Vol. 99: 58-63.

Graetz, G., & Michaels, G. (2017). Is modern technology responsible for jobless recoveries?. *American Economic Review*, 107(5), 168-173.

Impink, S. M. (2023). Both sides of cloud: More standardized tools and more diverse data analytics. Available at SSRN 4200675.

Jia, J., Jin, G. Z., & Wagman, L. (2021). The short-run effects of the general data protection regulation on technology venture investment. *Marketing Science*, 40(4), 661-684.

Kaplan, J. (2015). Humans Need Not Apply: A Guide to Wealth & Work in the Age of Artificial Intelligence. Yale University Press.

Katz, L. F. (1999). Changes in the wage structure and earnings inequality. *In Handbook of Labor Economics*, (Vol. 3, pp. 1463-1555). Elsevier.

Khan, L. M. (2017). Amazon’s antitrust paradox.

Lambrecht, N. and C. Tucker (2017). "Can Big Data Protect a Firm from Competition." *CPI Antitrust Chronicle*.

Manyika, J., Chui, M., Miremadi, M., Bughin, J., George, K., Willmott, P., & Dewhurst, M. (2017). A future that works: AI, automation, employment, and productivity. McKinsey Global Institute Research, Tech. Rep, 60, 1-135.

Miric, M., Ozalp, H., & Yilmaz, E. D. (2023). Trade-offs to using standardized tools: Innovation enablers or creativity constraints?. Strategic Management Journal, 44(4), 909-942.

Manyika, J., Lund, S., Chui, M., Bughin, J., Woetzel, J., Batra, P., ... & Sanghvi, S. (2017). Jobs lost, jobs gained: What the future of work will mean for jobs, skills, and wages.

Peukert, C., Bechtold, S., Batikas, M., & Kretschmer, T. (2022). Regulatory spillovers and data governance: Evidence from the GDPR. *Marketing Science*, 41(4), 746-768.

Rock, D. (2019). Engineering value: The returns to technological talent and investments in artificial intelligence. Available at SSRN 3427412.

Rosenberg, N. (1963). Technological change in the machine tool industry, 1840–1910. *The Journal of Economic History*, 23(4), 414-443.

Stucke, M. E., & Grunes, A. P. (2016). Introduction: big data and competition policy. Big Data and Competition Policy, Oxford University Press (2016).

Susskind, R. and D. Susskind (2015). The Future of the Professions. Oxford, Oxford University Press.

Varian, H. R. (2010). "Computer Mediated Transactions." *American Economic Review,* 100(2): 1-10.

Appendix

Appendix A.1. – Additional reporting on AI ethics responses

All data shown here comes from surveys 3-5 as these questions were not included in the first tow rounds of the survey. These figures need to be viewed in color.

Figure A.1. – Types of job augmented by AI products (versus being replaced)

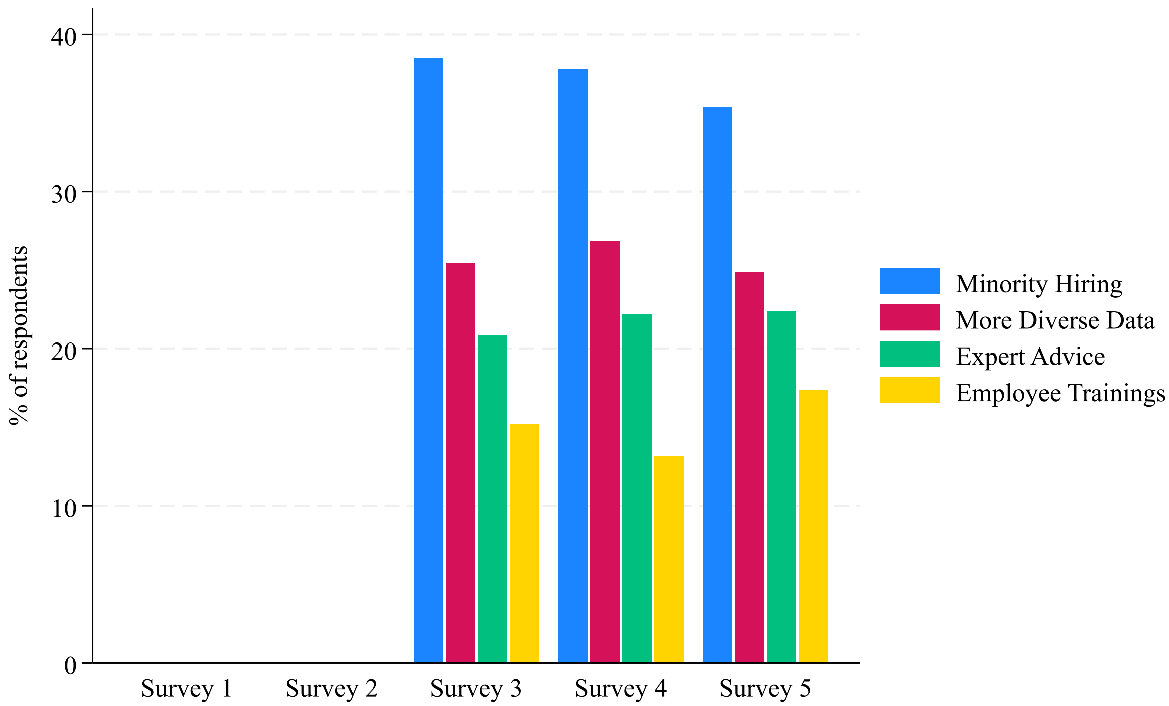


Figure A.2. – Types of job augmented by AI products (versus being replaced)

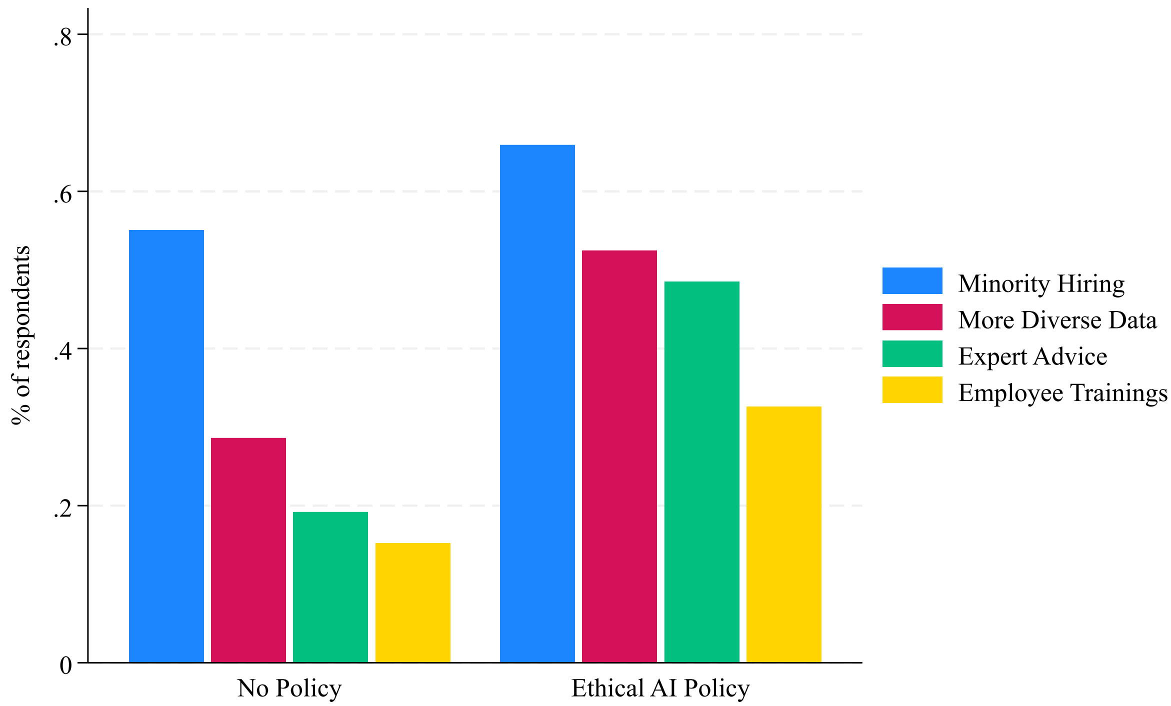


Figure A.3. – Types of job augmented by AI products (versus being replaced)

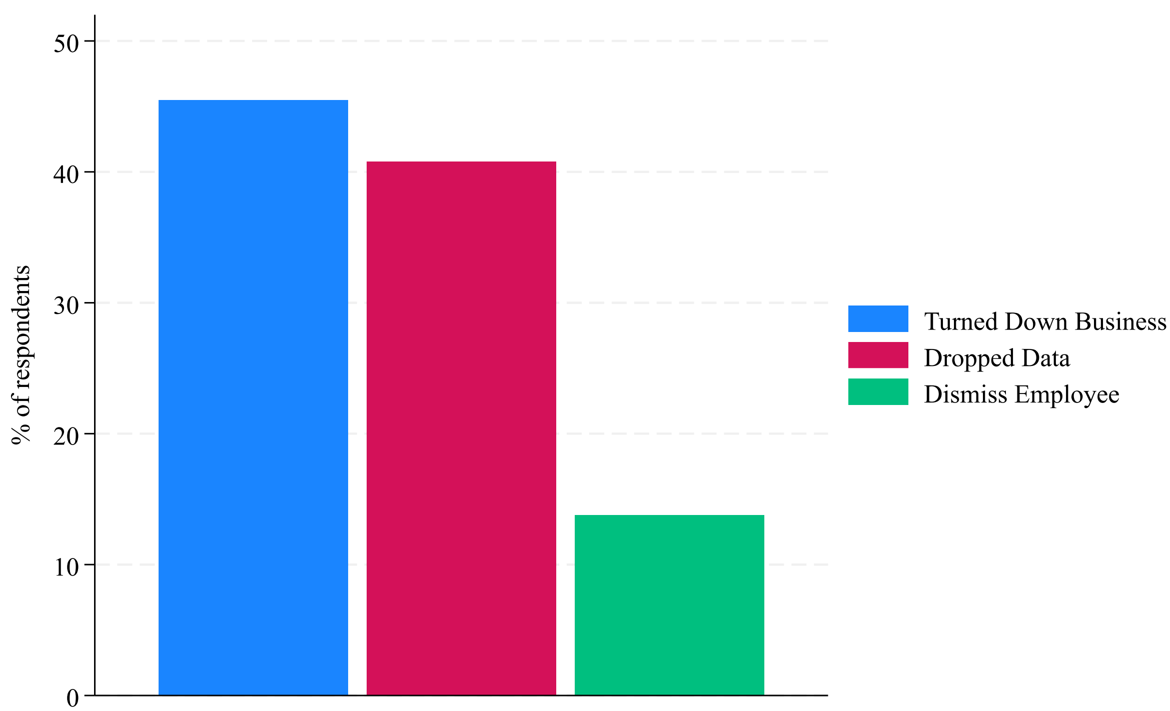


Figure A.4. – Types of job augmented by AI products (versus being replaced)

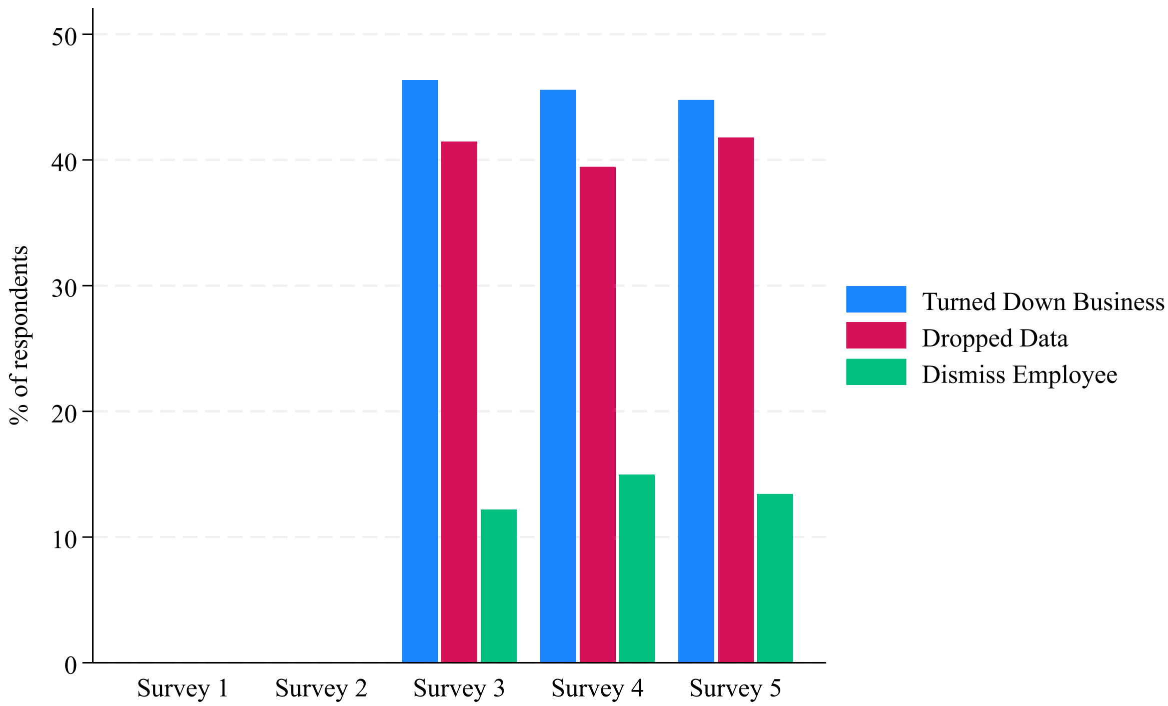


Figure A.5. – Types of job augmented by AI products (versus being replaced)

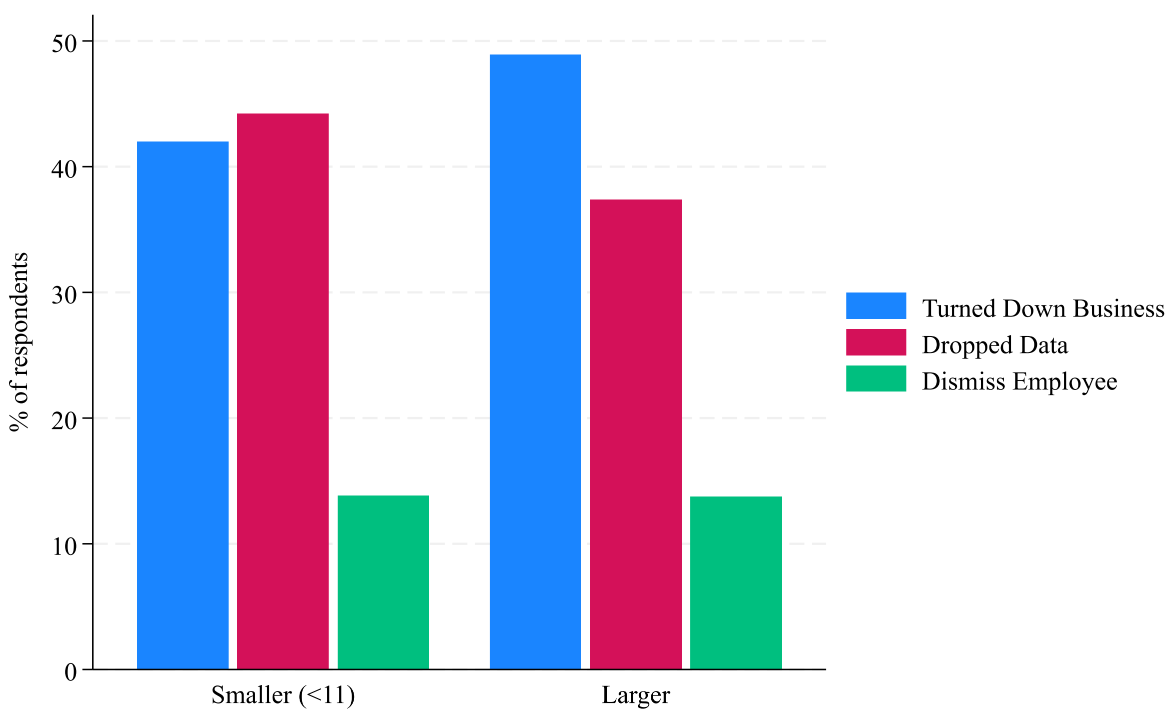


Figure A.6. – Types of job augmented by AI products (versus being replaced)

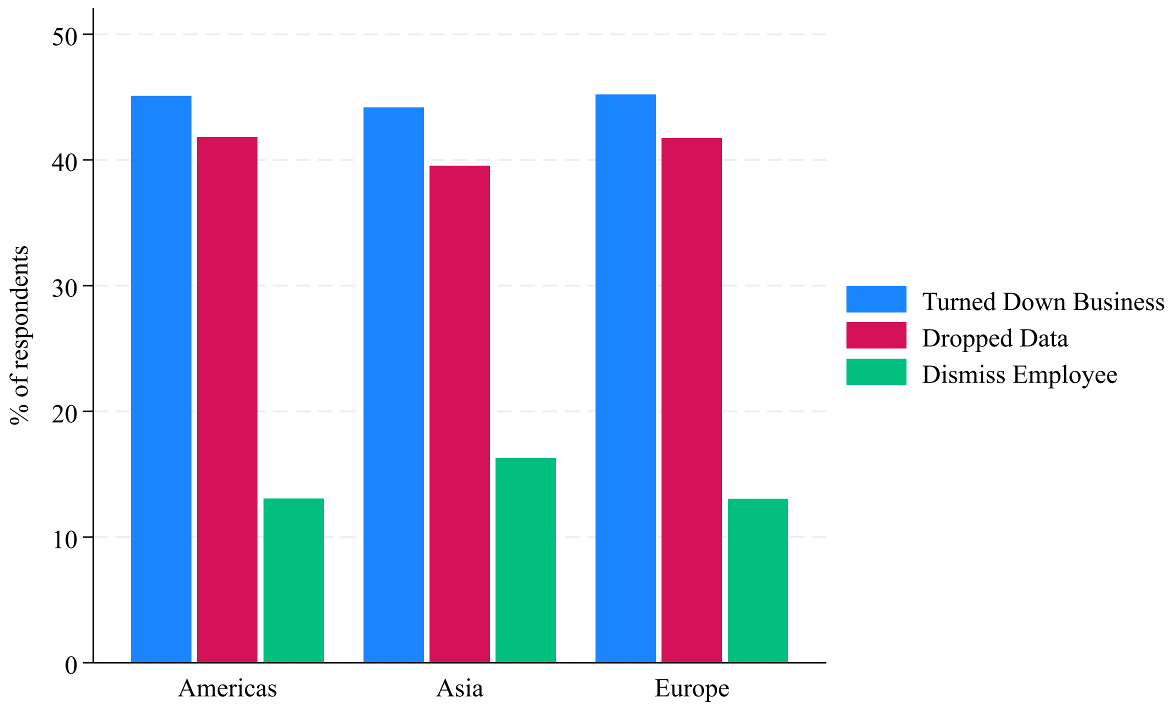


Figure A.7. – Types of job augmented by AI products (versus being replaced)