11-2018

The Business of AI Startups

James Bessen
Boston University School of Law

Stephen Michael Impink
New York University - Leonard N. Stern School of Business

Robert Seamans
New York University - Leonard N. Stern School of Business

Lydia Reichensperger

BU School of Law

THE BUSINESS OF AI STARTUPS

Boston University School of Law
Law & Economics Series Paper No. 18-28

November 2018

James E. Bessen
Boston University School of Law

The Business of AI Startups*

James Bessen, Technology & Policy Research Initiative, Boston University

Stephen Michael Impink, Stern School of Business, New York University

Lydia Reichensperger, Technology & Policy Research Initiative, Boston University

Robert Seamans, Stern School of Business, New York University

Abstract: New machine learning techniques have led to an acceleration of "artificial intelligence" (AI). Numerous papers have projected substantial job losses based on assessments of technical feasibility. But what is the actual impact? This paper reports on a survey of commercial AI startups, documenting rich detail about their businesses and their impacts on their customers. These firms report benefits of AI that are more often about enhancing human capabilities than replacing them. Their applications more often increase professional, managerial, and marketing jobs and decrease manual, clerical, and frontline service jobs. These startups sell to firms of different sizes, in different industries and nations, but the distribution of activity is distinct from that of larger firms. Firms serving EU customers appear to use higher levels of data protection.

JEL codes: O33, J21, L10

Keywords: artificial intelligence, automation, technology, labor-saving technological change, entry barriers, data protection

*Thanks to Victoria Smith for excellent research support. Thanks to Tim O'Reilly and O'Reilly Media, the Creative Destruction Lab, Philipp Hartmann, and Joachim Henkel for help in acquiring our sample. This work was funded by the Ewing Marion Kauffman Foundation. The contents of this publication are solely the responsibility of the authors.

Introduction

Breakthroughs in machine learning have led to an acceleration of "artificial intelligence" (AI) computer applications being developed for a wide range of uses. The potential of these new technologies has prompted assessments that these technologies will dramatically alter the economy. According to some commentators, we are entering a "Race Against the Machine" (Brynjolfsson and McAfee 2011), a "jobless future" (Ford 2015), and a "Fourth Industrial Revolution" (Schwab 2016). One paper estimates that 47% of jobs in the US will be "at risk of automation" in the next decade or so (Frey and Osborne 2017).

Much of this new literature on the economic consequences of AI has been based on estimated capabilities of the new technology, without taking into account possible difficulties in commercialization and adoption of the new technology. 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,[1] 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 2018).

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 applications. A look at the cutting-edge applications of the technology provides a window into likely outcomes over the next decade or so. To gain a peek through that window, this

[1] There is some reason to doubt the assessments. *None* of the occupations judged to be "completely automatable" 5 years ago has actually been completely automated nor even predominately automated. Interviews with people developing AI applications to automate the work of some of these occupations, such as accountants, are clearly focused on automating discrete tasks rather than the job as a whole.

paper reports on a global survey of startup firms developing and selling commercial applications based on AI. The survey questionnaire covers topics about the startup firms and their markets and customers, about their AI technologies, their use of data, and jobs and skills at their customers' firms.

One problem with surveys is that responses might be biased if respondents feel their self-interest is at stake. For example, if we asked whether their technology puts many jobs at risk, respondent answers might be biased downwards for fear of damaging the industry reputation. For this reason, we do not directly ask such questions. Instead, we ask indirect questions that reflect the impact on jobs. One set of questions concerns the benefits that their products provide customers, that is, the benefits that motivate customers to buy. Among the possible benefits is reducing customer labor costs as well as a wide variety of other possible benefits. We find that the benefits of current commercial AI applications are much more often about enhancing human capabilities than about reducing labor costs. Also, rather than ask whether their technology will eliminate jobs, we ask which occupational groups experience job losses and which experience job creation. Asked in this way, respondents have little reticence reporting that some occupations tend to lose jobs. Moreover, they report that other occupations show increased employment. The total picture revealed by these questions is at odds with some of the alarmist predictions. The actual effect of AI on jobs seems much more muted and varied by occupation, as developed below.

In addition to effects on employment, the survey also provides a window into other topics, including how AI startups address barriers to entry, 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, it is possible that large company investment may create barriers to entry in some markets. The survey asks respondents about several factors, such as data, hardware, software and firm size, which have previously been linked to barriers to entry. 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.

Hardware is less of a barrier to entry as the use of cloud-based technologies increases, and software is much more likely created and owned by the startup than licensed from larger firms. The AI startups in our sample disproportionately sell to mid-sized customers, signaling that their business models provide a comparative advantage for these firms.

In the following sections we first review related literature, then describe our survey, then discuss the results from the survey.

Literature

In addition to basic facts about the business of AI, the survey 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. 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 2018). 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, Gregory 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 et al. 2017a, 2017b) 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. Our survey provides some evidence about the actual impact of AI technologies based on employment changes at customer firms.

A few papers have looked at the impact of other technologies on jobs. 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. The Dauth el al (2017) 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 that the survey addresses 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 (2018).

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 and Autor 1999) and on job polarization (Goos , Manning, and Salomons 2009; Autor, Katz, and Kearney 2008). Bessen (2016) explores the association between computer use in an occupation and employment growth.

A third question the survey addresses is the extent of entry barriers into AI markets. It appears that investment in AI is currently dominated by large firms, especially a few large tech firms.[2] 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. 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 (2010) reports 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

---

[2] Using a survey, Bughin et al. (2017) estimate that businesses spent $18-27 billion on AI development in 2016 from internal funds, mainly from large companies. Venture capital, private equity, and other external sources invested $6-9 billion.

a certain point. For instance, Bajari et al. (2018) 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 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.

A fourth question the survey addresses is the use of data protections by AI startups. 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, Jin and Wagman 2018). 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 firms serving European vs non-European customers.

Survey and Sample

The survey was administered online using Qualtrics from May to September 2018. The 22 questions (excluding name and email questions) were pre-tested on academics and on roughly a half dozen firms with interviews. Potential respondents were contacted via email.

The main portion of our sample comes from the Crunchbase database of firms. Selecting firms tagged with a description containing "artificial intelligence," positive funding, and still in business in 2015, our base sample consists of 1,246 firms. Of these Crunchbase firms, 103 responded to the survey (Crunchbase response rate of 8.3%).

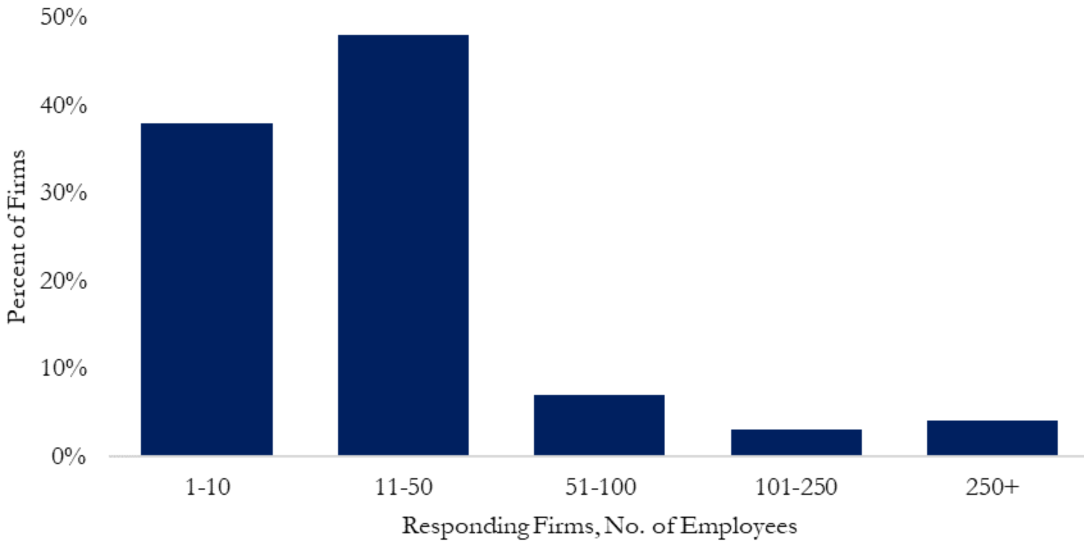
In addition to the Crunchbase sample, we received responses from three additional sources: alumni of the Creative Destruction Lab, a startup incubator, who were identified as working with AI; a list of machine learning companies obtained from Philipp Hartmann and Joachim Henkel (Hartmann and Henkel, 2018); also, O'Reilly Media ran a notice of the survey in its AI newsletter, providing a link to the online survey. In total, we obtained 179 completed responses.

Characteristics of respondents

We test the representativeness of our sample by comparing the characteristics of Crunchbase sample respondents and non-respondents. Comparing the two subsamples on number of employees (five size classes), age, number of investors, number of rounds of investment, and total amount of investment, we use t-tests on the differences in the subsample means between respondents and non-respondents. There are no differences significant at the 5% level (see Appendix Table A1). On average, the respondents tend to be slightly smaller and have slightly less capital.

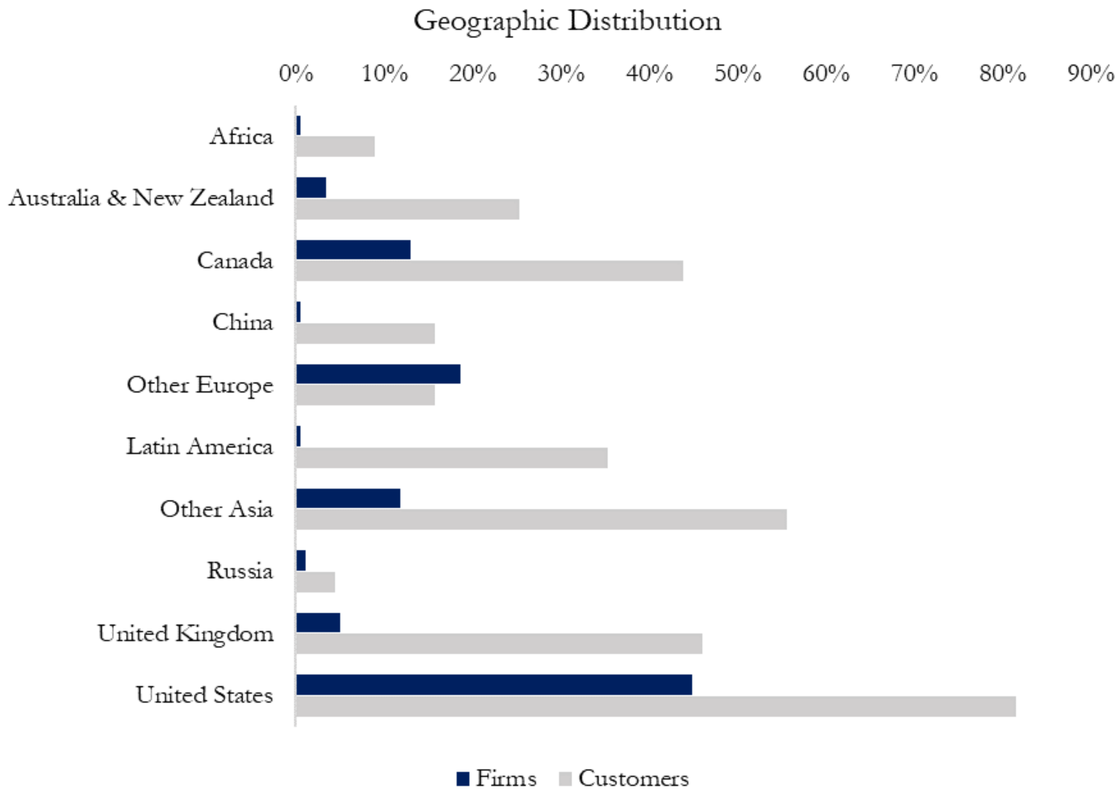
For the total sample of respondents, most firms are small with 50 employees or fewer (see Figure 1).

Figure 1. Employees per firm, respondents



Most of the firms are shipping product (67%), but 22% are in beta testing, and 10% are pre-beta. The respondents are located on every continent, but over half are in the US and Canada (Figure 2). Although the survey has global reach, response rates vary by country (Appendix Table A5). China and other emerging markets are likely to be underrepresented in the total population of AI startups given lower geographic coverage in the Crunchbase database. We use a Pearson's chi-squared test to determine that response rate and country are independent, so countries can be reported separately or aggregated. Response rates for the United States, where the majority of companies surveyed are located, is 7%.

Figure 2. Geographic Distribution of Startup Firms and Their Customers



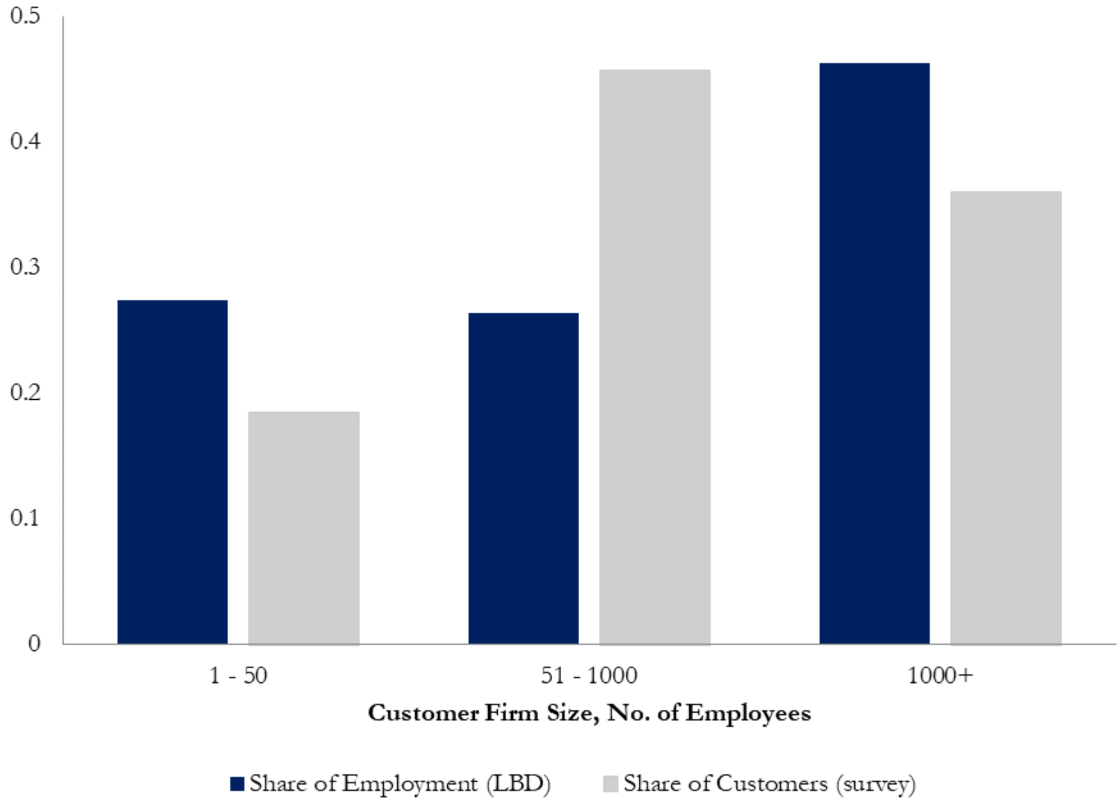
Findings

Markets

The surveyed startups also sell their products in all major regions (see Figure 2). The distribution of customers is generally broader, with over 50% of startup firms selling in the United States, in Asia, and in Europe.

Survey respondents sell disproportionately to mid-sized firms. Figure 3 shows the share of customers in each size class from the survey and also the share of employment for each size class for all US firms from the US Census Longitudinal Business Database (LBD). Almost half of the startups (46%) sell to firms in the middle category of 51-1000 employees, but these firms make up only 26% of the market as measured by employment size. The survey firms sell to both smaller and larger customers, but proportionately less than would be expected given the distribution of firms in the US, as indicated by the distribution of firms in the LBD.

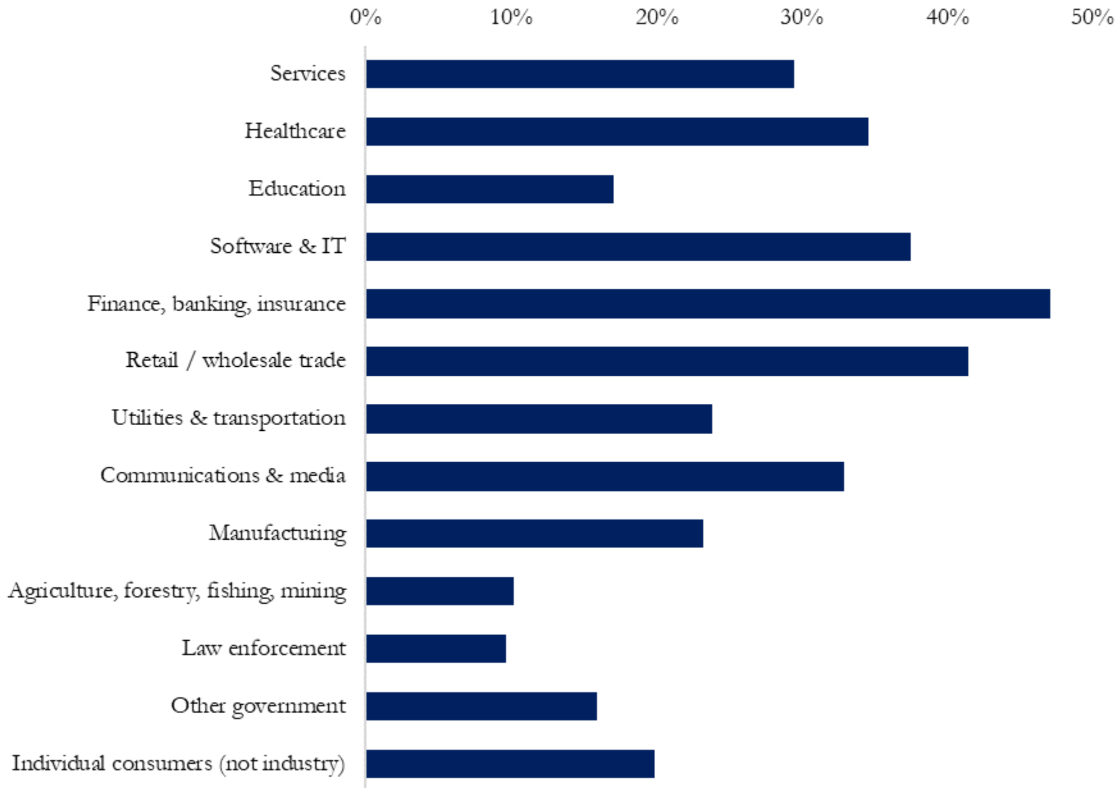
Figure 3. Startups Market Disproportionately to Mid-Sized Firms



Industry Distribution

The survey respondents sell to customers in all major industry sectors as well as to individual consumers as seen in Figure 4. 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).

Figure 4. Share of Firms Selling to Different Industries



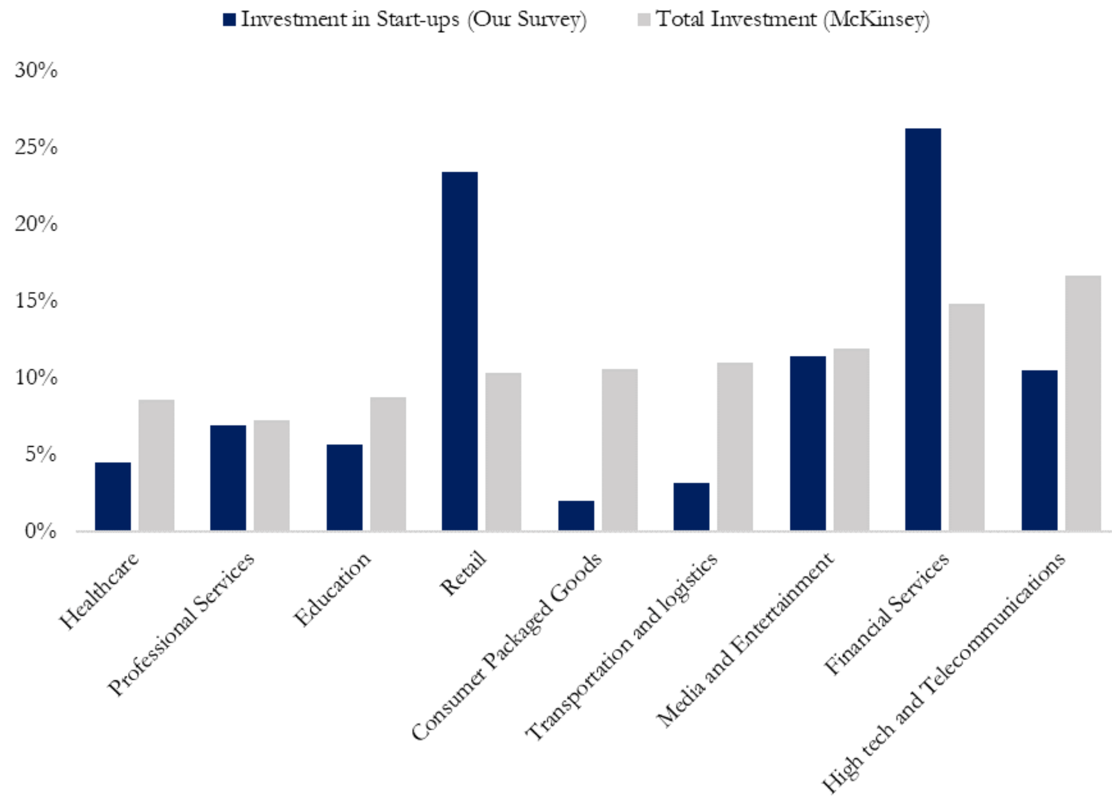
This broad distribution across industries suggests that the startup environment is healthy and many opportunities for entry exist. However, some industries may have relatively higher entry barriers than others. If there are no entry barriers, then both large and small firms will invest more in those industries that have the greatest technological opportunity. This means that without entry barriers, the distribution of AI development spending should look the same for small firms as it does for large ones. If, on the other hand, startups face significant entry barriers in an industry, then large firms would spend proportionately more in that industry. Comparing startup funding in different industries relative to total investment in AI in those industries helps identify those industries with possible entry barriers.[3]

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, Figure 5 shows the share of startup investment in each industry from our survey compared to the share of total investment from the McKinsey survey.[4]

[3] It is possible that these differences could also reflect inter-industry differences in access to funds, however, given that the differences in investment are so large, it is hard to imagine that credit differences are the main factor. In any case, credit constraints are a form of entry barrier.

[4] The Crunchbase sample includes total money raised by each startup. The comparison is only for those industries that we could match between our survey, where the relevant industry is the *customer* industry, and the industry reported by McKinsey.

Figure 5. Industry Share of Funds: Investment in Startups in Our Sample vs. Total Investment Figures from McKinsey



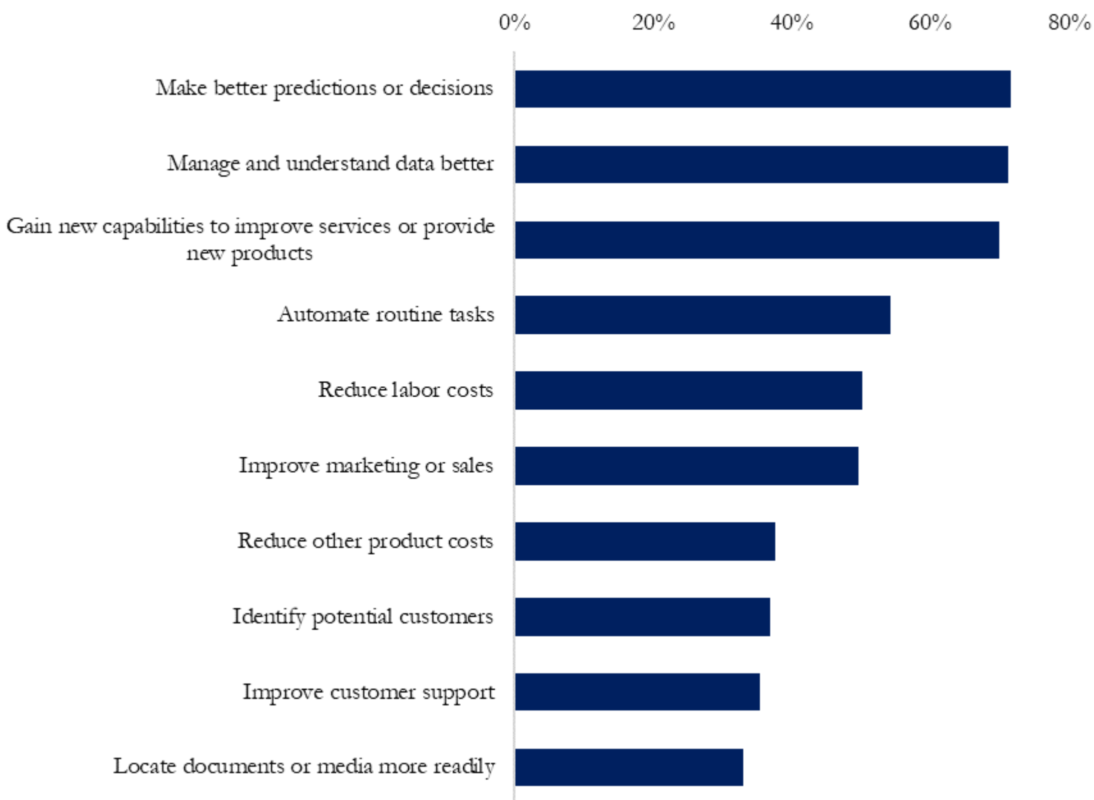
The chart shows that the share of investment going to startups and to all firms is roughly the same in many industries, but in three industries total investment—likely reflecting large firm investment—is disproportionately large: consumer packaged goods, transportation and logistics, and high tech and 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.

Augment or Replace?

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

(Figure 6). Each benefit is rated on a 5 point scale from "strongly disagree" to "strongly agree" (or "not applicable"). Figure 6 shows the share of responses ranked "strongly agree."[5]

Figure 6. Benefits to Customers



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, 54% strongly agree that their products automate routine tasks and 50% strongly agree that their products reduce labor costs. Taking these categories jointly, 55% of firms strongly agree that their products benefits

[5] Respondents tended to strongly agree if they agreed at all. The majority of answers were in the "strongly agree" category.

customers by replacing labor. In contrast, counting the remaining benefits as enhancing human capabilities, excluding "Reduce other product costs", 98% of firms strongly agree that their products benefit customers by enhancing capabilities. That is, AI appears to be much more about enhancing human capabilities than about replacing them. AI may very well eliminate jobs overall (something we explore further below), but because the technology appears to significantly augment human capabilities, it might well be associated with increased employment and wages for at least certain groups of workers.

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% strong agree. This difference across industry groups is highly significant (t-test probability value of 0.000).

Technology

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).

In building their products, the surveyed firms use the technology to perform a variety of functions. Some of these are developed internally and some are purchased from outside vendors (see Appendix Table A2). The most commonly used technology is natural language understanding and text analysis (62% 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 (see Appendix Table A3), 76% of reporting firms use neural networks including recurrent, convolutional, and generative adversarial neural networks. The next most common methods are clustering algorithms (59% of firms) and Bayesian or other methods of probabilistic inference (58% of firms). Other methods are used less frequently.

Thus, while startups use a wide variety of technologies to perform a variety of functions, a typical firm uses neural networks to do natural language text analysis for a cloud-based product.

Data

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.

Data Protection

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 (see Appendix Table A4).

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 7, 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.

On the other hand, there appear to be limited 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 appear slightly 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 7. Data Protection by Customer Location

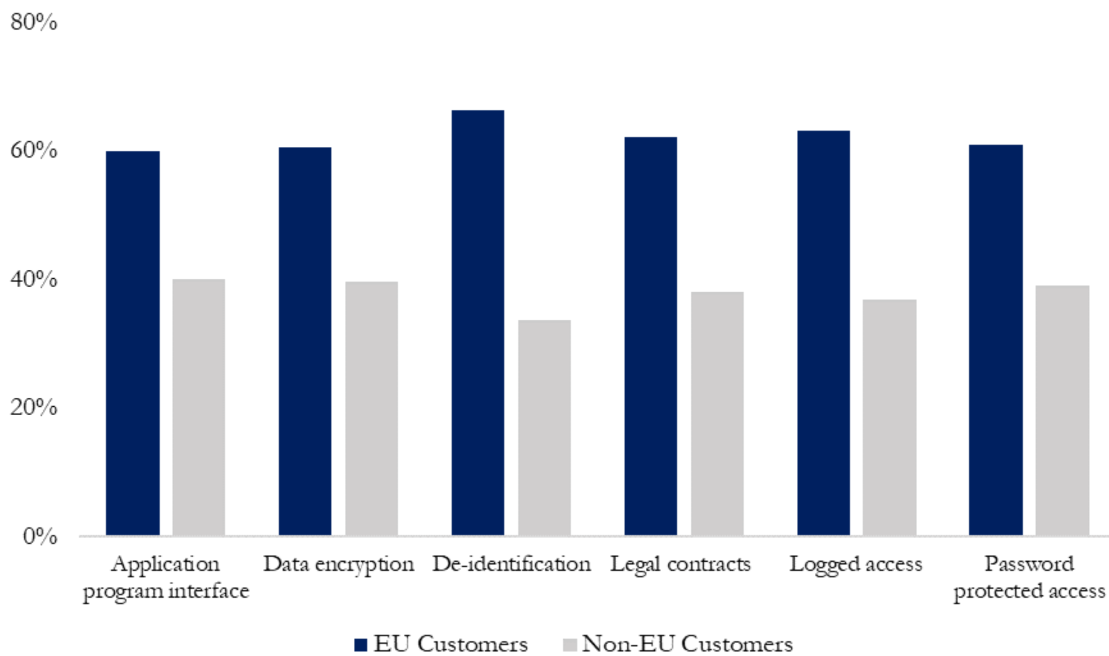
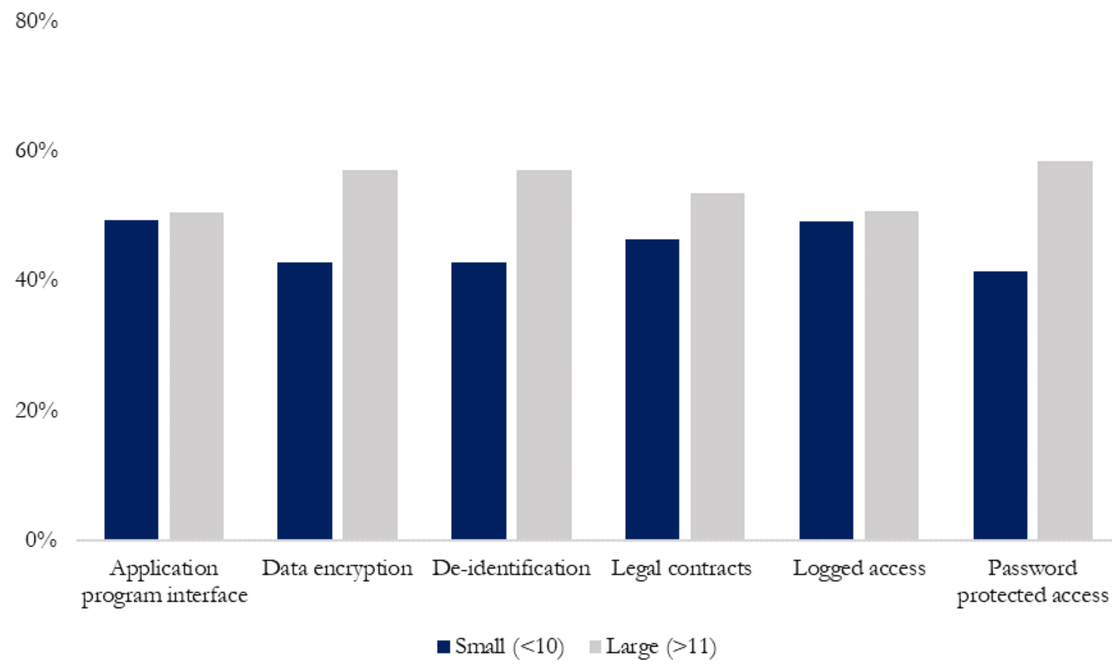


Figure 8. Data Protection by Firm Size

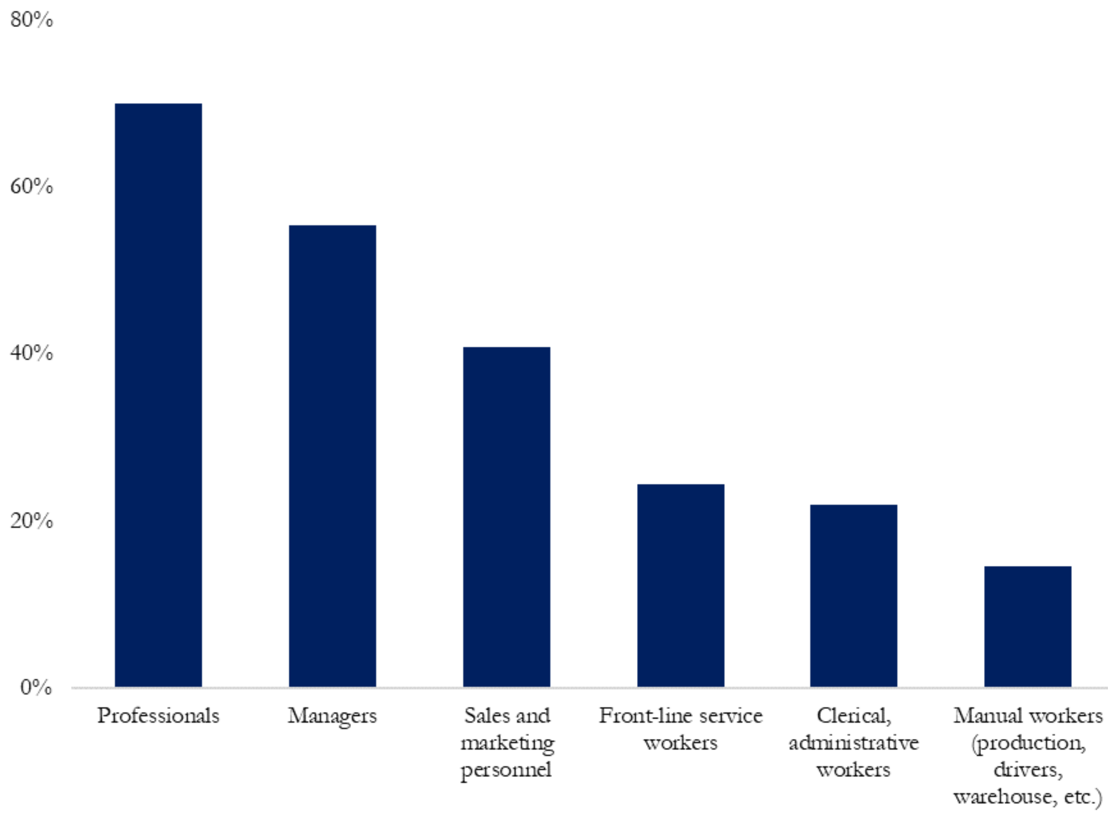


Workforce impacts

Skills and Training

Figure 9 reports the share of firms reporting whether each occupational group at their customer firms is a user or not (excluding those firms selling to individual consumers). White collar occupations are listed most frequently, including professionals (70%), managers (55%), and sales and marketing occupations (41%). 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 9. User Occupations



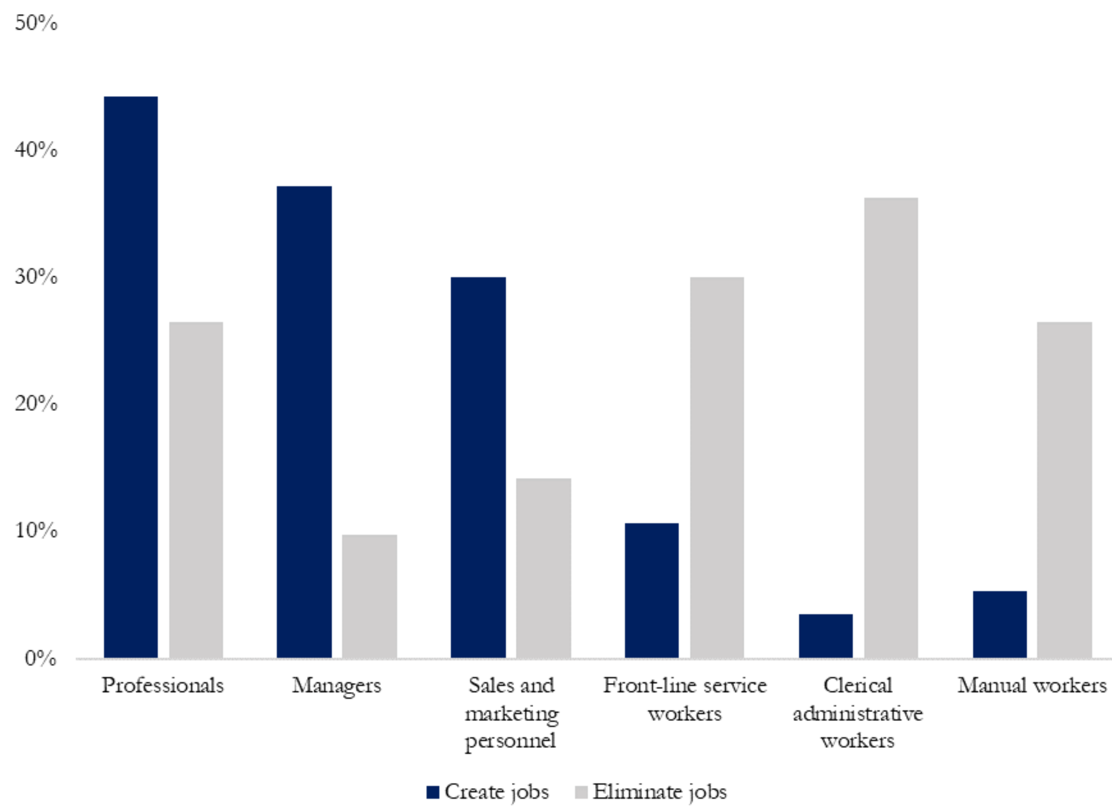
AI products appear to require a level of skill to the extent that the occupations involved have skill requirements. But does the use of these products require STEM skills or specialized training? The survey responses suggest that in most cases specialized computer skills or specific training demands 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.

Job Creation and Destruction

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 10 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 9.

Figure 10. Job Creation and Destruction by Occupational Group



Clearly, AI is not all about destroying jobs. In particular, those occupations that are most likely to use AI are also most likely to see jobs created. 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 10 reveals dramatic relative differences between occupational groups. Moreover, the results here are roughly consistent with the evidence above that only about half of the firms report labor cost reductions as a substantial customer benefit.

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:

- Data. 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.
- Software. 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.

- Firm size. 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. Data protection regulation has been described as a barrier that raises the cost of entry for startups. We find that data protections appear higher for startups with European customers, but little evidence of a difference in data protection between large and small firms.

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.

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 concerns 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, 2018). 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.” Available: https://www.accenture.com/us-en/insights/future-workforce/transforming-learning.

Acemoglu, D. and P. Restrepo (2018). "Artificial Intelligence, Automation and Work." NBER Working Paper.

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

Autor, D. and A. Salomons (2018). "Is automation labor-displacing? Productivity growth, employment, and the labor share." BPEA Conference Drafts.

Autor, D. H., et al. (2008). "Trends in U.S. Wage Inequality: Revising the Revisionists." The Review of Economics and Statistics **90**(2): 300–323.

Autor, D., Katz, L. “Changes in the Wage Structure and Earnings Inequality”. In: Ashenfelter O, Card D, editors. Handbook of Labor Economics. 3A. Amsterdam: North-Holland; 1999.

Bajari, P., et al. (2018). "The Impact of Big Data on Firm Performance: An Empirical Investigation." NBER Working Paper **24334**.

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

Bessen, J. (2017). "Automation and Jobs: When Technology Boosts Employment." Boston University School of Law, Law & Economics **No. 17-09**.

Brynjolfsson, E., McAfee, A. Race against the Machine. 2017

Brynjolfsson, E., Mitchell, T., & Rock, D. (2018). What can Machines Learn and What does it mean for Occupations and Industries. In *AEA Papers and Proceedings*, **108**.

Bughin, Jacques, E. Hazan, S. Ramaswamy, M. Chui, T. Allas, P. Dahlström, N. Henke, and M. Trench. "Artificial intelligence–the next digital frontier." McKinsey Global (2017).

Dauth, W., et al. (2018). "Adjusting to Robots: Worker-Level Evidence." Working Paper.

Felten, E. W., Raj, M., & Seamans, R. (2018). A Method to Link Advances in Artificial Intelligence to Occupational Abilities. In *AEA Papers and Proceedings*, **108**: 54-57.

Ford, M. . Rise of the Robots. London: OneWorld Publications. 2015

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. (2018). AI and the Economy. In *Innovation Policy and the Economy*, Volume 19. University of Chicago Press.

Goos, M., et al. (5863). "Job Polarization in Europe." American Economic Review **Vol. 99**: 58-63.

Graetz, G. and G. Michaels (2017). "Is Modern Technology Responsible for Jobless Recoveries?" IZA Discussion Papers **No. 10470**.

Hartmann, P. and J. Henkel. (2018)." Really the new oil? A resource-based perspective on data-driven innovation." Technical University of Munich working paper.

Jian Jia, Ginger Zhe Jin, Liad Wagman. 2018. "The Short Run Effects of GDPR on Technology Venture Investment." NBER Working Paper No. 25248.

Jin, Wang, and Kristina McElheran. "Economies Before Scale: Survival and Performance of Young Plants in the Age of Cloud Computing." (2017).

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

Manyika, James, Michael Chui, and Me Miremadi. "A future that works: automation, employment, and productivity." McKinsey Global Institute (2017a).

Manyika, James, Susan Lund, Michael Chui, Jacques Bughin, Jonathan Woetzel, Parul Batra, Ryan Ko, and Saurabh Sanghvi. "Jobs lost, jobs gained: Workforce transitions in a time of automation." McKinsey Global Institute (2017b).

Schwab, K. Fourth Industrial Revolution. New York: Crown Business. 2017

Sokol, D. Daniel, and Roisin Comerford. "Antitrust and Regulating Big Data." Geo. Mason L. Rev. 23 (2015): 1129, https://ssrn.com/abstract=2834611.

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

Table A1:

Variable	Not in Survey	Survey	H0: Diff = 0, Pr.
No. of Employees	1.85	1.64	0.052
Founding Year	2013.4	2013.6	0.525
No. of Rounds	2.27	2.27	0.996
No. of Investors	2.5	2.3	0.118
Money Raised ($M)	10.5	4.2	0.312

Table A2:

Which technologies did/do you use to develop your product? Were these developed internally or did they come from an external vendor (e.g., Amazon, Google, IBM, Microsoft or other)? (Please check all that apply.)

	Count	
Answer	Developed Internally	From External Vendor
Speech recognition	22	32
Visual recognition (including image, face and video)	57	27
Natural language understanding and text analysis	87	33
Natural language classification	77	24
Natural language translation	21	29
Natural language generation	43	12
Sentiment/emotion analysis	55	22
Decision management	90	7
Robotic process automation	45	7
Virtual agents/chatbots	27	11

Table A3:

What type of algorithms did/do you use to develop your AI (leave blank if you relied entirely on thrid party vendors)? (Please check all that apply.)

Answer	Count
Bayesian or other probabilistic inference	53
Clustering algorithms	50
Ensemble learning algorithms (e.g., random forest)	35
Genetic algorithms and evolutionary computing	11
Instance based algorithms (e.g., k-nearest neighbor, locally weighted learning)	30
Neural networks (including RNNs, CNNs, and GANs)	72
Reinforcement learning or transfer learning	39
Regression-based algorithms (e.g., logistic, LOES smoothing, stepwise regression)	38
Other	9

Table A4:

How is proprietary data handled between your firm and your customer? (Please check all that apply.)

Answer	Count
Legal contracts specify uses of the data	78
De-identification and other removal of personally identifiable information	46
Data encryption	55
Password protected access	52
Logged access	44
Application program interface	44
Other	4

Table A5:

	CB AI Targeted		Crunchbase		Response		% Responding
	Companies	% of Total	Companies	% of Total	Companies	% of Total	
United States	707	*57%*	229,861	*35%*	48	*47%*	*7%*
Canada	65	*5%*	19,051	*3%*	3	*3%*	*5%*
United Kingdom	106	*9%*	38,539	*6%*	7	*7%*	*7%*
France	35	*3%*	11,142	*2%*	4	*4%*	*11%*
Germany	35	*3%*	12,282	*2%*	3	*3%*	*9%*
Israel	51	*4%*	5,047	*1%*	1	*1%*	*2%*
China	26	*2%*	8,776	*1%*	-	*0%*	*0%*
Rest of World	218	*18%*	332,024	*51%*	36	*35%*	*17%*
Total	**1,243**		**656,722**		**102**		
Date Pulled	*6-Aug*		*30-Oct*				