**Proposal for Paper on Startups' Use of Data for Artificial Intelligence**

James Bessen and Robert Seamans

May 25, 2018

## Introduction

Throughout history, there has been a worry that automation would kill jobs and generate irreversible damage to the labor market. For example, John Maynard Keynes (1930) believed in technological unemployment and Wassily Leontief (1983), observing the dramatic improvements in the processing power of computer chips, worried that people would be replaced by machines, just like horses were made obsolete by the invention of internal combustion engines.

Despite the pessimism, there is ample evidence that automation fosters productivity growth (Brynjolfsson, Rock and Syverson, 2018). Economic research shows that, in the past, automation has often substituted for human labor in the short term, but in the long term has led to the creation of complementary jobs. Automation appears to have had different effects by occupation. For example, historically, it appears that middle-skill jobs are what get displaced by automation, leading to labor market polarization (though there is some evidence that labor market polarization is declining).

Projecting into the future is harder, but recent research suggests that occupations with many routine functions face a higher probability of automation (Frey and Osbourne, 2017). In particular, it has been argued that artificial intelligence (AI), which is able to learn from and replicate routine operations, may make many occupations redundant. However, large amounts of data are needed in order to train these virtual machines, and yet there has been no firm-level survey about how firms are collecting this data, let alone how they expect to use it to train their AI, or the impact they expect their AI to have on workers. Moreover, many of these firms are start-ups, which may lack the "big data" needed to train AI, or face other barriers to entry.

## Our Survey

To address this gap, we are conducting a survey of AI-enabled startup firms. The survey has been in the field for three weeks (started May 9, 2018), and generated about 100 responses to date. The questions on the survey are designed to, among other things, develop a better understanding about (1) startups' reliance on internally developed vs. externally sourced AI; (2) the types of algorithms used by startups that do internal development; (3) the sources of data on which the startups rely, and the extent to which this varies by the internal vs. external decision; (4) protections on this data, and the extent to which this varies by reliance on European customers, which would subject the startup to GDPR.

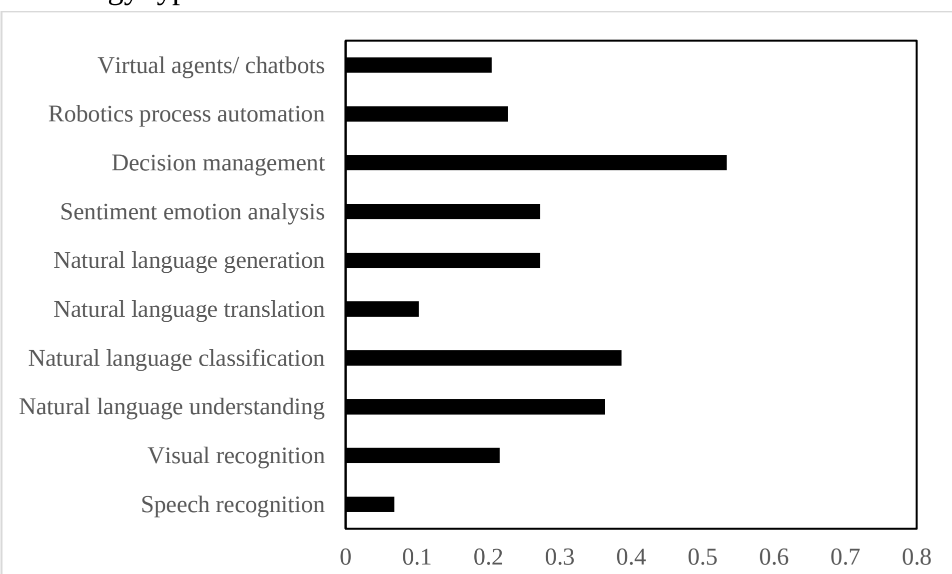
In addition, we plan to correlate the survey data with external data from Crunchbase on the startup valuation. Among other things, we plan to evaluate whether use of proprietary data correlates with higher valuations relative to use of third party data. If there are no returns to scale in data, as recently suggested by some researchers (e.g., Bajari et al, 2018; Varian, 2018) then there should be little difference.

**Initial Summary Statistics**

While our responses are still coming in, we thought it would be useful to take a "peek" at how GDPR is affecting (or not) startups by different types of technology. Our initial cut at the data is to focus only on those startups that rely in whole or in part, on their own internally developed technologies.

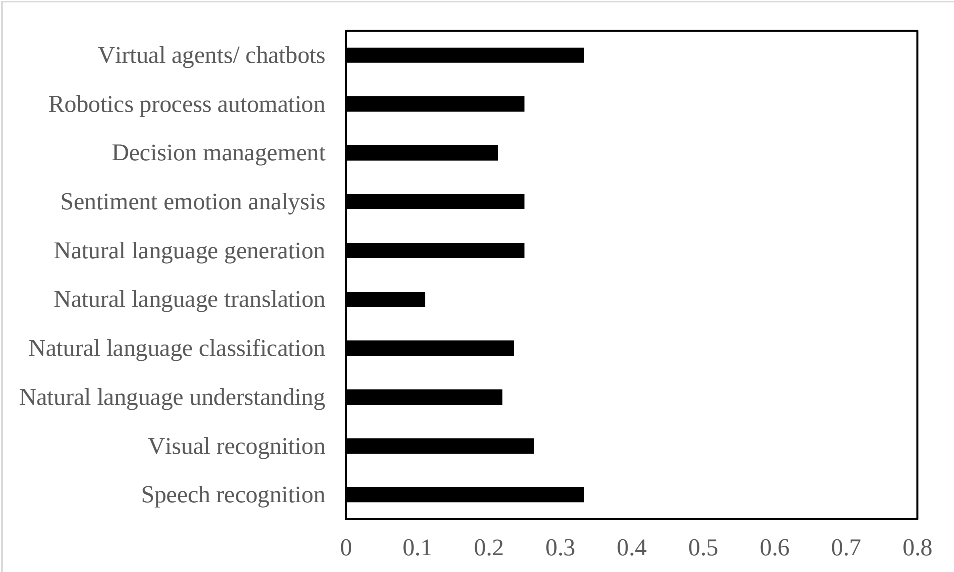
In Figure 1, we provide a breakdown of the technologies used by startups in our sample. As indicated, about 60% of the startups that rely on some of their own internally developed technology develop their own decision management technology (note these are not mutually exclusive; many startups rely on multiple types of technology). In contrast, only about 10% or less develop their own natural language translation or speech recognition.

Figure 1: Percent of respondents that rely entirely on internally developed technology, by technology type.




In Figure 2, we assess which of these startups report sales difficulties arising from GDPR. Interestingly, less than half (by technology type) report sales difficulties. This may be because the impact of GDPR is only just being felt. One thing on our "to do" list is to contrast this with the startups that rely on external providers of these technologies.

Figure 2: Percent of respondents that rely entirely on internally developed technology, that report difficulty arising from GDPR



## References

Bajari, P., Chernozhukov, V., Hortaçsu, A. and Suzuki, J., 2018. “The impact of big data on firm performance: An empirical investigation.” *National Bureau of Economic Research working paper No. w24334*.

Brynjolfsson, Erik, Daniel Rock, and Chad Syverson. 2017. “Artificial Intelligence and the Modern Productivity Paradox: A Clash of Expectations and Statistics.” in *The Economics of Artificial Intelligence*, eds. Ajay Agrawal, Joshua S. Gans and Avi Goldfarb, Chicago: University of Chicago Press (forthcoming).

Frey, Carl Benedikt, and Michael A. Osborne. 2017. “The Future of Employment: How Susceptible are Jobs to Computerisation?” *Technological Forecasting and Social Change* 114: 254–80.

Keynes, John Maynard. 1930. “Economic Possibilities for our Grandchildren.” In *Essays in Persuasion* (2010). London: Palgrave Macmillan.

Leontief, Wassily. 1983. “National Perspective: The Definition of Problems and Opportunities.” In *The Long-term Impact of Technology on Employment and Unemployment: a National Academy of Engineering Symposium*, June 30, 1983. Washington, DC: National Academy Press.

Varian, Hal. 2018. “Artificial Intelligence, Economics, and Industrial Organization.” in *The Economics of Artificial Intelligence*, eds., Ajay Agrawal, Joshua S. Gans and Avi Goldfarb, Chicago: University of Chicago Press (forthcoming).