NBER WORKING PAPER SERIES

COULD MACHINE LEARNING BE A GENERAL PURPOSE TECHNOLOGY? A COMPARISON OF EMERGING TECHNOLOGIES USING DATA FROM ONLINE JOB POSTINGS

Avi Goldfarb
Bledi Taska
Florenta Teodoridis

Working Paper 29767
http://www.nber.org/papers/w29767

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
February 2022

We acknowledge funding from SSHRC #502500 and the Sloan Foundation The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

At least one co-author has disclosed additional relationships of potential relevance for this research. Further information is available online at http://www.nber.org/papers/w29767.ack

Could Machine Learning be a General Purpose Technology? A Comparison of Emerging Technologies Using Data from Online Job Postings
Avi Goldfarb, Bledi Taska, and Florenta Teodoridis
NBER Working Paper No. 29767
February 2022
JEL No. O32,O33

ABSTRACT

General purpose technologies (GPTs) push out the production possibility frontier and are of strategic importance to managers and policymakers. While theoretical models that explain the characteristics, benefits, and approaches to create and capture value from GPTs have advanced significantly, empirical methods to identify GPTs are lagging. The handful of available attempts are typically context specific and rely on hindsight. For managers deciding on technology strategy, it means that the classification, when available, comes too late. We propose a more universal approach of assessing the GPT likelihood of emerging technologies using data from online job postings. We benchmark our approach against prevailing empirical GPT methods that exploit patent data and provide an application on a set of emerging technologies. Our application exercise suggests that a cluster of technologies comprised of machine learning and related data science technologies is relatively likely to be GPT.

Avi Goldfarb
Rotman School of Management
University of Toronto
105 St. George Street
Toronto, ON M5S 3E6
CANADA
and NBER
agoldfarb@rotman.utoronto.ca

Bledi Taska
Emsi Burning Glass
66 Long Wharf, Floor 2
Boston, MA 02110
bledi.taska@emsibg.com

Florenta Teodoridis
University of Southern California
701 Exposition Blvd,
HOH 503,
Los Angeles, CA, 900
florenta.teodoridis@marshall.usc.edu

I. Introduction

Knowing if an emerging technology is general purpose is of significant strategic importance for managers and policymakers. Such general purpose technologies (GPTs) are rare and hold potential for large scale economic impact because they push the production possibility frontier out several times (Bresnahan and Trajtenberg 1995). Examples of GPTs include the steam engine, electricity, computers, and the internet (Lipsey, Carlaw, and Bekar 2005). While theoretical models that explain the characteristics, benefits, and approaches to create and capture value from GPTs have advanced significantly, there is a scarcity of empirical methods to identify GPTs. The handful of attempts are customized to evaluate the GPT likelihood of certain technologies rather than provide a more universal evaluation approach and typically rely on the power of at least a decade of hindsight.

These limitations have at least three important implications. First, the lack of a more universal GPT evaluation approach makes it difficult to robustly evaluate the GPT likelihood of emerging technologies. Early approaches to identify GPTs use qualitative arguments and hence are context specific. Later quantitative approaches are also context specific. Specifically, a handful of studies (Moser and Nicholas 2004; Hall and Trajtenberg 2006; Feldman and Yoon 2012; Graham and Iacopetta 2014; Petralia 2020) evaluate the GPT likelihood of specific patents or patent classes that represent technologies hypothesized to be general purpose. While these patent-based approaches have some elements that can be generalized to other contexts, limitations remain. First, not all technologies sharply map onto a patent or class of patents. For example, the USPTO defines artificial intelligence (AI) as a collection of patents spread across a variety of patent classes. Second, identifying GPTs is not only about technology innovation, but also about technology adoption. Patents provide information about innovation activities, not adoption. Last, most patent-based methods require at least one measure based on forward citation data. This type of data is only meaningful after the passing of about a decade since the patented invention (e.g., Moser and Nicholas 2004; Hall and Trajtenberg 2006).

Second, reliance on the power of hindsight, such as that demanded by the citation-based measures, means the GPTs are identified only after the factors that influence the trajectory of GPT development play out (Lipsey, Carlaw, and Bekar 2005; Bresnahan 2010). By then, it is too late for managers to draw strategic insights from the classification because most of the value the technology holds would have already been captured and the magnitude of benefits that could have been captured can no longer be influenced by strategic decisions. Conceptually, this implies the possibility that, absent a methodology to identify the GPT likelihood of emerging technologies, some of the potential GPT benefits might never be realized or might be realized with a delay.

Third, the lack of a systemic and accessible way to evaluate the GPT likelihood of emerging technologies means that scholars and practitioners often speculate about these technologies' general purpose likelihood and base their research and business decisions on these speculations. For example, this is the case for several contemporaneously emerging technologies such as machine learning (e.g., Brynjolfsson, Rock, and Syverson 2019; Cockburn, Henderson, and Stern 2019; Trajtenberg 2019), cloud computing (e.g., Etro 2009), blockchain (e.g., Filippova 2019), nanotechnology (e.g., Forti, Munari, and Zhang 2019), 3D printing (e.g. Choi 2018), and the internet of things (e.g., Edquist, Goodridge, and Haskel 2019). By and large, the speculation is based on only one of the three characteristics of GPTs as identified in theoretical models, namely widespread interest in these technologies. However, the well-established theoretical definition of GPTs as conceived in Bresnahan and Trajtenberg (1995) and Bresnahan (2010) state that GPTs exhibit two additional characteristics: capable of ongoing technical improvement, and enabling of application sector innovations (Bresnahan, 2010, p. 764). This incomplete assessment of GPT classification could lead managers to pursue suboptimal technology strategies. For example, collaboration with academia and competitors is necessary to generate value from emerging technologies that are likely general purpose (e.g., Allen, 1983; Nuvolari, 2004; Cassiman and Veugelers 2006). If the technology is not general purpose, such investments are unnecessary and could even be harmful if a potential source of competitive advantage is shared or revealed in the process (e.g., Cohen, 2010).

In this paper, we propose a more systematic methodology to evaluate the GPT likelihood of emerging technologies that exploits job posting data. Our framework follows the well-established theoretical definition of GPTs (Bresnahan and Trajtenberg, 1995; Bresnahan, 2010). The approach is in line with the patent-based studies that engage in applying this theoretical framework to identify GPTs (Moser and Nicholas 2004; Hall and Trajtenberg 2006; Feldman and Yoon 2012; Graham and Iacopetta 2014; Petralia 2020). We focus on job posting data because human capital is an input into technology development and diffusion. Hence, the skills listed in job postings should reflect firms' intentions to engage with emerging technologies earlier than captured in patent data. Indeed, Tambe and Hitt (2012a,b) demonstrate that emergent technology diffusion can be measured using labor demand data.

We develop four measures to evaluate the GPT likelihood of emerging technologies with respect to the three GPT criteria in Bresnahan (2010). We evaluate widespread use in many industries by calculating a Gini coefficient for job postings referencing emerging technologies across industry sectors. We evaluate capable of ongoing technical improvement by estimating use in many research job postings through two measures: the number of research job postings and the fraction of postings using the technology that are research-focused. We evaluate the ability of emerging technologies to enable application sector innovations by calculating a Gini coefficient for research job postings referencing the emerging technologies across industry sectors. To examine GPT likelihood, we rank order emerging technologies across the four measures. Evaluating GPT likelihood by identifying which technologies rank high on all GPT criteria is a similar approach to prior quantitative studies that identify GPTs with historical data (e.g., Feldman and Yoon 2012).

We benchmark and demonstrate the GPT evaluation methodology we propose in the context of a set of emerging technologies and job posting data from 2010 to 2019. The job posting data come from Burning Glass Technologies and include postings from over 40,000 online job boards. According to Burning Glass, they have the near-universe of jobs that were posted online from 2010 through October 2019. We examine job postings that reference 21 emerging technologies: 3D printing, big data, blockchain, business intelligence (BI), cloud computing, CRISPR, data mining, data science, geographic information systems (GIS/GPS), internet of things (IoT), machine learning (ML), nanotechnology, natural language processing (NLP), polymer science, quantum computing, RFID, robotics, service-oriented architecture (SOA), telecommunications, virtual reality (VR), and Web 2.0. These represent all technologies from the Gartner "hype cycle" listed between its inception in 1995 and 2019 that are also listed as skills in the Burning Glass job posting data.

We start by benchmarking our method against Petralia's (2020) patent-based quantitative approach for evaluating GPT likelihood because that approach is at the frontier of such efforts, builds and extends prior efforts, and eliminates the need to rely on forward citation data that would require a much longer observation period than available for emerging technologies. We find that our job posting-based measures of whether an emerging technology is likely to be a GPT are strongly correlated with the patent-based measures. Moreover, we find that our measures predict future changes in the patent-based measures over-and-above using lagged patent-based measures alone. This suggests that our approach can be employed to glean early information about the GPT likelihood of emerging technologies, and hence help managers make more informed decisions while these technologies are diffusing.

Next, we apply our method to evaluate the GPT likelihood of our set of emerging technologies. We find that ML, alongside a set of complementary technologies (BI, big data, data mining, data science, and NLP) consistently ranks at or near the top. Because the boundaries of GPT candidates can span multiple complementary technologies (Field, 2008; Petralia, 2020), we interpret this result to suggest that these technologies are relatively likely to be a GPT. Cloud computing and robotics are also relatively prevalent in research job postings and widespread research use in application sectors.

Our interpretation is also informed by a comparison with telecommunications. Our results suggest that telecommunications is a potential GPT during our observation period (2010-2019), but less so than the ML cluster in 2019. Telecommunications has been identified by others as a GPT during a similar period (e.g., Liao et al., 2016; Strohmaier and Rainer, 2016; Petralia, 2020). Therefore, ML is relatively likely to be a GPT, as benchmarked against a technology that others have identified as a GPT. This benchmark, however, is imperfect because telecommunications in 2010 was likely a more mature technology than ML in 2019. For this reason, we prefer to focus our interpretation on the relative GPT likelihood of ML and the other emerging technologies in our dataset.

Our results also suggest that most of the other emerging technologies in our dataset are unlikely to be on the path to becoming GPTs in their current form.[2] Of course, several of these technologies have not received widespread speculation that they are GPTs (e.g., RFID and SOA).

[2] A limitation of our observation period is that we may catch some technologies before they are widely used. For example, it is possible that some early stage hyped technologies, such as quantum computing, may eventually prove to be GPTs, even if they are not yet widespread.

Nevertheless, our application of the job posting data method does not provide a definitive measure of which technologies in our dataset are GPTs. Without a long enough observation period that includes a clear threshold as to what constitutes a GPT, we can only identify the relative rank of the emerging technologies in our dataset in terms of the three characteristics of GPTs listed above. In this way, the application exercise in our paper suffers from many of the same weaknesses as the prior empirical attempts to identify GPTs. Like the prior literature, it can assess whether a technology is widely used and enables innovation across industries but there is no ground truth available for benchmarking. In addition, like much of the prior literature, our exercise is based on one data source. In addition, our set of emerging technologies is limited to those technologies that can be mapped to job posting skills in our data. For example, technologies such as tablets or mobile phones are too broad to be matched to job posting skills. The extent with which these limitations extend to the method we propose depend on the richness of available data and the time frame it covers. Given the increased richness of available digital data, we hope future applications of our methodology would be able to uncover more nuanced insights.

The paper proceeds as follows. In Section II, we discuss the strategic importance of general purpose technologies and review the advantages and shortcomings of existing approaches to identifying GPTs. We describe the state of empirical approaches to identify GPTs and our proposed method in Section III. We benchmark our method and provide an application in Section IV. We discuss limitations, contributions, and areas for future research in Section V.

II. General Purpose Technologies

As originally conceived in Bresnahan and Trajtenberg (1995), GPTs are “characterized by the potential for pervasive use in a wide range of sectors and by their technological dynamism” (p. 84). GPTs are transformative because they open up new opportunities for innovation and economic growth, linking the technical implementation of an innovation to its macroeconomic consequences. Specifically, GPTs generate productivity gains through an innovation loop. The commercial viability of a potential GPT is initially shown in academia or a producing industry. This spurs innovation in application industries. These application industry innovations, in turn, further advance producing industry innovation, and so on. This feedback loop provides the reason that GPTs unusually generate large productivity improvements. Most innovations will push out the production possibility frontier once, but then capital and labor adjustments will have diminishing returns. The feedback loop from GPTs means that each innovation pushes the production possibilities frontier out several times, increasing productivity substantially over a sustained period of time.

This implies that the benefits of a GPT are large but occur with a long lag (Bresnahan, Brynjolfsson, and Hitt 2002). GPTs require complementary innovations that take advantage of the capabilities of the technology (e.g., Greenwood and Yorukoglu 1997; Aral, Brynjolfsson, and Wu 2012; Tambe, Hitt, and Brynjolfsson 2012). Bresnahan (2010) describes how his and Trajtenberg's independent prior research on accounting and CT scans led them to appreciate the importance of complementary innovation in application sectors and motivated the GPT research. Complementary innovation also means long run financial investments. For example, electricity's early uses focused on street lighting and street railways (Lipsey, Carlaw, and Bekar 2005). Over time, innovation occurred in a wide range of sectors, from upstream advances in power generation to downstream development of household appliances such as washing machines, vacuum cleaners, and refrigerators. Electrification also led to the reorganization of factories (David 1990). Importantly, it was primarily these later innovations that drove productivity growth both within companies and at a macroeconomic level.

Because the benefits of GPTs occur with a lag, it is difficult to identify these technologies early on. Bresnahan (2010, p. 764) defines a GPT as a technology that "(1) is widely used, (2) is capable of ongoing technical improvement, and (3) enables innovation in application sectors." Wide use, ongoing improvement, and follow-on innovation are difficult to measure without the benefit of hindsight. Therefore, most studies that seek to identify GPTs use data that is only available with a lag of more than a decade, including evidence of widespread adoption, such as patent citations, and evidence of productivity impact. The examples in Bresnahan and Trajtenberg's original article and in Jovanovic and Rousseau's (2005) review are backward-looking, emphasizing that the impact can be seen after many years. Lipsey, Carlaw, and Bekar (2005) take a qualitative approach, using historical examples to identify GPTs using a millennia-long time scale, starting with the domestication of plants. More systematic studies of particular technologies, such as Moser and Nicholas's (2004) study of electricity and Feldman and Yoon's (2012) examination of the purposeful recombination of genetic material (rDNA), use decades of data on patent citations. By this point, the GPT classification is too late to be managerially useful. Therefore, it is valuable for managers to have an early sense of whether an emerging technology is likely a GPT for managers to make informed decisions about their organization's technology strategy.

First, recognizing early that a technology is a GPT would allow organizations to engage in managing the necessary coordinated innovation process (e.g., Greenwood and Yorukoglu 1997; Aral, Brynjolfsson, and Wu 2012; Tambe, Hitt, and Brynjolfsson 2012) by, for example, investing in developing internal R&D capabilities, engaging in collaboration with academic researchers, forming alliances with other organizations in the producing industry and with other organizations in applications industries, including competitors (e.g., Allen, 1983; Nuvolari, 2004; Cassiman and Veugelers 2006; Cohen 2010), and considering third-party service providers (Attewell 1992; Bresnahan and Greenstein 1996). If the technology is not a GPT, such large-scale investments in innovation coordination are unnecessary and could even be harmful if a potential source of competitive advantage is shared or revealed in the process (e.g., Cohen 2010).

Second, having an early sense of GPT likelihood would help organizations structure their activities to maximize the value they can capture from such investments in innovation. As Gambardella et al (2020) emphasize, when a technology can be employed to generate innovations that apply to a wide range of industries, the profits often accrue to those who own complementary assets. Conti, Gambardella, and Novelli (2019) demonstrate that intermediaries often arise that focus on these complementary assets and capture much of the value. In other words, when a technology is general purpose, it is relatively difficult for innovators to capture the lion's share of the value. Investing in complementary assets becomes particularly important. Alternatively, a more profitable path may be to focus on investing only in complementary capital and skills. This path, however, requires others to invest in innovation while capturing less of the value. Complementary assets play a less central role in value capture if the technology is not general purpose (Gambardella et al. 2020).

Third, because further innovation in the industries that use the technology is essential for the GPT to have an impact, financial investment horizons are likely to be long term. Thus, organizations looking to benefit from GPTs need to prepare for the financial burden by, for example, ensuring management buy in and securing budgets that are not subject to the same expectations of day to day operations. This approach would allow organizations to experiment with different potential areas where the GPT could be incorporated, identify collaborators, and learn how business processes need to change (Bresnahan and Greenstein 1996). In contrast, if the technology is not a GPT, then the investment horizon is shorter, and the need for organizational change is narrower.

It is also possible to engage in misplaced strategic actions driven by wrong evaluations of emerging technologies being GPT. A variety of recent innovations are claimed to represent GPTs, including machine learning (e.g., Brynjolfsson, Rock, and Syverson 2019; Cockburn, Henderson, and Stern 2019; Trajtenberg 2019), cloud computing (e.g., Etro 2009), blockchain (e.g., Filippova 2019), nanotechnology (e.g., Forti, Munari, and Zhang 2019), 3D printing (Choi 2018), and the internet of things (e.g., Edquist, Goodridge, and Haskel 2019). These claims are generally based on evaluating the widespread use criterion of a GPT, while disregarding the other GPT criteria. A better assessment of GPT likelihood is thus necessary because it could reduce the uncertainty of managerial strategy with respect to innovation, organizational change, and investment.

III. Method to Identify General Purpose Technologies

III.1. Current state of established methods to identify GPTs

The early empirical approaches to identify GPTs used qualitative arguments. Bresnahan and Trajtenberg (1995) identify examples of technologies that are widespread and involve complementary innovation. Lipsey, Carlaw, and Bekar (2005) provide detailed histories of a variety of technologies that they label as GPTs. Jovanovic and Rousseau (2005) provide a narrative description of the diffusion of electricity and IT. Such methods are context specific and hence cannot be generalized to evaluate emerging technologies.

More recently, scholars turned to patent-based quantitative approaches to evaluate GPTs. The advantages of patents are the rich information they contain about firms' technological innovations, the nature of those innovations, and the relationships between them. There are however significant disadvantages in relying on patents to empirically evaluate the GPT likelihood of emerging technologies. One such limitation stems from the type of data captured in patents. Patents are a great paper trail of technological innovation but not of technological adoption. This is important because the characteristics of GPTs, as documented in the well-established theoretical definition (Bresnahan and Trajtenberg, 1995; Bresnahan, 2010), span both innovation and adoption patterns.

For example, wide use of the technology, the first GPT criterion, includes both attempts to innovate by building on the GPT, and attempts to adopt either the GPT or the innovations that build on the GPT. These latter actions are not reflected in patents. Moreover, the patenting process is lengthy, hence it takes years to observe informative GPT-based innovation attempts in patents.

Other limitations are directly related to the choices made by scholars who developed the handful of quantitative methods to identify GPTs. One such limitation is the choice to restrict to technologies whose boundaries are equivalent to those of a patent or a class of patents. Unfortunately, not all technologies can be mapped onto a patent or class of patents. For example, the USPTO defines artificial intelligence (AI) as a collection of patents spread across a variety of patent classes. Another limitation is reliance on patent forward citation data to evaluate the breadth of influence and diffusion of GPTs across sectors. Meaningful forward citation data is available only after the passing of a long enough period of time. Specifically, four out of the five total studies that develop a patent-based approach to evaluate the GPT likelihood of certain technologies rely on at least one measure based on forward citation data. Moser and Nicholas (2004) measure how patents from the 1920s were cited between 1976 and 2002. Hall and Trajtenberg (2006), Feldman and Yoon (2012) and Graham and Iacopetta (2014) construct GPT measures based on forward citations over about a decades-long timespan. Hall and Trajtenberg (2006) and Graham and Iacopetta (2014) warn that forward citation data younger than about a decade out likely does not carry sufficient meaning. Last, Hall and Trajtenberg (2006) warn that the patent-based measures they pioneered do not give internally consistent estimates of GPT likelihood, hence more research is needed to identify robust approaches to identify GPTs empirically. More specifically, in a recent review of advantages and disadvantages of citation-based measures to capture attributes of technological generality, Jaffe and de Rassenfosse (2019) summarize the issues raised by Hall and Trajtenberg (2006): “The article concludes that the identified measures, although promising, give contradictory messages when taken separately and that it is not obvious how to combine those measures to choose a sample of GPT patents. The fundamental difficulty is that we don’t have measures of how general-purpose a technology is other than broad conceptions of GPT technologies. Thus, although it seems plausible that general-purposeness would be reflected in citation patterns, it is hard to pin such patterns down or test their validity” (p.27).

Most recently, the only other empirical study that engages with identifying GPTs, Petralia (2020), takes on the citation limitation challenges to develop a patent-based method that does not rely on forward citations. He notes that aside from the disadvantages highlighted above, forward citation data are a poor measure for capturing the breadth of use and diffusion of technologies because patent citations are "more concerned with delimiting the scope of the invention rather than comprehensively accounting for its knowledge composition" (Petralia, 2020, p. 49). Indeed, he shows that citation-based measures remain invariant over a long period of time relative to his method that is able to capture the dynamics of technology diffusion and evolution.

III.2. Our method

We build on the most recent effort of assessing the GPT likelihood of technologies as documented in Petralia (2020) to propose a method based on job posting data. We depart from patent data to address some of the limitations of patent-based methods as highlighted in the prior section. Specifically, we note that human capital is an input into technology development and diffusion, and hence the skills listed in job posting data should reflect firms' intentions to engage with a certain technology earlier than captured in patent data. Indeed, Tambe and Hitt (2012a,b) demonstrate that emergent technology diffusion can be measured using labor demand data.

Job postings specify industry, skills needed, and whether the job is a research position. Therefore, it is possible to get an informative (though imperfect) early measure of whether a technology is widely used and involves innovation, even in application sectors, by analyzing the skill requirements in job postings across industry sectors. The approach assumes that an informative mapping of emergent technologies to job posting skills can be constructed. The ability to evaluate the GPT likelihood of technologies for which such a mapping cannot be developed remains a limitation of our approach.

We base our methodology on the well-established theoretical definition of GPTs: "(1) is widely used, (2) is capable of ongoing technical improvement, and (3) enables innovation in application sectors." (Bresnahan, 2010, p. 764). We evaluate widespread use by calculating the Gini coefficient on job postings across industry sectors. We assess if a technology is capable of ongoing technical improvement by evaluating whether the technology shows up in many research job postings, because ongoing technical improvement implies continuous innovation in the technology. We evaluate the ability of technologies to enable innovation in application sectors by calculating the Gini coefficient on research job postings across industry sectors.

The Gini coefficient is a measure of statistical dispersion that ranges between 0 and 1, with 0 meaning all values are perfectly equal and 1 meaning only one observation has all of the measured factor. Gini coefficients are typically used to measure economic inequality, but they are also useful in measuring statistical dispersion, particularly when small and zero values for observations are informative. In contrast to the Hirshman-Herfindahl index, which measures whether a small number of observations have most of the share, the Gini coefficient captures whether a large number of observations have little share (e.g., Fleder and Hosanagar 2009; Cui, Orhun, and Duenyas 2018). In their patent-based approach for identifying GPTs, Graham and Iacopetta (2014) recognize and demonstrate the value of the Gini index for GPT evaluation. Low Gini values calculated based on the percentage of technology skill types in job postings across industries suggest widespread use of the technology across industries. Similarly, low Gini values calculated based on the percentage of technology skill types in research job postings across industries suggest the ability of technologies to enable innovation across application sectors.

We evaluate the extent to which a technology shows up in many research job postings using two measures: the number of research job postings with skills that map onto the technology and the fraction of such postings that are research-focused. The assumption is that a technology that is capable of ongoing technical improvement would have many research job postings and a disproportionate fraction of such postings relative to the total number of jobs for the technology. It is possible that the research-related jobs use the technology as an input but do not drive innovation in the technology directly. Just as many researchers may use microscopes without leading to innovation in microscope technology, many researchers may use ML, CRISPR, or nanotechnology without improving the technology. However, we argue that a low number of research job postings and a low fraction of such postings relative to the total number of jobs for the technology suggest the technology is not capable of ongoing technical improvement. In other words, we argue that many research job postings and a disproportionate fraction of such postings relative to the total number of jobs for the technology is not a sufficient but a necessary condition for evaluating if a technology is capable of ongoing technical improvement.

To evaluate the GPT likelihood of emerging technologies, we propose to rank the technologies along the four measures. Those technologies that rank near the top in all measures would have a higher likelihood to be GPT. The approach to rank-order the technologies is in line with some of the other patent-based empirical studies that evaluate the GPT likelihood of certain technologies (e.g., Feldman and Yoon 2012).

IV. Our Method in Action: Benchmarking and an Application

Our goal is to use job posting data to assess the GPT likelihood of emergent technologies. To demonstrate our proposed method, we apply it to a set of emerging technologies. To benchmark our approach, we evaluate the correlation with and predictive power of our proposed measures relative to those developed by Petralia (2020), the most recent patent-based methodology to evaluate GPTs.

IV.1. Data description

Our job posting data come from 202,049,236 electronic job postings in the US from January 1, 2010 to October 31, 2019 collected by Burning Glass Technologies, which describes itself as "the world's leading provider of real-time labor market data products and analysis". As described in Hershbein and Kahn (2018), this dataset aggregates, parses, and deduplicates millions of job postings into machine-readable form. Hershbein and Kahn provide a systematic analysis of the usefulness of this data set to understand the US job market. They demonstrate that the dataset is generally representative of the US job market, though somewhat biased to jobs requiring more skills than the average US job. Given our focus on research jobs and jobs that require technology skills, this bias is unlikely to affect our conclusions.

Our goal is to use job posting data to assess the GPT likelihood of emergent technologies. Gartner has documented the set of emergent technologies in their "hype cycle" lists since 1995. We collect all technologies listed in the Gartner "hype cycle" from 1995 to 2019 and identify the subset of technologies that are also listed as skills in the Burning Glass job posting data for a total of 21 technologies. Some technologies listed in the Gartner "hype cycle" are too broad to be mapped to skills in the Burning Glass data (e.g., tablets, mobile phones and drones), while others refer to broad technology concepts that also do not map to specific job skills (e.g., smart workplaces, digital security and collective intelligence). Nevertheless, the 21 technologies we examine should be seen as a non-exhaustive list of emerging technologies.[3]

We identify jobs postings that demonstrate labor demand for a particular technology based on skills and skill clusters characterizing the postings. Each job posting in the Burning Glass database is associated with a set of skills. Burning Glass groups these skills into skill clusters. There are 17,422 skills and 644 skill clusters in the Burning Glass data. We define job postings representative of a certain technology based on the mention of that technology either as a skill or skill cluster in a job posting.[4] For example, we define a job as ML if the posting has at least one skill which is in the skill cluster of "machine learning." Appendix 1 provides the definitions we use for each technology, the job posting counts, and examples for each technology.

We also use the skill and skill cluster data to identify research job postings as distinct from non-research job postings. We define a posting as a research job if it includes at least one skill in the Burning Glass - defined research skill clusters ("research methodology", "laboratory research", "medical research" and "clinical research") to a total of 4,269,779 research job postings in the data. While a handful of these research job postings relate to background research and have little to do with corporate "research & development" that generates innovation, the skill clustering approach ensures that job postings are classified as research if the skills listed in the job posting describe the job functions of a researcher. This is different from classifying a job posting as research based on weaker indicators such as the title of the job postings. Such an approach could lead to false positives because there are instances of job postings where "research" is listed in the title although the position is not research in the scholarly sense (e.g., market research job postings). At the same time, this approach means several of the jobs classified as non-research might be reasonably seen as research job postings. While this means we may undercount research job postings, this will not affect the comparison between technologies as long as any potential undercounting of research jobs applies to all technologies in similar ways.

[3] We also explored alternative strategies for identifying emerging technologies. For example, we identified all technologies that appeared on the cover of *Science* or *Nature* between 2000 and 2015. This generated a subset of the technologies listed in Gartner (specifically machine learning, GIS, CRISPR, quantum computing, robotics, nanotechnology, internet of things, and cloud computing) in addition to one technology not listed (fracking).
[4] Some job postings list skills from multiple technologies. We classify these job postings as requiring both skills.

IV.2. GPT criteria - Measures

Our goal is to use the job posting data to assess whether the 21 emerging technologies are (1) used in many industries, (2) used in many research job postings, and (3) the research job postings are spread across many industry sectors. We evaluate widespread use by calculating the Gini coefficient on job postings by 3-digit NAICS,[5] from 2010 to 2019. We evaluate a technology's use in many research job postings using the two measures: the number of research job postings and the fraction of postings using the technology that are research-focused, from 2010 to 2019. We evaluate the ability of technologies to enable innovation in other sectors by calculating the Gini coefficient on research job postings by 3-digit NAICS, from 2010 to 2019.

IV.3. Benchmarking

We evaluate if the job posting measures we develop correlate with prevailing patent-based measures. We focus on the measures developed in Petralia (2020) because the approach is at the frontier of such efforts, builds and extends prior efforts, and eliminates the need to rely on forward citation data that would require a much longer observation period than available for emerging technologies. Specifically, at least an additional decade of forward-looking data would be needed to construct informative citation-based metrics.

Petralia's approach evaluates electricity and computer & communication patent classes from 1993 to 2014 to identify the classes that are relatively likely to be GPTs. The GPT-ness of patent classes is based on three measures that map onto the three criteria established in Bresnahan (2010):

1) Widespread use. Petralia (2020) describes this criterion as potential use in a wide variety of products and processes and measures it as the count of the number of 3-digit patent classes with patents that include keywords describing the GPT candidates of interest – electricity and computer & communications.

[5] "A NAICS (pronounced NAKES) Code is a classification within the North American Industry Classification System. The NAICS System was developed for use by Federal Statistical Agencies for the collection, analysis and publication of statistical data related to the US Economy." (https://www.naics.com/what-is-a-naics-code-why-do-i-need-one/). We include measures based on the 2-digit NAICS in Appendix 2. All our findings continue to hold using this alternative measure.

2) Capable of ongoing technical improvement. Petralia (2020) describes this criterion as wide scope for improvement and evaluates it through the growth in number of patents classified in the 3-digit patent classes representing the GPT candidates of interest – electricity and computer & communications.
3) Enables innovation in application sectors. Petralia (2020) describes this criterion as strong complementarity with existing and new technologies and measures it as the count of the number of other 3-digit patent classes that are listed in patents from the 3-digit patent classes representing the GPT candidates of interest – electricity and computer & communications.

Since Petralia's data and time period largely precedes ours, we cannot directly examine whether our measures predict his as provided in his paper. Therefore, we apply his method, using updated patent data (grant dates from 2010 to 2019) and our 21 categories of technologies. This required some minor changes to his approach. Specifically, Petralia (2020) can construct the first and the third measures distinct from one another because his approach is conditional on being able to identify patents describing the GPT candidates of interest based on both 3-digit patent classes and keywords. We identify patents for our technologies using keywords only because most of our technologies are not well-represented by 3-digit patent classes. For example, while there are 3-digit patent classes for telecommunications, no such classes are available for most other technologies on our list. Moreover, to our knowledge, there is no established approach to identifying a collection of patent classes that represents several of the technologies on our list. For example, in the most recent report on AI and ML patents, the USPTO does not rely on patent classes to identify ML patents, and instead uses a proprietary machine learning approach to identify the patents based on information contained in the patent text.[6]

We draw the keywords from the Burning Glass skills that we used to identify the relevant technologies in our Burning Glass job posting data. We search patents using these keywords in Google Patents to take advantage of the Google search engine that scans both the abstract and the full text of the patents.[7] We then match the collected patent IDs into PatentsView.org, a database that tracks detailed patent data, including 3-digit patent classes. From these patents data, it is then

[6] https://www.uspto.gov/sites/default/files/documents/OCE-DH-AI.pdf
[7] We exclude polymer science because the only keyword available from the Burning Glass data, "polymer science," is too generic to identify an informative set of patents that represents this technology.

straightforward to calculate the wide scope for improvement and the potential use in a variety of products and processes measures described in Petralia (2020). Specifically, we measure wide scope for improvement by the number of yearly patents granted for each technology over our observation period, 2010-2019. Petralia (2020) does not suggest a precise formula for measuring growth in patents. The paper cites work that uses either a count of patents over various time periods or a growth rate. In the application exercise included in his paper, Petralia calculates a growth rate over five years. Given the relative short length of our observation period, we focus on a yearly count of patents. We measure potential use in a variety of products and processes by counting the number of 3-digit patent classes under which each set of technology patents granted in a focal year are classified. We consider all 3-digit patent classes listed on each patent.

We benchmark our job-posting-based method against the Petralia (2020)-inspired patent-based method in two steps. First, we test the correlation between our measures and the patent-based ones. If our measures are suited to identify GPT likelihood, then they should be correlated with the more established patent-based measures for identifying GPTs. Table 1a and 1b show the correlations. In Table 1a, we include panel Poisson estimates with year fixed-effects and robust standard errors clustered at the year level. We chose models with year fixed effects because GPT evaluation methods are preoccupied with evaluating the cross-technology variation in GPT likelihood. The approach is in line with other quantitative studies that evaluate the GPT likelihood of certain technologies (Moser and Nicholas 2004; Feldman and Yoon 2012). Columns 1 and 2 show that the Gini coefficient on all job postings across industries is negatively correlated with the growth in the yearly number of patents and with more co-occurring patent classes. Since a lower Gini coefficient means more widespread use of the technology, this suggests a strong correlation between our measures and the Petralia-inspired patent measures. Columns 3 and 4 show that more research job postings are positively correlated with the patent-based measures, as expected. Columns 5 and 6 show a negative correlation between the percentage research job postings measure and the patent measures. Percentage job postings is the only measure for which we do not have a prior for the direction of the correlation. Last, columns 7 and 8 show that the Gini coefficient on research job postings is negatively correlated with the patent measures, as expected. Table 1b shows the same correlation patterns on a yearly basis.

Next, we test if our measures have five-year predictive power of the patent measures, over-and-above a lagged version of the patent measures. Because the Petralia-inspired patent measures are not citation-based, neither method requires a decade of hindsight. In other words, the hypothesized benefit of the predictive power of our method relative to the measures in Petralia is not related to the lag between patent grant year and the time of informative citation accumulation (i.e., about a decade (Hall and Trajtenberg, 2006; Graham and Iacopetta, 2014)), but to the lag between the time searching for technology skills in the labor market and the time being granted patents on that technology. The average patent evaluation period from application to grant is three years (e.g., Graham and Iacopetta 2014).

Table 2 shows the predictive power of our measures five-years out.[8] As before, all models include data from 2015 to 2019 (i.e., 2010-2019 data truncated based on lags as appropriate) and are panel Poisson with year fixed effects and robust standard errors clustered at the year level. The predictors are the job posting measures lagged five years and the patent-based measures lagged five years, in all pair-wise combinations. We include lagged patent measures because past patent trends are predictive of future patent trends. Hence, we want to check if our job posting measures have any predictive power above and beyond the temporal correlation between patent measures. We find that our job posting measures are predictive of patent trends five years later. The predictive power of our measures is strongest for the capable of ongoing technical improvement patent metric, and somewhat weaker for the widespread use measure, although directionally consistent. In Appendix 3, we show robustness to using four- and six-years lags. The same patterns persist for the patent count measure. For the number of patent classes measure, the results do not hold with six-year lags though they do hold and are somewhat stronger with four-year lags.

Taken together, Tables 1 and 2 suggest the job posting measures are strongly correlated with patent-based approaches for estimatinsg GPT-ness. Furthermore, the job posting measures predict future values of patent-based measures at least five years out over-and-above the predictive power of lagged patent measures.

[8] Descriptive statistics are provided in Appendix 3.

IV.3. An application

We apply our method to evaluate the GPT likelihood of the emerging technologies in our data. We start by examining the co-occurrence of the technologies in job postings to identify closely related technologies. There is no requirement in the Gartner process for the technologies to be mutually exclusive. Our methodology, however, requires us to draw clear lines between the technologies we analyze because we aim to assess whether specific technologies are relatively likely to be GPTs.

To draw clear lines between the technologies we analyze, we need to identify which technologies are most closely related and then choose a surrogate for each group of technologies. We identify closely related technologies using data on co-occurrence of technologies in job postings. If two technologies frequently appear in the same job postings, we argue that they are likely to represent the same underlying tools.[9] Table 3a presents the overlap in job postings in 2010; Table 3b presents the same overlap in 2019. Each number represents the fraction of job postings that mention the technology in the row that also mention the technology in the column. For example, Table 3a the bottom row of column 1 shows that 0.9% of RFID jobs also mention ML. In contrast, the last column of row 1 shows that 0.3% of ML jobs also mention RFID. As a reminder, we define a job posting to represent a technology if the posting has at least one skill requirement that represents the technology. Thus, job postings can be classified as representing more than one technology if the posting lists skills representing multiple technologies. Thus, Tables 3a and 3b are not symmetric because it is possible that there is a larger or smaller fraction of technology A jobs that also list technology B skills, than the fraction of technology B jobs that also list technology A skills.

We observe that there is a strong overlap between ML, BI, big data, data mining, data science, and NLP. For each of these six technologies, the others are generally in the top five in terms of co-occurrence in both 2010 and 2019. For example, the top five co-occurring technologies in job postings that list ML in 2010 are the other five listed technologies. In 2019, it is big data, data mining, data science, NLP, and cloud computing. In 2019, 36% of data science job postings

[9] From job posting data alone, it is difficult to assess whether the combined appearance of these technology skill clusters is a result of (1) different names for the same underlying technology, (2) gradual substitution of an earlier technology for a new one, or (3) complementary technologies. For our purposes, the key takeaway is that these technologies cannot be considered separately when assessing whether they are GPTs.

mention ML, 54% of NLP job postings mention ML, and 27% of data mining job postings mention data science. Clearly, these technologies are closely related. While other technologies are sometimes connected, the connections are not symmetric as with 13% of RFID job postings mentioning telecommunications in 2010, but only 0.1% of telecommunications job postings mention RFID. Other technologies that may seem related on the surface, such as robotics and machine learning, do not have a strong overlap.

Of these six technologies that are most connected, we first focus on ML because it is the technology most hypothesized to be a GPT (e.g., Brynjolfsson, Rock, and Syverson 2019; Cockburn, Henderson, and Stern 2019; Trajtenberg 2019). Later, we show the results for the other five technologies and demonstrate that these technologies display similar patterns. Put differently, and consistent with prior work on other major technological changes (e.g., Rosenberg 1963 on technological change in 19th century machine tools), our results will suggest that it is not appropriate to evaluate the relative GPT likelihood of a technology independent of its cluster of related technologies.

IV.3.1. Evaluating the three criteria for GPT likelihood

For ease of exposition, we focus on comparing the values for the job posting-based measures in 2010 and 2019, the first and the last year of our observation period. We have no reason to expect anything other than linear trends over time. Indeed, the patterns from 2011 through 2018 change roughly monotonically over time. We include data for all years in Appendix 4.

a) Widespread use

We evaluate evidence for widespread use by calculating the Gini coefficients by 3-digit NAICS (Table 4). We show the data for 2019 and 2010, and rank order the technologies by the values in 2019. We also show data on the total number of job postings that mention the technologies in 2019 and 2010.

Column 1 shows the Gini coefficients in 2019. Telecommunications and robotics are most widespread, followed by cloud computing, SOA, and ML. The technologies that are not widespread are Web 2.0, VR, polymer science, nanotechnology, and especially CRISPR and quantum computing. Column 2 shows the same values for 2010. ML, robotics, 3D printing, and

IoT were not widely diffused in the earlier period. In contrast, Web 2.0 was less concentrated in 2010 than in 2019. This change over time demonstrates that, had we undertaken this exercise in 2010, ML would not have appeared as a likely GPT in terms of widespread use. The contrast in number of jobs between columns 3 and 4 explains why. It was too early in 2010. ML was not yet widespread. This is consistent with the timing of the commercial development of deep learning, the underlying technique that propelled the ML hype and GPT speculation. The commercial opportunities in deep learning became apparent in 2012 because of the ImageNet competition that year (Agrawal, Gans, and Goldfarb 2018). This suggests some caution in interpreting the results on those technologies that are relatively immature and not widespread such as quantum computing. It is possible that over time, they will become more widespread.

b) Many research job postings

We capture widespread use in research by the total number of research job postings and by the percentage of research job postings out of total job postings per technology (Table 5). These job postings include all research-related jobs that mention the technology skill cluster. We order the data by the total number of research job postings in 2019 (column 1).

ML has the most research job postings in 2019, followed by cloud computing, robotics, and telecommunications. These four technologies also have the most job postings overall as seen in Table 4. Column 3 therefore examines research jobs as a fraction of total job postings. Of these four relatively widely diffused technologies, only ML has over 10% of job postings as research jobs. Of the other categories with a high percentage of job postings as research jobs, only 3D printing was not near the bottom of Table 4 in terms of widespread use.

The data from 2010 provide further insight. Even as the number of job postings in ML and cloud computing grew substantially, the proportion of those postings that are research jobs remained relatively flat (columns 3 and 4). These technologies do not seem to have moved away from research as they have diffused. For ML in particular, this suggests that research use is a key aspect of its application.

Overall, we observe that ML is the only technology near the top in both number and percentage of research job postings over time. The other technologies are near the top in one or the other.

However, this does not suggest ML is more likely to be a GPT in this dimension. We do not know of a formal literature on whether "potential for innovation" focuses on whether there are many researchers working on a technology or whether a large fraction of people working on a technology are researchers. As such, we cannot reject cloud computing or telecommunications as potential GPTs based on the "potential for innovation" criterion. They satisfy it under one metric but not the other.

c) Research job postings across application sectors

Table 6 examines whether the technologies we study are used for research in a range of application sectors by repeating the 3-digit NAICS analysis restricted to research job postings. The data is again ordered by the 2019 values.

The set of technologies that were widespread across research job postings in 2019 (column 1) includes cloud computing, robotics, IoT, telecommunications, and especially ML. We interpret this to suggest that these five technologies are relatively likely to be GPTs, as defined by the criterion of enabling innovation in applications sectors. Of all of the technologies examined, only telecommunications was widespread in many research job postings in 2010 (column 2). Thus, of all the technologies listed in the Gartner hype cycle since 1995 that are also listed as Burning Glass skill clusters, only telecommunications displayed somewhat similar values in 2010 to ML, cloud, and robotics in 2019, suggesting that the technology had a high GPT likelihood in 2010.

d) Ranking the three GPT criteria

Table 7 presents the relative rank of the technologies in each of the three GPT criteria in 2019. While there is no established formula for weighing our four measures evaluating the three criteria for GPTs, we interpret our findings to suggest that, of all emerging technologies in our data, ML is relatively likely to be a GPT. ML ranks consistently near the top in all measures and ranks at the top in most research job postings, and degree of widespread of those postings across industry sectors. Since innovation in applications sectors is a key distinguishing feature of GPTs (Bresnahan and Trajtenberg 1995), we view the widespread use in research as particularly important. Cloud computing, robotics, and telecommunications are also relatively prevalent in research job postings and widespread research use. Although their fraction of research job postings is relatively low, we cannot exclude the possibility that these technologies are potential GPTs.

It is also informative to focus on the 2010 data. In that year, only telecommunications was relatively widespread, with a large number of research job postings in a wide range of industries. In other words, looking at the 2010 data, telecommunications stood out as a possible GPT, though with values in the research-related categories that were less likely to be GPTs than ML did in 2019. Telecommunications has been previously identified as a GPT (e.g., Liao et al., 2016; Strohmaier and Rainer, 2016; Petralia, 2020). We view this as supportive that our approach is informative in evaluating a relative GPT likelihood of emerging technologies. Moreover, the fact that ML looks relatively more likely to be a GPT in 2019 then telecommunications suggests that ML might be a GPT. However, we caution against drawing a bold conclusion about ML since telecommunications was likely relatively more mature compared to ML and the other emerging technologies during our observation period.

All other ranking data suggest that, aside from ML, cloud computing, robotics, and telecommunications, most of the other technologies listed are much less likely to be on the path to becoming GPTs in their current form.

IV.3.2. The three GPT criteria for the ML cluster

In section IV.3., we identified six technologies that appear in many of the same job postings and that all relate to the use of data: ML, BI, big data, data mining, data science, and NLP. In Table 8 (a, b, c), we repeat the above analysis for these six technologies. The results show that all of the technologies other than NLP have aspects of GPTs. Furthermore, ML, data science, and to a lesser extent big data have a large number of research job postings. We interpret this to suggest that, taken together, these technologies are likely to be a GPT, compared to the other technologies examined above. Furthermore, there was a notable change between 2010 and 2019. In 2010, none of these data-focused technologies had all three features of GPTs: widespread, with a large number or proportion of research jobs, and used for research in a wide variety of industries. By 2019, data science and ML clearly meet these criteria relative to the other technologies examined. Big data, data mining, and BI also can be seen as relatively likely to be GPTs. Prior literature highlighted

ML as a GPT; however, Table 8 suggests an important nuance. It is this cluster of technologies that together are relatively more likely to represent a GPT.

V. Discussion and Conclusion

GPTs are different from other technologies. They hold potential for substantial economic impact, but the impact is not guaranteed. Economic actors need to engage in appropriate strategies that solve the canonical GPT problem; the large productivity gains from GPTs occur when there is a coordinated positive feedback loop in innovation between producing and application industries. Thus, application-industry organizations looking to benefit from GPTs need to develop processes for research collaborations with industry, with academia and with companies in producing industries. Furthermore, these processes take time, are costly, and the productivity benefits may require several years to appear. Some of these benefits may accrue to intermediaries or end-users rather than to the innovators (Gambardella et al., 2020).

It is then beneficial to have an early sense if a technology is GPT. When new technologies appear, it is not unusual to find claims that these technologies are general purpose. The speculations are generally informed by observed widespread engagement with the emerging technologies. Although this is just one characteristic of GPTs, the speculations emerge because a more comprehensive evaluation of GPT-ness is generally available with a lag, after the benefits of the technology have been realized. We propose an approach that evaluates all three GPT criteria (Bresnahan, 2010) to reveal information about the GPT likelihood of emerging technologies while they emerge.

Empirically, we leverage the insight that early trends of technology diffusion and adoption can be observed in job posting data (Tambe and Hitt, 2012a,b). We construct measures that capture the three GPT criteria for a set of 21 emerging technologies that we can map to job posting skills. We benchmark our approach against the latest patent-based quantitative method for evaluating GPT likelihood inspired by Petralia (2020). We find that our measures of whether an emerging technology is likely to be a GPT are correlated with the patent-based measures. Moreover, we find that our measures predict future values of the patent-based ones over-and-above the predictive power of lagged patent measures.

Next, we apply our method to compare the relative GPT likelihood of our set of emerging technologies. Our results suggest that a suite of data-focused technologies—often represented by ML—are relatively likely to be a GPT. Cloud computing and robotics also display some characteristics of GPTs. We based this interpretation, at least in part, on a comparison with telecommunications. Telecommunications, an established GPT (e.g., Liao et al., 2016; Strohmaier and Rainer, 2016; Petralia, 2020) has a high GPT likelihood throughout our observation period, although less than the ML cluster in 2019.

Our results also show that several technologies are unlikely to be general purpose. For some of these technologies, this is unsurprising; both scholars and practitioners did not engage in speculations about the general purpose likelihood of these technologies. Thus, we view the result that RFID, Web 2.0, and SOA are relatively unlikely to be GPTs as evidence that our method has power to distinguish between technologies. For other technologies, the results suggest that some claims of technologies being general purpose seem unlikely in their current stage; notably for 3D printing, nanotechnology, IoT, and blockchain.

An important limitation of this particular application of our proposed method is that our measurement focuses on a particular time window in the diffusion and adoption of emerging technologies. It does not capture GPTs that have already widely diffused and no longer have significant innovation in the applications sectors as captured by skills in job postings. This is the case for GPTs such as electricity and the internet. The application also does not capture technologies in very early stages. For example, quantum computing may someday become a GPT, given that it represents the next generation of computational devices, but the technology is not yet sufficiently developed. A longer-run observation period can mitigate some of these issues. With a longer observation period, it might be possible to exploit the trajectory of job postings over time, or the relative importance of research jobs and breadth of industries to identify the potential of very early stage technologies to become GPTs. It is reasonable to assume that job posting data will continue to accumulate, allowing future research to derive more nuanced insights from and further develop this GPT evaluation method.

Another limitation of our application example is that it does not prove whether an emerging technology is a GPT, but rather provides a relative GPT likelihood ranking of emerging technologies. In essence, during our observation period, we do not observe an objective threshold above which a technology could be labeled general purpose. A longer-run dataset can solve the issue if the data covers the period of emergence of at least one established GPT because the established GPTs would provide the threshold. The closest we get in our dataset is to compare against telecommunications, a technology that is identified as general purpose in other studies (e.g., Liao et al. 2016; Strohmaier and Rainer 2016; Petralia 2020). However, our dataset does not cover the emergence period of telecommunications. This suggests that, when employed in future studies, the method we propose will likely render better insights for future emerging technologies.

A longer-run data set would also enable better benchmarking of our method. Because the ten-year observation period does not capture job posting data about established GPTs at a similar stage of development, we benchmark our approach against patent-based measures inspired by established quantitative approaches to identify GPTs over the same period of time. Future research could revisit the issue once enough time has passed to enable the power of hindsight. It is important to note that a better assessment of the predictive power of the job posting approach, relative to a patent-based approach, is not sufficient in of itself. The patent-based methods also rely on relative assessments for GPT likelihood; they do not compare against an objective threshold to identify GPTs (e.g., Moser and Nicholas, 2004, Feldman and Yoon, 2012). The advantage is that available patent data cover a longer period of time than the job posting data in this paper. Future research should compare across methods to analyze if an objective threshold for GPTs could be identified.

There are also several measurement limitations when using job posting data. Job postings do not represent hires. Also, job postings are not directly capturing technology usage; this measurement concern might be exacerbated for skills that are listed to mainly signal intent for technology usage. The job posting methodology is also limited to evaluating technologies that can be mapped to skills. Some technologies are too broad to be mapped to job posting e.g., tablets, mobile phones and drones, while others refer to broad technology concepts rather than products e.g., smart workplaces, digital security, and collective intelligence.

More broadly, focusing on job postings has some advantages over established approaches exploiting patent data, but also drawbacks. This suggests the possibility that other data sources could be employed to balance the disadvantages. These could include unstructured data, such as text; advances in natural language processing techniques lower the cost of extracting information about the diffusion and adoption of emerging technologies from such data types. For example, future research could exploit data about entrepreneurship patterns, VC investment, news articles describing engagement with emerging technologies, and search trends. Each type of data will have some advantages and some limitations in tracing technology diffusion and adoption, but some will likely be complementary to job postings and patents. Hence, the combined approaches should provide a better understanding of where each type of data might succeed or fail and hence, an overall more accurate GPT assessment of emerging technologies.

Despite these limitations, even relative comparisons using our approach have an action-focused interpretation. An emerging strategy literature focuses on 'enabling technologies' (e.g. Teece 2018; Gambardella et al 2020; Rathje and Katila 2021), which Rathje and Katila (2021, p. 1) call "junior GPTs". In this literature, technologies can be represented on a continuum of their propensity to enable other innovation, and GPTs represent the most important enabling technologies (Rathje and Katila 2021)**.** Our ranking stacks technologies based on their enabling propensity. Therefore, an alternative interpretation of our method is that the higher the ranking, the higher the enabling ability and hence relative GPT likelihood. Having an early sense of enabling propensity is important because many of the managerial consequences of GPTs apply to enabling technologies, albeit at a smaller scale. For example, managers need to anticipate the innovation externalities that are needed by collaborating with other organizations and understand that any substantial benefits depend on further innovations over a relatively long time horizon. Some of these benefits may accrue to others, whether intermediaries who specialize in key complements or consumers (Gambardella et al. 2020).

Certainly, these insights would be strengthened by future research efforts towards improving our ability to identify GPTs early on. The various limitations of our method suggest there is still considerable work to do, related to alternative data sources, longer time horizons, and alternative ways to combine data. Gambardella et al. (2020) note that the benefits of GPTs tend not to accrue to those investing in the complex and costly innovation process required to generate value. Thus, absent appropriate policy interventions, innovators likely underinvest in GPTs relative to a social optimum. These dynamics are clearly complicated by a poor understanding of which technologies are general purpose.

References

Agrawal, Ajay, Joshua Gans, Avi Goldfarb. 2018. *Prediction Machines: The Simple Economics of Artificial Intelligence*. Harvard Business Review Press, Boston MA.

Allen, Robert. C. 1983. Collective Invention. *Journal of Economic Behavior & Organization* 4(1): 1-24.

Aral, Sinan, Erik Brynjolfsson, Lynn Wu. 2012. Three-Way Complementarities: Performance Pay, Human Resource Analytics, and Information Technology. *Management Science* 58(5), 913-931.

Attewell, P., 1992, Technology diffusion and organizational learning: the case of business computing, *Organization Science* 3, 1-19.

Bresnahan, Timothy. 2010. General Purpose Technologies. In Bronwyn Hall, Nathan Rosenberg Eds. *Handbook of the Economics of Innovation*. Chapter 18, 761-791.

Bresnahan, Timothy, Erik Brynjolfsson, Lorin Hitt. 2002. Information Technology, Workplace Organization, and the Demand for Skilled Labor: Firm-Level Evidence, *The Quarterly Journal of Economics*, 117(1), 339–376.

Bresnahan, T., S. Greenstein. 1996. Technical Progress and Co-invention in Computing and in the Uses of Computers. *Brookings Papers on Economic Activity, Microeconomics* 1996; 1-83.

Bresnahan, T., M. Trajtenberg. 1995. General Purpose Technologies 'Engines of Growth'? *Journal of Econometrics*. 65, 83-108.

Brynjolfsson, Erik, Daniel Rock, Chad Syverson. 2019. Artificial Intelligence and the Modern Productivity Paradox: A Clash of Expectations and Statistics. In Agrawal, Gans, Goldfarb Eds. *The Economics of Artificial Intelligence: An Agenda*. University of Chicago Press, 23-57.

Cassiman. B., Veugelers. R., 2006. In search of complementarity in innovation strategy: Internal R&D and External Knowledge Acquisition. Management Science 52(1): 68-82

Choi, Jongmin. 2018. The rise of 3D printing and the role of user firms in the U.S.: Evidence from patent data. *Technology Analysis & Strategic Management* 30(1), 1195-1209.

Cockburn, Iain, Rebecca Henderson, Scott Stern. 2019. The Impact of Artificial Intelligence on Innovation. In Agrawal, Gans, Goldfarb Eds. *The Economics of Artificial Intelligence: An Agenda*. University of Chicago Press, 115-146.

Conti, Raffaele, Alfonso Gambardella, Elena Novelli. 2019. Specializing in Generality: Firm Strategies When Intermediate Markets Work. *Organization Science* 30(1), 126-150.

Cohen, W. 2010. Fifty years of empirical studies of innovation activity and performance. In the Handbook of the Economics of Innovation, 1, 129-213.

Cui, Yao, Yeşim Orhun, Izak Duenyas. 2018. How Price Dispersion Changes When Upgrades Are Introduced: Theory and Empirical Evidence from the Airline Industry. *Management Science* 65(8), 3835-3852.

David, Paul. 1990. The Dynamo and the Computer: An Historical Perspective on the Modern Productivity Paradox. *American Economic Review Papers & Proceedings* 80(2), 355-361.

Edquist, Harald, Peter Goodridge, Jonathan Haskel. 2019. The Internet of Things and Economic Growth in a Panel of Countries. *Economics of Innovation and New Technology*.

Etro, Federico. 2009. The Economic Impact of Cloud Computing on Business Creation, Employment, and Output in Europe: An Application of the Endogenous Market Structures Approach to a GPT Innovation. *Review of Business and Economics* 0(2): 179-208.

Feldman, Maryann, JiWoong Yoon. 2012. An empirical test for general purpose technology: an examination of the Cohen-Boyer rDNA technology. *Industrial and Corporate Change* 21(2), 249-275.

Filippova, Evgeniia. 2019. Empirical Evidence and Economic Implications of Blockchain as a General Purpose Technology. IEEE Technology & Engineering Management Conference, Atlanta, GA.

Fleder, Daniel, Kartik Hosanagar. 2009. Blockbuster Culture's Next Rise or Fall: The Impact of Recommender Systems on Sales Diversity. *Management Science* 55(5):697-712.

Forti, Enrico, Federico Munari, Chunxiang Zhang. 2019. “Does VC Backing Affect Brand Strategy in Technology Ventures?” *Strategic Entrepreneurship Journal* 14(2): 265-286.

Gambardella, Alfonso, Sohvi Heaton, Elena Novelli, David Teece. 2020. Profiting From Enabling Technologies. Working paper, City University of London Business School.

Graham, Stuart J. H., Maurizio Iacopetta. "Nanotechnology and the Emergence of a General Purpose Technology." *Annals of Economics and Statistics*, no. 115/116 (2014): 25-55.

Greenwood, Jeremy, Mehmet Yorukoglu. 1997. "1974" Carnegie-Rochester Conference Series on Public Policy 46: 49-95.

Hall, Bronwyn H., Manuel Trajtenberg, 2006. "Uncovering General Purpose Technologies with Patent Data," Chapters, in: Cristiano Antonelli & Dominique Foray & Bronwyn H. Hall & W. Edward Steinmueller (ed.), New Frontiers in the Economics of Innovation and New Technology, chapter 14, Edward Elgar Publishing.

Hershbein, Brad, Lisa B. Kahn. 2018. "Do Recessions Accelerate Routine-Biased Technological Change? Evidence from Vacancy Postings." *American Economic Review* 108(7): 1737-72.

Jaffe, Adam, B., Gaetan de Rassenfosse. 2019. "Patent Citation Data in Social Science Research: Overview and Best Practices." In Ben Depoorter, Peter Menell and David Schwartz *Research Handbook on the Economics of Intellectual Property Law* Volume 2, Chapter 2, 20-46.

Jovanovic, Boyan, Peter Rousseau. 2005. General Purpose Technologies. In Philippe Aghion, Steven N. Durlauf Eds. *Handbook of Economic Growth* Volume 1B, Chapter 18, 1181-1224.

Lipsey, Richard, Kenneth Carlaw, Clifford Bekar. 2005. *Economic Transformations: General Purpose Technologies and Economic Growth*. Oxford University Press, Oxford UK.

Moser, Petra, Tom Nicholas. 2004. Was electricity a general purpose technology? Evidence from historical patent citations. *American Economic Review Papers & Proceedings* 94(2), 388-394.

National Academies of Sciences (NAS), Engineering, and Medicine, "Quantum Computing: Progress and Prospects" (Washington, DC: The National Academies Press, 2019).

Nuvolari, Alessandro. 2004. Collective Invention During the British Industrial Revolution: The Case of the Cornish Pumping Engine. *Cambridge Journal of Economics* 28(3): 347-363.

Petralia, Sergio. 2020. Mapping general purpose technologies with patent data. *Research Policy* 49(7).

Rathje, Jason, Katila Riitta. 2021. Enabling Technologies and the Role of Private Firms: A Machine Learning Matching Analysis. *Strategy Science*, forthcoming.

Rosenberg, Nathan. 1963. Technological Change in the Machine Tool Industry, 1840-1910. *Journal of Economic History*, 414-443.

Tambe, Prasanna, Lorin Hitt. 2012a. Now IT's Personal: Offshoring and the Shifting Skill Composition of the U.S. Information Technology Workforce. *Management Science* 58(4), 678-695.

Tambe, Prasanna, Lorin Hitt. 2012b. The Productivity of Information Technology Investments: New Evidence from IT Labor Data. *Information Systems Research* 23(3-part 1), 599-617.

Tambe, Prasanna, Lorin Hitt, Erik Brynjolfsson. 2012. The Extroverted Firm: How External Information Practices Affect Innovation and Productivity. *Management Science* 58(5), 843-859.

Trajtenberg, Manuel. 2019. AI as the next GPT: A Political Economy Perspective. In Agrawal, Gans, Goldfarb Eds. *The Economics of Artificial Intelligence: An Agenda*. University of Chicago Press, 175-186.

Table 1a: Correlation between our job posting measures and the patent-based measures.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
2010-2019	Count of patents	Co-occurring patent classes	Count of patents	Co-occurring patent classes	Count of patents	Co-occurring patent classes	Count of patents	Co-occurring patent classes
Widespread use	-3.672***	-1.485***						
(Gini by 3-digit industry)	(0.201)	(0.080)						
Many research jobs			0.007***	0.004***				
(count of research jobs in hundreds)			(0.002)	(0.001)				
Disproportionate research jobs					-5.364***	-0.979***		
(fraction of research jobs)					(0.443)	(0.148)		
Widespread research use							-3.346**	-1.675***
(Gini by 3-digit industry)							(0.578)	(0.195)
LL	-793.32	-9,288.82	-851.64	-9,709.74	-824.43	-9,977.35	-829.98	-9,365.55
Observations	200	200	200	200	200	200	200	200

Note: Unit of observation is the technology-year. Dependent variables are based on patent data 2010-2019. Independent variables are based on job posting data 2010-2019. All columns show Poisson regressions with year fixed-effects and robust standard errors clustered at the year level. In columns 1, 3, 5 and 7 the dependent variable is count of patents in thousands. In columns 2, 4, 6 and 8 the dependent variable is count of co-occurring patent classes. *significant at 10%, **significant at 5%, ***significant at 1%

Table 1b: Correlation between our job posting measures and the patent-based measures, by year.

Panel A:

DV=Count of patents	Widespread use (Gini by 3-digit industry)	Many research jobs (count of research jobs in hundreds)	Disproportionate research jobs (fraction of research jobs)	Widespread research use (Gini by 3-digit industry)
2010	-5.435***	0.060***	-1.553	-9.116***
	(1.226)	(0.010)	(2.820)	(1.426)
2011	-3.770***	0.047***	-4.143*	-10.950***
	(0.970)	(0.007)	(2.487)	(1.738)
2012	-3.971***	0.040***	-5.466**	-6.760***
	(0.903)	(0.006)	(2.466)	(1.270)
2013	-4.534***	0.033***	-6.060***	-2.283**
	(0.878)	(0.005)	(2.189)	(1.081)
2014	-3.830***	0.023***	-6.077***	-1.241
	(0.858)	(0.005)	(2.045)	(1.019)
2015	-3.632***	0.015***	-5.606***	-2.952***
	(0.811)	(0.003)	(1.725)	(0.793)
2016	-3.833***	0.011***	-6.701***	-4.988***
	(0.786)	(0.003)	(1.842)	(0.917)
2017	-3.356***	0.007***	-7.016***	-2.678***
	(0.723)	(0.002)	(1.776)	(0.734)
2018	-3.532***	0.004***	-5.294***	-3.037***
	(0.685)	(0.001)	(1.581)	(0.783)
2019	-2.868***	0.004***	-3.976***	-2.061***
	(0.580)	(0.001)	(1.137)	(0.540)
Obs.	20	20	20	20

Note: Unit of observation is the technology. Dependent variables are based on patent data. Independent variables are based on job posting data. All columns show cross-sectional Poisson regressions. The dependent variable is count of patents in thousands. *significant at 10%, **significant at 5%, ***significant at 1%

Panel B:

DV= Co-occurring patent classes	Widespread use (Gini by 3-digit industry)	Many research jobs (count of research jobs in hundreds)	Disproportionate research jobs (fraction of research jobs)	Widespread research use (Gini by 3-digit industry)
2010	-1.941***	0.021***	3.789***	-3.527***
	(0.125)	(0.001)	(0.264)	(0.164)
2011	-1.456***	0.016***	0.472**	-4.084***
	(0.101)	(0.001)	(0.214)	(0.184)
2012	-1.572***	0.013***	-0.186	-2.170***
	(0.103)	(0.001)	(0.204)	(0.153)
2013	-1.883***	0.012***	-1.273***	-1.496***
	(0.111)	(0.001)	(0.165)	(0.143)
2014	-1.729***	0.010***	-1.379***	-1.470***
	(0.113)	(0.001)	(0.141)	(0.133)
2015	-1.343***	0.006***	-1.067***	-1.165***
	(0.104)	(0.000)	(0.109)	(0.107)
2016	-1.454***	0.005***	-1.191***	-1.810***
	(0.105)	(0.000)	(0.104)	(0.120)
2017	-1.464***	0.004***	-1.495***	-1.427***
	(0.102)	(0.000)	(0.118)	(0.102)
2018	-1.257***	0.002***	-1.017***	-1.569***
	(0.096)	(0.000)	(0.106)	(0.109)
2019	-1.150***	0.002***	-1.004***	-1.227***
	(0.087)	(0.000)	(0.098)	(0.083)
Obs.	20	20	20	20

Note: Unit of observation is the technology. Dependent variables are based on patent data. Independent variables are based on job posting data. All columns show cross-sectional Poisson regressions. The dependent variable is count of co-occurring patent classes. *significant at 10%, **significant at 5%, ***significant at 1%

Table 2: Job posting measures predict patent-based measures five years later

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
2010-2019	Count of patents	Co-occurring patent classes	Count of patents	Co-occurring patent classes	Count of patents	Co-occurring patent classes	Count of patents	Co-occurring patent classes
Widespread use lagged 5 years	-1.360***	-0.044						
	(0.412)	(0.064)						
(Gini by 3-digit industry)								
Many research jobs lagged 5 years			0.010***	-0.001				
			(0.003)	(0.001)				
(count of research jobs in hundreds)								
Disproportionate research jobs lagged 5 years					-1.838***	-0.097		
					(0.321)	(0.357)		
(fraction of research jobs)								
Widespread research use lagged 5 years							-2.518***	-0.238***
							(0.735)	(0.075)
(Gini by 3-digit industry)								
Dependent variable lagged 5 years	0.090***	0.004***	0.086***	0.004***	0.094***	0.004***	0.092***	0.003***
	(0.011)	(0.000)	(0.010)	(0.000)	(0.014)	(0.000)	(0.008)	(0.000)
LL	-290.30	-2,231.45	-290.41	-2,228.95	-293.24	-2,230.97	-280.76	-2,224.45
Observations	100	100	100	100	100	100	100	100

Note: Unit of observation is the technology-year. Dependent variables are based on patent data 2010-2019. Independent variables are based on job posting data 2010-2019. All columns show Poisson regressions with year fixed-effects and robust standard errors clustered at the year level. In columns 1, 3, 5 and 7 the dependent variable is count of patents in thousands. In columns 2, 4, 6 and 8 the dependent variable is count of co-occurring patent classes. *significant at 10%, **significant at 5%, ***significant at 1%

Table 3a: Overlap of technology in job postings in 2010

	ML	BI	Big Data	Data Mining	Data Science	NLP	Cloud	Telecom	GIS	Quantum	Robotics	Nanotech	IoT	CRISPR	VR	3D Print	Polymer	Block-chain	Web2.0	SOA	RFID
ML	100	8.4	10.1	40.2	27.9	13.8	5.3	3.5	0.9	0.1	3.0	0.0	0.0	0.0	0.7	0.0	0.0	0.0	1.2	0.5	0.3
BI	0.3	100	0.4	3.5	1.1	0.1	2.1	2.9	0.6	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.4	1.5	0.0
Big Data	7.5	9.3	100	10.0	3.4	3.9	13.7	1.9	0.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	2.2	0.9	0.0
Data Mining	8.6	22.7	2.9	100	10.9	2.6	2.6	4.8	1.1	0.0	0.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.7	1.0	0.2
Data Science	14.5	16.7	2.3	26.3	100	3.4	2.4	3.5	1.4	0.0	0.5	0.0	0.0	0.0	0.1	0.0	0.1	0.0	0.4	0.5	0.0
NLP	21.9	3.8	8.3	19.2	10.6	100	7.2	20.3	0.5	0.1	0.3	0.3	0.0	0.0	0.0	0.0	0.0	0.0	1.9	0.7	0.1
Cloud	0.4	5.2	1.5	1.0	0.4	0.4	100	7.1	0.4	0.0	0.1	0.0	0.1	0.0	0.0	0.0	0.0	0.0	1.8	1.2	0.1
Telecom	0.1	2.7	0.1	0.7	0.2	0.4	2.6	100	0.9	0.0	0.1	0.0	0.2	0.0	0.0	0.0	0.0	0.0	0.3	0.4	0.1
GIS	0.3	5.9	0.1	1.7	0.9	0.1	1.6	10.3	100	0.0	0.2	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.5	0.8	0.1
Quantum	6.3	0.0	0.0	1.1	5.3	5.3	5.3	6.3	0.0	100	0.0	1.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Robotics	1.2	0.6	0.0	0.3	0.4	0.1	0.4	1.7	0.3	0.0	100	0.2	0.0	0.0	0.2	0.0	0.0	0.0	0.0	0.1	0.2
Nanotech	0.1	0.8	0.0	0.2	0.0	1.1	0.3	0.6	0.0	0.1	2.6	100	0.0	0.0	0.0	0.1	1.2	0.0	0.6	0.0	0.0
IoT	0.1	2.8	0.0	0.6	0.1	0.0	3.0	23.7	0.1	0.0	0.2	0.0	100	0.0	0.0	0.0	0.0	0.0	0.2	1.8	1.6
CRISPR	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100	0.0	0.0	0.0	0.0	0.0	0.0	0.0
VR	10.1	0.8	0.0	2.6	3.8	0.4	2.0	4.3	2.4	0.0	6.5	0.0	0.0	0.0	100	0.0	0.0	0.0	2.4	0.0	1.4
3D Print	0.0	0.0	0.0	0.0	0.0	0.0	0.0	5.1	0.5	0.0	2.0	0.5	0.0	0.0	0.0	100	0.0	0.0	1.0	0.0	0.0
Polymer	0.1	0.3	0.0	0.1	0.6	0.0	0.0	1.3	0.0	0.0	0.2	1.0	0.0	0.0	0.0	0.0	100	0.0	0.0	0.0	0.0
Blockchain	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100	0.0	0.0	0.0
Web2.0	0.4	4.5	1.0	1.2	0.3	0.4	7.8	3.5	0.5	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.0	100	2.9	0.0
SOA	0.2	15.4	0.4	1.6	0.3	0.2	5.3	4.8	0.9	0.0	0.1	0.0	0.1	0.0	0.0	0.0	0.0	0.0	2.8	100	0.1
RFID	0.9	3.7	0.0	3.2	0.2	0.2	2.6	13.4	0.7	0.0	1.4	0.0	1.1	0.0	0.3	0.0	0.0	0.0	0.2	1.2	100

Note**:** Each number represents the fraction of job postings that mention the technology in the row that also mention the technology in the column. The shade of blue darkens with higher numbers.

Table 3b: Overlap of technology in job postings in 2019

	ML	BI	Big Data	Data Mining	Data Science	NLP	Cloud	Telecom	GIS	Quantum	Robotics	Nanotech	IoT	CRISPR	VR	3D Print	Polymer	Block-chain	Web2.0	SOA	RFID
ML	100	9.8	32.5	12.8	44.8	14.1	19.2	1.9	0.6	0.6	6.3	0.0	5.7	0.1	1.5	0.3	0.0	2.2	0.0	0.7	0.1
BI	4.4	100	10.7	5.9	9.0	0.6	10.3	2.0	0.5	0.1	0.7	0.0	0.9	0.0	0.1	0.0	0.0	0.3	0.0	0.7	0.0
Big Data	19.1	14.0	100	6.5	21.3	3.8	27.6	2.8	0.5	0.2	0.9	0.0	3.2	0.0	0.2	0.0	0.0	0.7	0.1	1.0	0.2
Data Mining	20.7	21.1	17.8	100	26.6	6.2	7.0	2.2	1.1	0.3	1.1	0.0	1.1	0.1	0.2	0.0	0.0	0.4	0.1	0.2	0.1
Data	36.2	16.1	29.3	13.3	100	7.6	13.5	1.9	1.0	0.2	1.5	0.0	2.7	0.1	0.3	0.1	0.0	0.7	0.0	0.3	0.2
NLP	54.3	5.4	25.1	14.7	36.3	100	14.2	3.9	0.4	0.5	4.4	0.0	3.1	0.0	0.4	0.0	0.0	2.1	0.1	0.5	0.0
Cloud	4.9	5.9	12.1	1.1	4.3	1.0	100	4.2	0.3	0.2	0.6	0.0	2.7	0.0	0.1	0.0	0.0	0.6	0.1	1.2	0.0
Telecom	0.7	1.6	1.8	0.5	0.9	0.4	6.1	100	0.7	0.1	0.4	0.0	1.3	0.0	0.1	0.0	0.0	0.1	0.1	0.1	0.2
GIS	2.0	3.5	2.5	2.2	3.8	0.3	3.9	6.2	100	0.0	0.6	0.0	0.5	0.0	0.1	0.1	0.0	0.0	0.0	0.6	0.2
Quantum	14.9	4.3	8.0	4.3	5.6	2.9	22.4	3.0	0.2	100	4.3	0.1	74.3	0.1	0.5	0.0	0.0	13.2	0.0	0.6	0.0
Robotics	9.1	2.1	2.1	1.0	2.8	1.7	3.4	1.5	0.3	0.3	100	0.1	1.8	0.1	1.1	1.9	0.0	1.0	0.0	0.1	0.2
Nanotech	5.6	0.9	2.0	0.2	3.3	0.2	1.2	6.1	0.3	0.5	9.3	100	3.6	1.1	0.7	3.2	1.9	0.0	0.2	0.0	0.0
IoT	13.7	4.9	13.3	1.7	8.1	1.9	24.9	8.4	0.4	7.6	2.9	0.1	100	0.0	1.1	0.3	0.0	6.2	0.0	0.8	0.7
CRISPR	2.0	0.2	0.7	2.3	2.5	0.1	0.5	0.1	0.0	0.2	3.0	0.3	0.0	100	0.2	0.1	0.0	0.2	0.0	0.0	0.0
VR	16.6	1.7	3.2	1.1	3.9	1.3	4.7	2.0	0.4	0.2	8.8	0.1	5.2	0.1	100	2.7	0.0	2.4	0.1	0.2	0.1
3Dprint	3.4	0.4	0.5	0.3	1.0	0.0	0.9	0.8	0.3	0.0	16.3	0.3	1.7	0.0	2.9	100	0.4	0.2	0.0	0.0	0.3
Polymer	0.5	0.2	0.0	0.7	0.4	0.0	0.1	0.2	0.0	0.0	0.7	1.3	0.0	0.0	0.2	2.9	100	0.0	0.0	0.0	0.1
Blockchain	21.4	5.5	11.2	2.3	8.0	5.1	23.1	2.6	0.1	5.4	6.8	0.0	24.5	0.0	2.0	0.2	0.0	100	0.0	0.8	0.3
Web2.0	1.4	4.0	8.1	2.0	0.8	0.8	21.1	6.5	0.4	0.0	0.3	0.1	0.2	0.0	0.3	0.1	0.0	0.1	100	2.0	0.0
SOA	3.9	9.1	9.4	0.7	1.8	0.8	26.3	2.1	1.1	0.1	0.3	0.0	1.8	0.0	0.1	0.0	0.0	0.4	0.3	100	0.0
RFID	0.6	1.0	4.1	0.4	3.5	0.1	2.3	6.0	0.7	0.0	2.1	0.0	3.9	0.0	0.1	0.3	0.0	0.3	0.0	0.0	100

Note**:** Each number represents the fraction of job postings that mention the technology in the row that also mention the technology in the column. The shade of blue darkens with higher numbers.

Table 4: Evidence of widespread use

Technology	(1) Gini all jobs 2019 (3 digit NAICS)	(2) Gini all jobs 2010 (3 digit NAICS)	(3) Total jobs 2019	(4) Total jobs 2010
Telecommunications	0.55	0.56	411,262	244,240
Robotics	0.57	0.72	105,108	18,136
Cloud computing	0.62	0.68	590,189	88,591
Service-oriented architecture (SOA)	0.63	0.66	27,608	20,794
ML	0.65	0.78	152,002	7,255
GIS	0.74	0.69	48,662	21,389
3D printing	0.75	0.91	12,532	196
Internet-of-things	0.76	0.87	63,072	1,590
Blockchain	0.77	n/a	15,829	0
RFID	0.79	0.83	11,754	2,352
Web 2.0	0.83	0.71	3,627	20,246
Virtual Reality	0.85	0.83	13,299	507
Polymer Science	0.90	0.87	1,677	1,469
Nanotechnology	0.90	0.90	1,081	1,143
CRISPR	0.95	n/a	3,670	0
Quantum Computing	0.95	0.93	6,436	95

Table 5: Number of research jobs

Technology	(1) Total research jobs in 2019	(2) Total research jobs in 2010	(3) % research in 2019 (out of total per tech)	(4) % research in 2010 (out of total per tech)
ML	19,772	989	13.0	13.6
Cloud computing	11,274	867	1.9	1.0
Robotics	6,405	1,963	6.1	10.8
Telecommunications	6,354	4,044	1.6	1.7
CRISPR	2,609	0	71.1	n/a
GIS	1,797	480	3.7	2.2
Internet-of-things	1,779	48	2.8	3.0
3D printing	1,569	20	12.5	10.2
Virtual Reality	1,075	47	8.1	9.3
Polymer Science	618	486	36.9	33.1
Blockchain	524	0	3.3	n/a
Nanotechnology	441	251	40.8	22.0
Quantum Computing	378	11	5.9	11.6
Service-oriented architecture (SOA)	348	165	1.3	0.8
RFID	146	69	1.2	2.9
Web 2.0	137	309	3.8	1.5

Table 6: Research job postings across applications sectors

Technology	(1) Gini research jobs 2019 (3 digit NAICS)	(2) Gini research jobs 2010 (3 digit NAICS)
ML	0.60	0.86
Cloud computing	0.63	0.87
Robotics	0.64	0.81
Internet-of-things	0.72	0.96
Telecommunications	0.74	0.67
3D printing	0.82	0.98
GIS	0.83	0.88
Polymer Science	0.88	0.91
Blockchain	0.89	n/a
Virtual Reality	0.90	0.96
CRISPR	0.90	n/a
Service-oriented architecture (SOA)	0.91	0.96
Nanotechnology	0.94	0.95
RFID	0.95	0.95
Quantum Computing	0.95	0.99
Web 2.0	0.98	0.93

Table 7: GPT likelihood relative rank in 2019

Technology	(1) Widespread use (Gini 3-digit NAICS)	(2) Many research jobs (count of research jobs)	(3) Disproportionate research jobs (fraction of research jobs)	(4) Widespread research use (Gini 3-digit NAICS)
ML	5	1	4	1
Cloud Computing	3	2	13	2
Robotics	2	3	7	3
Internet of Things	8	7	12	4
Telecommunications	1	4	14	5
3D Printing	7	8	5	6
Geographic Information Systems	6	6	10	7
Polymer Science	13	10	3	8
Blockchain	9	11	11	9
Virtual Reality	12	9	6	10
CRISPR	15	5	1	11
Service-oriented architecture	4	14	15	12
Nanotechnology	14	12	2	13
RFID	10	15	16	14
Quantum Computing	16	13	8	15
Web 2.0	11	16	9	16

Table 8a: Evidence of widespread use for data-related technologies

Technology	(1) Gini for all jobs 2019 (3-digit NAICS)	(2) Gini for all jobs 2010 (3-digit NAICS)	(3) Total jobs 2019	(4) Total jobs 2010
BI	0.42	0.48	338,615	221,120
Data mining	0.49	0.57	94,205	33,730
Data science	0.56	0.66	188,092	14,013
Big data	0.63	0.81	258,761	9,680
ML	0.65	0.78	152,002	7,255
NLP	0.67	0.78	39,386	4,563

Table 8b: Number of research jobs for data-related technologies

Technology	(1) Total research in 2019	(2) Total research in 2010	(3) % research in 2019 (out of total per tech)	(4) % research in 2010 (out of total per tech)
Data science	26,527	2,161	14.10	15.42
ML	19,772	989	13.01	13.63
Data mining	13,499	3,899	14.33	11.56
Big data	12,540	148	4.85	1.53
BI	10,921	3,302	3.23	1.49
NLP	4,250	182	10.79	3.99

Table 8c: Research job postings across applications sectors for data-related technologies

Technology	(1) Gini research jobs 2019 (3-digit NAICS)	(2) Gini research jobs 2010 (3-digit NAICS)
Data mining	0.51	0.84
Data science	0.55	0.86
BI	0.56	0.69
ML	0.60	0.86
Big data	0.64	0.94
NLP	0.73	0.94

Online Appendix 1: Classification of job postings

Table A1.1: Our definition of the different types of job postings and examples for each

Technology category	**First year in Gartner hype-cycle**	**Definition**	**Example research job**	**Example non-research job**	**Count research job postings (2010 – 2019)**	**Count non-research job postings (2010 – 2019)**	**Count total job postings (2010-2019)**
ML	2007	At least one skill in the BG defined skill cluster "Machine Learning"	ID: 38317996020 Title: Oncology Bioinformatics/Data Science Roles Employer: Astrazeneca Degree-level: PhD Skills: Python, Machine Learning, Artificial Intelligence, Clinical Research, Mathematical Modeling, Somatic, Data Analysis, Natural Language Processing, Next Generation Sequencing (NGS), Bioinformatics, Big Data, Data Management, UNIX, Time Series Models, Molecular Targets, Cancer knowledge, Biomarkers, Drug Discovery, Biotechnology, Deep Learning, Communication Skills, Genomics, Data Science, Oncology, Bayesian Modeling, Biology, Immunology	ID: 38413121409 Title: Senior Technical Product Manager - Mulesoft Employer: Salesforce Degree-level: Master's Skills: Data Warehouse Processing, Quick Learner, Data Science, Oral Communication, Data Warehousing, Analytical Skills, Product Management, Technical Writing / Editing, Mulesoft, Extraction Transformation and Loading (ETL), Machine Learning, Product Development, Writing, Network Troubleshooting, Software Engineering, Target Market, Communication Skills, Artificial Intelligence, Troubleshooting, Product Sales, Creative Problem Solving, Prioritizing Tasks, MuleSoft Anypoint, Customer Acquisition, Creativity	68,552	469,132	537,684
Business Intelligence (BI)	2012	At least one skill in BG defined skill cluster "Business Intelligence" or "Business Intelligence Software"	ID: 38472330759 Title: Data & Applied Scientist Employer: Parkland Health Degree-level: Master's Skills: Pentaho, Social Services, Data Science, Data Analysis, Predictive Models, SAS, Meeting Deadlines, Tableau, Model Building, Data Visualization, Machine Learning, Writing,	ID: 38472246468 Title: Systems Analyst Big Data/Hadoop Employer: (not available) Degree-level: Master's Skills: Microsoft Visio, Extraction Transformation and Loading (ETL), Data Warehousing, Business Intelligence, Systems Analysis, Apache Hadoop, Microsoft Office,	64,335	2,842,778	2,907,113

			Scikit-learn, Statistical Methods, Natural Language Processing, Experiments, SPSS, R, Pattern Recognition, Research, Critical Thinking, SQL, D3.js, SAP BusinessObjects, Qlikview, WEKA	Big Data Analytics, Software Installation, Data Management, Big Data, Information Technology Industry Knowledge			
Big Data	2011	At least one skill in BG defined skill cluster "Big Data"	ID: 38472636425 Title: Quantitative Research Analyst Employer: (not available) Degree-level: (not available) Skills: Deep Learning, Fixed Income, Communication Skills, Big Data, Investment Strategy, Research, Quantitative Research, Natural Language Processing, Machine Learning, Risk Management, Investment Management, Decision Making, Business Development	ID: 38472244950 Title: Enterprise Database Administrator/Developer Employer: General Mills Degree-level: Bachelor's Skills: Microsoft Active Directory, SAP, Oracle, Teradata DBA, Authentication, Problem Solving, Domain Name System (DNS), Database Administration, Clustering, SAP HANA, Apache Hadoop, VMware, Python, MongoDB, Linux, Ansible, MySQL, SQL	55,298	1,148,248	1,203,546
Data Mining	1996	At least one skill in BG defined skill cluster "Data Mining"	ID: 38316930989 Title: Research Scientist Employer: Point Blank Solutions, Inc Degree-level: PhD Skills: Experiments, Experimental Design, Writing, Physical Abilities, Computer Literacy, Technical Writing / Editing, Scheduling, Project Design, Typing, Engineering Design, Data Mining, Risk Assessment, Risk and Mitigation Analysis, Product Development, Chemical Engineering, Failure Analysis, Process Improvement, Research, Initiative, Planning, Root Cause Analysis, New Product Development, Engineering Design and Installation, Simulation, Financial Analysis, Project Management, Polymer Science, Numerical analysis, Technical Training	ID: 38472290861 Title: Business Intel Engineer II Employer: Amazon Degree-level: Bachelor's Skills: Business Intelligence, Data Engineering, Predictive Models, Data Mining, Teamwork / Collaboration, Data Science, Problem Solving, Presentation Skills, Decision Making, Physics, Retail Industry Knowledge, Machine Learning, Amazon Web Services (AWS), Amazon Web Services (AWS), Data Validation, Big Data Analytics, Python, Data Quality, Economics, Program Development, SQL, Creativity, Research, Big Data	82,406	573,979	656,385

Data Science	2004	At least one skill in BG defined skill cluster "Data Science"	ID: 38472583705 Title: Senior Data Scientist, Evaluations Employer: Quartet Health Degree-level: PhD Skills: Mental Health, Data Science, Predictive Models, Medical Coding, Problem Solving, Teamwork / Collaboration, Primary Care, Multi-Tasking, Biostatistics, Data Transformation, Customer Contact, Software Development, Applied Statistics, Extraction Transformation and Loading (ETL), Python, Software Engineering, Communication Skills, Economics, Git, Epidemiology, Statistical Programming, Information Technology Industry Knowledge, Bioinformatics, Data wrangling, Experiments, Statistical Methods, Research, Creativity	ID: 38472253223 Title: Hris Analyst Employer: Fluke Networks Degree-level: Bachelor's Skills: Cost per hire, Detail-Oriented, Human Resource Management Industry Knowledge, Data Manipulation, Oracle Business Intelligence Enterprise Edition (OBIEE), Oracle, Teamwork / Collaboration, Problem Solving, Organizational Skills, Analytical Skills, Data Analysis, Business Intelligence Reporting, Data Science, Time Management, Communication Skills, Microsoft Excel, Taleo, Oracle HCM Assessments, HR Metrics, Project Management, Microsoft Office, Root Cause Analysis, Sales, Research, Microsoft Sharepoint, Creativity, Human Resource Information System (HRIS), Critical Thinking	104,062	643,609	747,671
Natural Language Processing (NLP)	1995	At least one skill in BG defined skill cluster "Natural Language Processing (NLP)"	ID: 38321530432 Title: Principal Analyst, Quantitative Research - Advanced Analytics Employer: FINRA Degree-level: PhD Skills: Predictive Models, Decision Trees, Research, Machine Learning, Data Collection, Self-Starter, Economics, Natural Language Processing, Organizational Skills, Financial Industry Knowledge, Surveillance, Random Forests, Quantitative Research, Derivatives, Securities, Risk Management, Meeting Deadlines, Fixed Income, Writing, Pattern Recognition, Data Science, Statistical Methods, Equities,	ID: 38472293179 Title: Principal Technical Program Manager Tpm Alexa - Product Knowledge Employer: Amazon Degree-level: Bachelor's Skills: Program Management, Natural Language Processing, Amazon Web Services (AWS), Amazon Web Services (AWS), Planning, Web Application Development, Total productive maintenance, Product Knowledge, Multi-Tasking, Amazon Alexa, Quality Management, Product Management	15,737	148,173	163,910

			Stress Testing, Predictive Analytics				
Cloud Computing	2008	At least one skill in BG defined skill clusters “Cloud Computing”, “Cloud Solutions”, “Cloud Storage”	ID: 38413077222 Title: Cloud Architect/Research Technologist Employer: (not available) Degree-level: Master’s Skills: OpenStack, Cloud architecture, Teamwork / Collaboration, Chef Infrastructure Automation, AWS Simple Storage Service (S3), Microsoft Azure, Writing, ServiceNow, ServiceNow, Configuration Management, Troubleshooting, Linux, CEPH (Software), CloudStack, Linux Scripting, VMware, Virtualization, Communication Skills, Experiments, Xen, UNIX Shell, Google Compute Engine (GCE), Kubernetes, Research, UNIX, Puppet, Creativity, Hyper-V	ID: 38448369509 Title: Systems Administrator Employer: (not available) Degree-level: Bachelor’s Skills: Scalability Design, Network Switches, Cisco, MacIntosh OS, Cloud Computing, Cisco Switching, Ubuntu, Virtual Private Networking (VPN), Secure Shell, Good Clinical Practices (GCP), Caching, Linux, Network Administration, Ethernet, Kubernetes, Graphics Processing Units (GPU), System Administration	42,723	2,737,205	2,779,928
Telecommunications	1995	At least one skill in BG defined skill cluster “Telecommunications”	ID: 38321785505 Title: Decision Analyst Employer: Huntington National Bank Degree-level: PhD Skills: SQL, Risk Management, Microsoft Excel, Data Mining, SPSS, Digital Marketing, Direct Marketing, Direct Mail, Business Intelligence, R, Microstrategy, SAS, Retail Industry Knowledge, SQL Server, S-Plus, Economics, Statistics, Apache Hadoop, Python, Machine Learning, Problem Solving, Research, Oracle, Telecommunications, Java, Writing, Experimental Design, JavaScript, PERL Scripting Language, Communication Skills, SAP BusinessObjects, C++, Tableau,	ID: 38472246535 Title: Regional Field Support Engineer Employer: Wayfair Degree-level: Bachelor’s Skills: Troubleshooting, Windows Server, Linux, Group policy, Voice over IP (VoIP), Communication Skills, Hypertext Preprocessor (PHP), Physical Abilities, VBScript, Creativity, UNIX, CentOS, SQL, Microsoft PowerShell, Warehouse Operations, FreeBSD, Microsoft Active Directory, Microsoft Windows, Dynamic Host Configuration Protocol (DHCP), Switchgear, Detail-Oriented, Energetic, Problem Solving, Repair, Cisco, Network Switches, Domain Name System (DNS), E-Commerce, It Support, Team Building, Hardware and Software Configuration, Planning	53,241	3,364,855	3,418,096

			Microsoft Powerpoint, Verbal / Oral Communication, Microsoft Word, Statistical Analysis, Teradata				
GIS	2002	At least one skill in BG defined skill cluster "Geographic Information System (GIS) Software"	ID: 38414176209 Title: Research Specialist II Employer: County Riverside Degree-level: PhD Skills: QDA Miner, Survey Analysis, Staff Management, SPSS, Qualtrics, ArcGIS, Research, SQL, SQL Server, Statistics, Microsoft Excel, Project Management, Data Collection, Program Evaluation, Database Design, Public Health and Safety, Economics, Data Warehousing, Presentation Skills, Social Services, Data Analysis, Statistical Analysis, Natural Sciences, Case Management, Public administration, SAS, Planning, Microsoft Access, Report Writing, Writing, Research Design	ID: 38444284893 Title: Coast Finance Manager – Forestry Employer: (not available) Degree-level: Master's Skills: Information Systems, Organizational Skills, Positive Disposition, Geographic Information System (GIS), Customer Contact, Planning, Self-Motivation, Land Development, Forestry Operations, Verbal / Oral Communication, Communication Skills, Finance, Property Management, Accounting, Budgeting	11,756	315,985	327,741
Quantum Computing	1999	At least one skill in BG defined skill cluster "Quantum Computing"	ID: 38426245831 Title: Quantum Scientist, Lead Employer: Booz Allen Hamilton Inc. Degree-level: PhD Skills: Physics, Research Design, Machine Learning, Data Visualization, Customer Service, Data Science, Leadership, Research, Scheduling, Quantum Computing, Strategic Planning, Project Management	ID: 38444716328 Title: Associate Partner Security Strategy Risk and Compliance Employer: IBM Degree-level: Master's Skills: Technical Writing / Editing, Thought Leadership, Sales Leadership, Systems Integration, Professional Services Marketing, Quantum Computing, Internet of Things (IoT), Management Consulting	1,309	12,095	13,404
Robotics	2007	At least one skill in BG defined skill cluster "Robotics"	ID: 38416313106 Title: Human Systems Engineer - Elsys Employer: Georgia Institute of Technology Degree-level: PhD	ID: 38444276675 Title: PLC Programmer Employer: Diedre Moire Corporation Degree-level: (not available) Skills: Variable Frequency Drives (VFDs), Programmable Logic Controller (PLC) Programming,	35,705	514,053	549,758

			Skills: Computational Modeling, Experimental Design, Industrial Engineering, Software Development, Simulation, Robotics, Autonomous Systems, Computer Engineering, System Architecture, Decision Making, Surveys, Research, Human Computer Interaction, Systems Engineering, Industrial Engineering Industry Expertise, Psychology, Avionics	Human Machine Interface (HMI), Compliance with Customer Specifications, Electrical Systems, C++, Visual Basic, Software Development, Automation Systems, Technical Support, Servo Drives / Motors, Machinery, Rockwell Automation, Debugging, Microsoft C#			
Nanotechnology	2002	At least one skill in BG defined skill cluster ”Nanotechnology”	ID: 38446874232 Title: Associate Scientists I Employer: Black Diamond Structures, Llc Degree-level: Master’s Skills: Lifting Ability, Nanotechnology, Chemistry, Research, PH Meters, Microsoft Office, Java, Materials Science, Mechanical Engineering, X-Rays, Detail-Oriented, Laboratory Safety and Chemical Hygiene Plan, Data Analysis, Organizational Skills, Microscope, Laboratory Equipment, Technical Support, Tableau, Lab Safety	ID: 38452828314 Title: High Vacuum Technician Employer: Texstars Llc Degree-level: Bachelor’s Skills: Detail-Oriented, Manufacturing Processes, Quality Management, Plumbing, Repair, Robotics, Purchasing, Technical Support, Electronic Schematics, Programmable Logic Controller (PLC) Programming, Preventive Maintenance, Equipment Repair, Predictive / Preventative Maintenance, Quality Assurance and Control, Nanotechnology, Windows Programming	4,101	7,279	11,380
Internet-of-things (IoT)	2011	At least one skill in BG defined skill cluster “Internet of Things (IoT)”	ID: 38426874176 Title: Senior Staff Rf And Electrical Engineer Advanced Development Employer: Eargo Degree-level: PhD Skills: FDA Regulations, Firmware, Scheduling, Research, Experiments, Initiative, Quality Assurance and Control, Budgeting, Emissions Testing, Software Testing, Compliance Testing, Electrical Systems, Communication Skills, Engineering Design, Circuit Board, Embedded Firmware,	ID: 38444285332 Title: Product Marketing Manager Employer: (not available) Degree-level: Bachelor’s Skills: Pricing Strategy, Demand Forecasting, Product Research, Product Design, Research, Market Strategy, Key Performance Metrics, New Product Development, Internet of Things (IoT), Microsoft Excel, Competitive Analysis, Time Management, Written Communication, Product Management, Market Research, Software Development, Retail Industry Knowledge, Writing,	5,835	198,806	204,641

			Internet of Things (IoT), Schematic Design, Configuration Management, Verbal / Oral Communication, Schematic Diagrams, Power Supplies, Detail-Oriented, Teamwork / Collaboration, Design for Manufacture/Design for Assembly (DFM/DFA), Electrical Engineering, Organizational Skills, Simulation, Engineering Design and Installation, Electrical Design, Oscilloscopes, Electrical Control, Digital Signal Processing (DSP), Test Equipment	Product Development, Planning, Software as a Service (SaaS), Software as a Service (SaaS), Product Marketing			
CRISPR (DNA logic and/or editing)	2005	At least one skill in BG defined skill clusters "CRISPR" or "CRISPR-DM"	ID: 38444688851 Title: Scientist - Drug Discovery Biology & Pharmacology Employer: (not available) Degree-level: (not available) Skills: Biochemical and Cell-Based Assays, Pharmacology, Research, Experiments, CRISPR, Biotechnology, Drug Discovery, Repair, Biology, Assay Development, Remodeling	ID: 38414235752 Title: Flow Cytometry Technical Sales Specialist Employer: Nanocellect Biomedical Degree-level: Master's Skills: Cell Cloning, Sales Forecasting, Sales, Flow Cytometry, Sales Planning, Market Planning, CRISPR, Biotechnology, Client Base Retention, Product Sales, Description and Demonstration of Products, Technical Sales, Genomics, Product Knowledge, Problem Solving, Strategic Sales, Leadership, Editing, Customer Service, Biology, Customer Contact, Sales Strategy, Market Dynamics, Business Acumen, Lead Generation	7,549	3,339	10,888
Virtual Reality (VR)	1995	At least one skill in the BG defined skill cluster "Augmented Reality/Virtual Reality (AR/VR)" or skill "Augmented Reality (AR)"	ID: 38488240063 Title: Vice President, Strategy & Innovation Research Employer: Synchrony Financial Degree-level: Master's Skills: Project Management, Consumer Insights, Business Strategy, Strategic Planning, Psychology, Communication Skills, Regression Analysis, Economics, New Product	ID: 38486938318 Title: Technology Analyst Employer: Infosys Degree-level: (not available) Skills: Requirements elicitation, Virtual Reality (VR), Information Technology Industry Knowledge, Opportunity Identification, Software Development, Employee Training, Level design	3,112	43,738	46,850

			Development, Consumer Segmentation, Creativity, People Management, Budget Management, Research, Virtual Reality (VR), Creative Problem Solving, Quantitative Research, Budgeting, Building Effective Relationships, Market Research, Presentation Skills, Consumer Behavior, Teamwork / Collaboration, Marketing Communications, Research Design, Focus groups, Benchmarking, Planning, Consumer Research				
3D printing	2007	At least one skill in the BG defined skill cluster “3D Printing/Additive Manufacturing (AM)”	ID: 38496736012 Title: Research and Development Mechanical Engineer Employer: Sandia Corporation Degree-level: Master’s Skills: Mechanical Design, Critical Thinking, Research, Creativity, Prioritizing Tasks, Laboratory Testing, Finite Element Analysis, Kinematics, Packaging, Computational Fluid Dynamics, Materials Science, Mechanical Engineering, Aerodynamics, Radar Systems, Remote Sensing, Fluid Mechanics, Problem Solving, Teamwork / Collaboration, Novel Materials, Microfluidics, Autonomous Systems, Materials Selection, 3D Printing / Additive Manufacturing (AM), Simulation, Systems Integration, Product Development, Physics, Nondestructive Testing (NDT)	ID: 38485145033 Title: Value Stream Manager Employer: United Technologies Corporation Degree-level: Bachelor’s Skills: 3D Printing / Additive Manufacturing (AM), Problem Solving, Facebook, Supervisory Skills, Cost Control, Scheduling, Process Improvement	5,734	39,904	45,638
Polymer Science	2003	At least one BG defined skill “Polymer Science”	ID: 38491999087 Title: Industrial Researcher - Post-Doctoral Employer: Evonik Degree-level: PhD	ID: 38491786895 Title: Process Engineer Employer: Cps Degree-level: Bachelor’s Skills: Machinery, Mechanical Engineering, Polymer Science,	6,425	12,354	18,779

			Skills: UV-Vis, Polymer Synthesis, 3D Printing / Additive Manufacturing (AM), Microscope, Cytotoxicity, Materials Science, Tissue Engineering, Personal Protective Equipment (PPE), Extrusion, Viscometers, Biomaterials, Research, Creativity, Chemistry, Polymer Science, Microsoft Office, Clinical Development, Communication Skills	Process Engineering, Chemical Engineering			
Blockchain	2016	At least one BG defined skill "Blockchain" or "Bitcoin"	ID: 38493674160 Title: Blockchain Researcher Employer: Anchorage Degree-level: (not available) Skills: Structured Methods, Economics, Research Reports, Cryptography, Algebra, Blockchain, Creativity, Onboarding, Research, Teamwork / Collaboration, Anti Money Laundering (AML), Analytical Skills, Educational Materials, Detail-Oriented, Oracle, Writing, Calculus	ID: 38485086384 Title: Marketing Specialist Employer: Intlmaec Degree-level: Bachelor's Skills: Bilingual, Social Media, Editing, Marketing, Training Programs, Infographics, Marketing Automation, CPT Coding, English, Chinese, Deep Learning, Creativity, Blockchain, Big Data	1,382	36,869	38,251
Web 2.0	2006	At least one BG defined skill "Web 2.0"	ID: 37840019691 Title: Research and Development Cybersecurity Employer: Sandia Corporation Degree-level: Master's Skills: Creativity, Cyber Security Knowledge, Analytical Skills, Intrusion detection, Critical Thinking, Information Extraction, Research, Python, Authentication, Information Assurance, Cryptography, Web 2.0, Vulnerability analysis, Apache Webserver, Simulation, Network Engineering, Agile Development, Experiments, Network Security, Software	ID: 38515713994 Title: Engineer 4 Network Engineering Data Center Employer: Comcast Degree-level: (not available) Skills: Technical Support, JNCIE, Network Troubleshooting, Cisco, Engineering Design and Installation, PERL Scripting Language, Technical Training, Network Switches, Web 2.0, Building Effective Relationships, Network Infrastructure (Edge POE Devices), System/Network Configuration, Juniper Networks, Ansible, Kubernetes, Network Engineering, Network Testing, Virtualization, Traffic Engineering, Engineering Design, Communication Skills, Python, SDN, Next Generation	1,653	101,416	103,069

			Engineering, Technology Transfer, Written Communication, Writing, System Design, Planning, Web Servers, Routers, PERL Scripting Language, Security Vulnerability & Penetration Testing, Vulnerability assessment	Data Center, Troubleshooting, Routing Optimization			
Service-Oriented Architecture (SOA)	2004	At least one BG defined skill "Service-Oriented Architecture (SOA)"	ID: 38540054216 Title: Emerging Technologies Lead Systems/Software Engineer Employer: MITRE Corporation Degree-level: PhD Skills: Prototype Design Development, Written Communication, AJAX, Systems Engineering, Python, Software Engineering, C++, Service-Oriented Architecture (SOA), Java, Unified Modeling Language (UML), Object-Oriented Programming, Creativity, JavaScript, Rhapsody, DevOps, Extensible Markup Language (XML), Software Architecture, Experiments, Computer Engineering, SysML, Application Lifecycle Management, Software Development, Agile Development, Mentoring, Web Services Architecture, Microsoft C#, XML Schemas, Scrum, Business Development, Extensible Stylesheet Language XSL, Internet Technologies, Project Planning and Development Skills	ID: 38529201989 Title: Senior Software Developer .Net Developer Employer: Comtech Global Degree-level: Bachelor's Skills: .NET, Oracle, Detail-Oriented, Microsoft Project, Oracle SOA Suite, Computer Engineering, Microsoft Edge, Web 2.0, Writing, Software Development, Service-Oriented Architecture (SOA), Communication Skills	2,586	252,861	255,447
RFID	2003	At least one BG defined skill "Radio Frequency Identification (RFID)"	ID: 38475177245 Title: Biological Threat Analyst, Mid Employer: Booz Allen Hamilton Inc. Degree-level: PhD	ID: 38472278381 Title: Secure Mobile Systems Engineer, Senior Employer: Booz Allen Hamilton Inc. Degree-level: Master's	853	62,788	63,641

			Skills: Detail-Oriented, Virology, Customer Service, Data Science, Biodefense, Analytical Skills, Problem Solving, Threat Analysis, Intelligence Analysis, Epidemiology, Empower, Splunk, Radio Frequency Identification (RFID), Immunology, Microbiology, Infectious Disease, Telematics, Graphics Processing Units (GPU)	Skills: JavaScript, Microsoft PowerShell, Hardware Experience, System Administration, Puppet, System Design, Ansible, C++, Communication Skills, Virtualization, Java, Python, Written Communication, Cryptography, Systems Engineering, VMware, Radio Frequency Identification (RFID), Swift (Programming Language), Transmission Control Protocol / Internet Protocol (TCP / IP), Certification & Accreditation, Configuration Management, Objective C, Linux, Information Systems, Bash, Microsoft C#, Building Effective Relationships, Software Customizations, Ruby, Software Development, Chef Infrastructure Automation, Hardware and Software Configuration, Systems Management			

Note: Research jobs defined as such if at least one skill in BG defined skill clusters labeled as "…Research…" and referencing scholarly-type research (i.e. "Research Methodology", "Laboratory Research", "Medical Research" and "Clinical Research"

Online Appendix 2: Job postings by NAICS 2-digit industry sector (2010 and 2019)

Table A2.1. Number of jobs in data by industry and technology (2010)

NAICS 2	11	21	22	23	31-33	42	44-45	48-49	51	52	53	54	55	56	61	62	71	72	81	92
ML	12	120	12	0	748	27	389	49	1211	558	31	1518	5	165	291	135	15	32	11	93
3D printing	0	2	0	1	57	0	0	1	5	3	0	17	0	3	32	2	0	0	1	1
Blockchain	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Cloud	30	77	101	145	6279	569	1945	505	10442	3729	421	23968	130	4892	859	1780	123	1941	339	726
CRISPR	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
GIS	120	350	218	164	2330	182	160	285	1571	402	242	5317	27	623	858	243	49	161	316	1389
IoT	0	1	1	7	396	7	10	9	185	23	1	373	0	54	12	29	1	7	1	26
Nanotech	0	2	5	3	199	7	7	3	13	15	4	202	0	7	264	18	0	0	3	261
Polymer	0	25	2	6	709	12	18	2	39	5	1	293	3	42	64	20	0	2	6	6
Quantum Computing	0	0	0	0	13	0	0	1	11	0	0	34	0	0	24	0	0	1	3	2
RFID	13	7	3	14	572	29	61	44	73	26	15	528	3	106	27	89	8	20	6	34
Robotics	24	71	73	264	6240	565	329	119	400	111	100	2183	13	686	866	1474	99	44	99	381
SOA	5	28	65	30	1406	101	459	280	1326	1675	39	5938	28	1170	202	248	31	299	96	285
Telecom	87	395	878	1949	19370	788	4453	1662	45906	9761	2233	50085	429	18163	6065	9821	393	2654	2954	6855
VR	0	0	0	3	40	3	6	2	34	7	3	111	0	23	94	22	1	1	1	18
Web 2.0	1	16	26	18	1298	63	382	153	1899	811	120	4627	29	937	804	286	62	167	137	202
BI	73	536	794	510	16665	1499	5625	2130	15210	21366	1603	53752	421	12577	3879	6879	392	3727	994	1940
BigData	0	34	5	7	456	41	243	101	1333	659	65	1486	6	346	83	1173	38	126	23	127
Data Mining	29	508	107	46	3240	292	1468	322	3204	4428	170	7027	53	1453	871	1907	98	490	159	383
Data Science	25	152	71	17	1684	66	494	83	1148	2437	104	3261	24	417	520	447	42	118	65	167
NLP	0	45	4	3	211	1	103	34	707	269	21	823	1	162	131	548	21	34	14	70

Table A2.2. Number of jobs in data by industry and technology (2019)

NAICS 2	11	21	22	23	31-33	42	44-45	48-49	51	52	53	54	55	56	61	62	71	72	81	92
ML	183	421	297	188	13184	613	8508	1790	12356	16756	926	27392	223	14250	5141	3209	195	793	598	1915
3D printing	14	41	21	64	4703	90	211	74	230	67	30	1687	10	577	1342	304	38	166	48	168
Blockchain	13	27	6	25	412	11	159	67	723	2047	41	6004	11	1700	209	62	21	47	41	67
Cloud	761	817	906	1460	27744	2110	21171	4732	44397	39693	4034	123784	963	62274	7105	8780	1076	5522	1702	6836
CRISPR	0	0	0	1	572	10	0	3	6	9	0	1088	0	41	789	424	0	5	1	21
GIS	290	362	1031	800	1390	1216	230	666	1481	1291	735	9655	73	2891	2415	604	142	209	505	5492
IoT	111	137	238	169	8074	683	2874	451	11375	1674	256	16989	63	4018	692	678	65	320	108	382
Nanotech	0	2	0	12	172	1	1	0	4	10	2	231	2	23	381	33	4	0	2	28
Polymer	0	19	1	4	728	9	45	3	18	8	3	266	5	82	148	12	1	2	24	9
Quantum Computing	0	0	0	0	63	19	7	12	174	55	0	5576	0	36	313	1	0	0	0	26
RFID	3	16	16	101	993	28	6198	265	139	52	34	1052	5	334	80	286	37	88	35	407
Robotics	93	241	243	1387	25299	679	3431	1317	2648	4477	1614	13873	187	6082	4580	10314	258	804	627	1784
SOA	20	34	58	56	1104	44	975	242	1146	2118	204	5919	61	3275	311	360	61	232	162	348
Telecom	157	948	2233	7129	19142	1341	21695	4506	50677	17285	6201	56555	811	29306	10515	26306	989	3730	9812	16637
VR	6	24	22	146	2478	19	753	79	2146	141	53	1934	8	823	942	218	149	227	29	237
Web 2.0	0	0	4	2	93	0	111	5	247	159	3	602	5	613	223	59	14	22	6	58
BI	266	937	1282	1426	25533	2637	18810	4832	17018	45374	3682	58125	842	25854	8721	13605	1011	5276	3322	5171
BigData	175	380	490	208	13128	731	7936	1649	18403	28922	979	53968	265	28865	2985	2656	403	1408	714	2881
Data Mining	59	500	302	331	7906	592	4065	1487	4456	20002	821	13964	228	5445	3405	5979	269	785	411	1990
Data Science	287	551	604	281	14146	746	7475	2517	12969	25268	1428	35320	299	14487	7392	5946	421	3460	1051	3036
NLP	31	52	30	78	1610	82	1321	217	2920	5488	189	7181	103	3203	1121	3978	59	543	209	1095

Table A2.3. Number of research jobs in data by industry and technology (2010)

NAICS 2	11	21	22	23	31-33	42	44-45	48-49	51	52	53	54	55	56	61	62	71	72	81	92
ML	6	24	0	0	123	4	58	3	153	117	3	225	0	11	61	49	0	2	0	12
3D printing	0	0	0	0	7	0	0	0	0	1	0	2	0	0	10	0	0	0	0	0
Blockchain	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Cloud	0	1	3	3	89	10	4	2	177	29	1	254	1	20	20	38	0	11	1	2
CRISPR	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
GIS	16	2	3	3	50	1	5	7	14	8	8	86	0	3	72	30	3	0	8	66
IoT	0	0	0	0	7	0	0	0	6	0	0	8	0	0	0	1	0	1	1	14
Nanotech	0	2	0	0	46	2	0	1	3	12	0	72	0	1	66	5	0	0	1	4
Polymer	0	5	0	0	251	4	6	1	22	1	0	91	2	15	21	7	0	0	3	1
Quantum Computing	0	0	0	0	0	0	0	0	2	0	0	8	0	0	1	0	0	0	0	0
RFID	0	0	0	0	13	0	1	0	2	1	0	25	0	0	7	4	0	3	0	0
Robotics	4	3	5	3	627	4	6	2	24	4	0	510	1	46	137	221	0	1	2	31
SOA	0	0	0	0	29	0	2	1	9	3	0	69	0	3	0	14	0	0	0	0
Telecom	0	5	3	5	850	19	33	22	404	95	13	978	3	120	118	712	0	16	20	71
VR	0	0	0	0	4	0	0	0	1	0	0	14	0	6	6	2	0	0	0	3
Web 2.0	0	0	0	0	37	3	6	0	42	4	9	81	0	11	36	7	0	0	3	1
BI	6	6	4	4	501	45	107	11	250	300	12	826	1	73	145	400	6	27	30	39
BigData	0	0	0	0	19	0	6	1	43	5	0	27	0	9	3	5	0	3	0	1
Data Mining	5	31	2	1	724	40	108	17	288	341	10	858	1	88	154	704	4	32	16	41
Data Science	6	10	1	1	427	13	46	9	135	308	7	533	3	51	147	140	0	5	7	46
NLP	0	2	0	0	18	0	3	0	30	9	0	46	0	7	14	31	0	0	0	1

Table A2.4. Number of research jobs in data by industry and technology (2019)

NAICS 2	11	21	22	23	31-33	42	44-45	48-49	51	52	53	54	55	56	61	62	71	72	81	92
ML	22	41	40	21	2045	57	930	287	1688	2682	127	3027	36	1473	1252	1006	11	123	66	291
3D printing	0	10	3	2	468	18	26	3	16	5	5	312	1	50	240	38	9	1	6	30
Blockchain	0	0	0	4	30	0	6	2	33	100	0	115	0	47	21	7	1	5	10	3
Cloud	21	11	11	38	790	47	412	75	1135	1215	53	2343	12	813	373	398	28	86	43	121
CRISPR	0	0	0	1	452	10	0	1	3	7	0	753	0	30	553	337	0	1	0	11
GIS	13	7	8	12	68	7	5	14	18	45	17	239	1	58	361	93	17	3	36	301
IoT	6	8	9	2	291	16	63	15	133	61	7	542	0	117	37	31	2	22	21	22
Nanotech	0	0	0	0	78	1	1	0	0	5	0	123	1	7	126	21	0	0	0	14
Polymer	0	5	1	3	249	4	17	1	13	4	0	138	1	25	45	8	1	2	0	5
Quantum Computing	0	0	0	0	15	0	0	0	14	6	0	280	0	6	30	0	0	0	0	5
RFID	0	0	1	0	25	0	3	0	0	0	0	24	0	21	5	10	0	0	0	15
Robotics	6	21	10	7	1716	22	120	31	148	99	6	1325	10	317	523	568	7	12	18	195
SOA	0	0	1	0	15	2	22	10	21	43	0	74	0	24	4	10	0	3	0	2
Telecom	0	0	18	8	1053	14	100	59	475	229	12	1058	4	344	433	720	3	56	35	337
VR	0	2	0	1	339	0	21	6	153	5	2	180	0	49	128	25	3	2	12	19
Web 2.0	0	0	0	0	0	0	0	0	0	0	0	1	0	94	11	6	0	1	0	0
BI	23	12	42	15	901	32	608	182	641	1449	74	1885	12	536	609	857	50	144	187	240
BigData	6	12	26	14	992	34	631	139	1277	2455	64	2121	11	1007	520	346	12	56	44	125
Data Mining	8	22	35	12	1994	41	360	91	633	1101	55	2372	16	578	1032	2569	19	80	77	214
Data Science	34	38	55	35	2555	59	1100	506	2072	3669	180	4117	39	1876	1545	1613	51	208	249	432
NLP	3	8	6	7	234	4	155	38	435	962	27	631	4	325	176	189	5	17	16	40

Online Appendix 3: Comparison with patent-based approaches, robustness

Table A3.1: Descriptive statistics for Table 2

	(1) **Mean**	(2) **St. Dev.**	(3) **Min**	(4) **Max**
Widespread use lagged 5 years (Gini by 3-digit industry)	0.75	0.15	0.40	1
Many research jobs lagged 5 years (count of research jobs in hundreds)	14.64	19.22	0	67.72
Disproportionate research jobs lagged 5 years (fraction of research jobs)	0.08	0.10	0	0.65
Widespread research use lagged 5 years (Gini by 3-digit industry)	0.88	0.10	0.63	1
Count of patents (thousands)	5.35	8.50	0	44.18
Count of patents (thousands) lagged 5 years	2.60	5.25	0	31.98
Co-occurring patent classes	266.64	144.59	0	534
Co-occurring patent classes lagged 5 years	179.31	127.29	0	463

Table A3.2: Job posting measures predict patent-based measures four years later

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
2010-2019	Count of patents	Co-occurring patent classes	Count of patents	Co-occurring patent classes	Count of patents	Co-occurring patent classes	Count of patents	Co-occurring patent classes
Widespread use lagged 4 years (Gini by 3-digit industry)	-1.457*** (0.427)	-0.069 (0.081)						
Many research jobs lagged 4 years (count of research jobs in hundreds)			0.009*** (0.003)	0.000 (0.001)				
Disproportionate research jobs lagged 4 years (fraction of research jobs)					-1.778*** (1.246)	-0.052 (0.241)		
Widespread research use lagged 4 years (Gini by 3-digit industry)							-2.134** (0.850)	-0.314*** (0.059)
Dependent variable lagged 4 years	0.087*** (0.008)	0.004*** (0.000)	0.085*** (0.007)	0.004*** (0.000)	0.090*** (0.011)	0.004*** (0.000)	0.088*** (0.008)	0.004*** (0.000)
LL	-326.48	-2,346.80	-326.88	-2,348.01	-330.04	-2,347.63	-321.81	-2,331.06
Observations	120	120	120	120	120	120	120	120

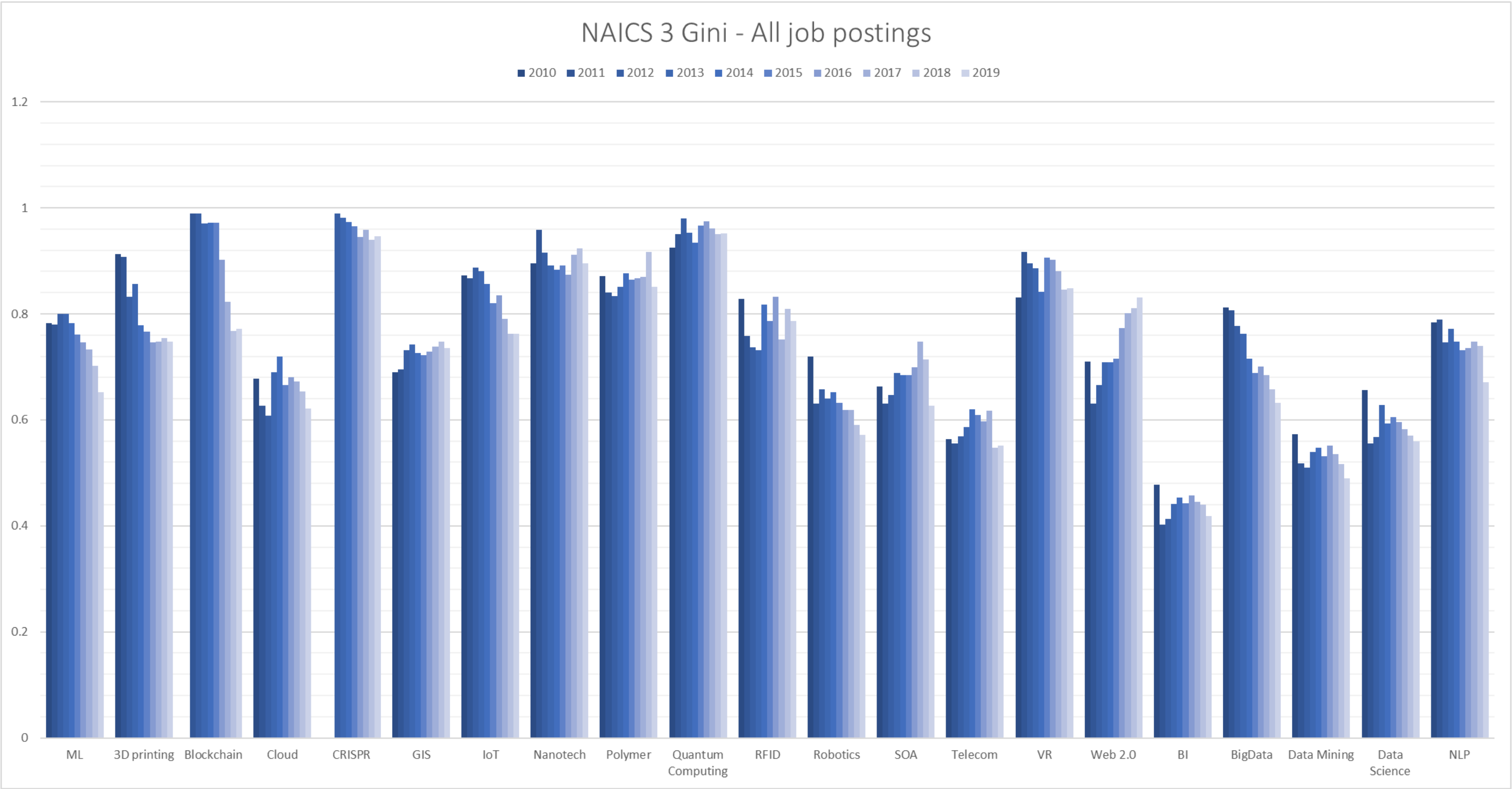
Note: Unit of observation is the technology-year. Dependent variables are based on patent data 2010-2019. Independent variables are based on job posting data 2010-2019. All columns show Poisson regressions with year fixed-effects and robust standard errors clustered at the year level. In columns 1, 3, 5 and 7 the dependent variable is count of patents in thousands. In columns 2, 4, 6 and 8 the dependent variable is count of co-occurring patent classes. *significant at 10%, **significant at 5%, ***significant at 1%

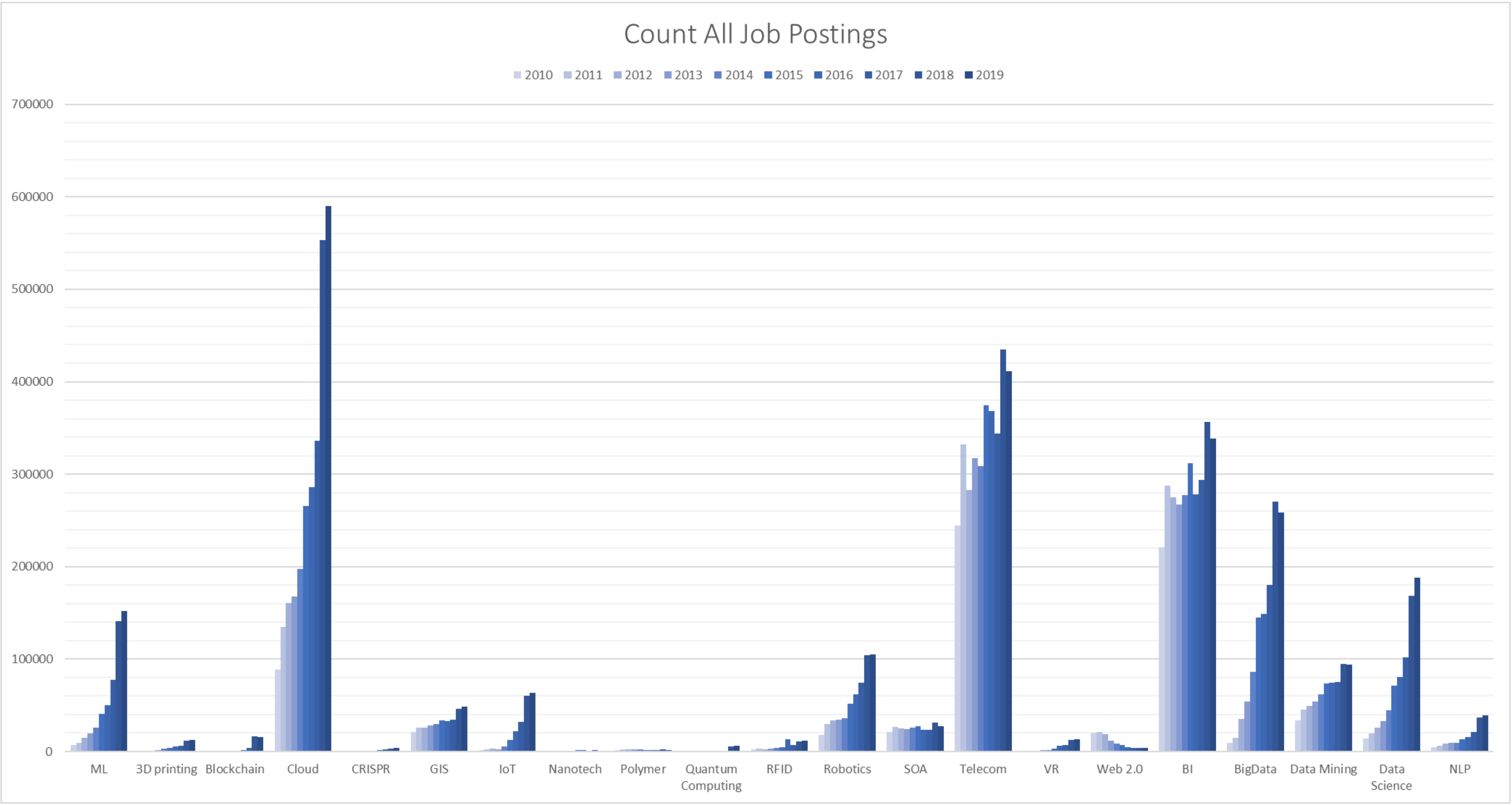
Table A3.3: Job posting measures predict patent-based measures six years later

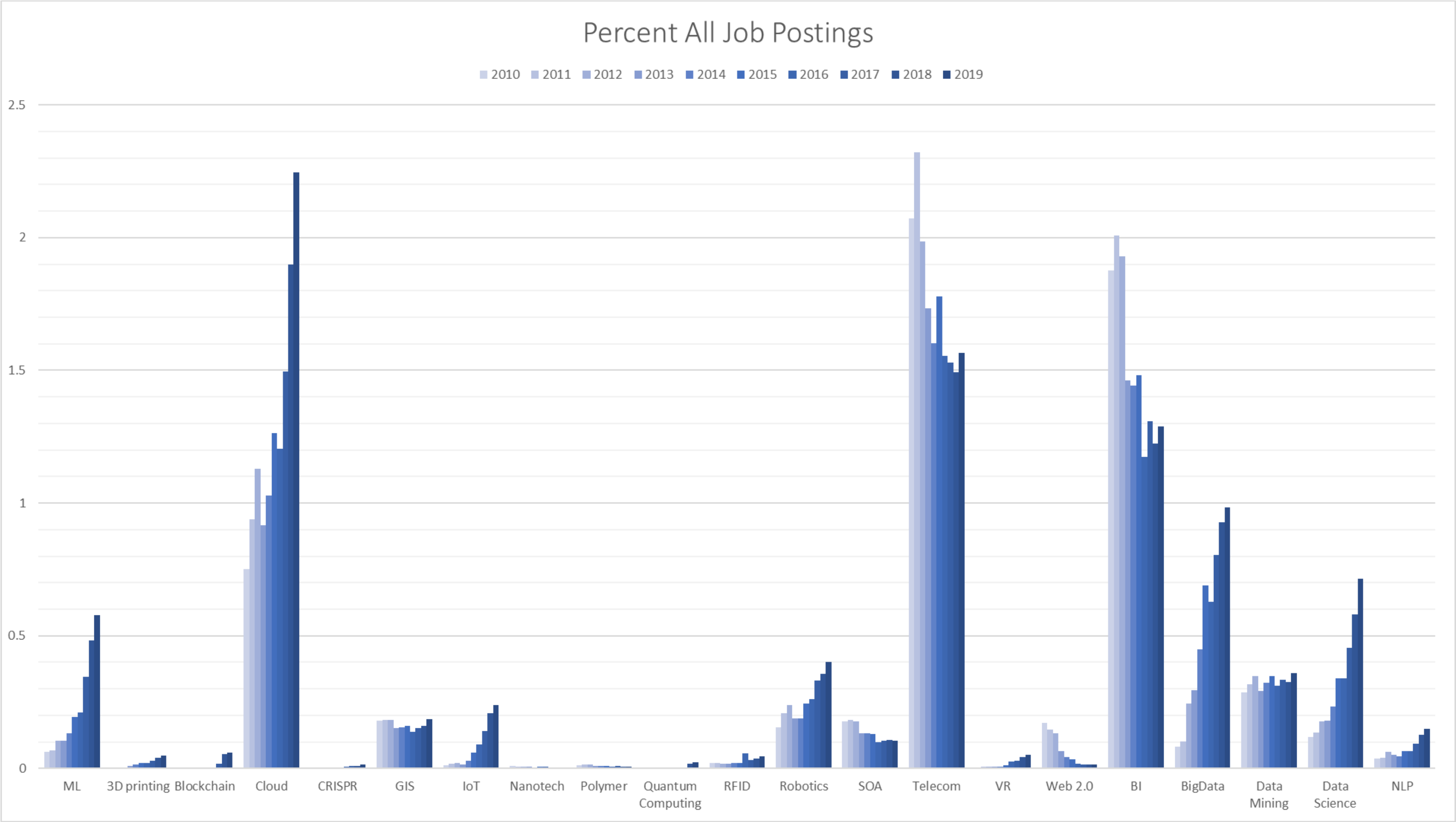
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
2010-2019	Count of patents	Co-occurring patent classes	Count of patents	Co-occurring patent classes	Count of patents	Co-occurring patent classes	Count of patents	Co-occurring patent classes
Widespread use lagged 6 years (Gini by 3-digit industry)	-1.113*** (0.348)	0.121 (0.068)						
Many research jobs lagged 6 years (count of research jobs in hundreds)			0.005*** (0.002)	-0.001* (0.001)				
Disproportionate research jobs lagged 6 years (fraction of research jobs)					-2.025*** (0.463)	0.090 (0.540)		
Widespread research use lagged 6 years (Gini by 3-digit industry)							-2.129*** (0.687)	-0.100 (0.076)
Dependent variable lagged 6 years	0.100*** (0.013)	0.003*** (0.000)	0.100*** (0.015)	0.003*** (0.000)	0.104*** (0.017)	0.003*** (0.000)	0.098*** (0.007)	0.003*** (0.000)
LL	-242.81	-1,982.17	-245.66	-1,979.57	-243.34	-1,984.50	-238.02	-1,984.06
Observations	80	80	80	80	80	80	80	80

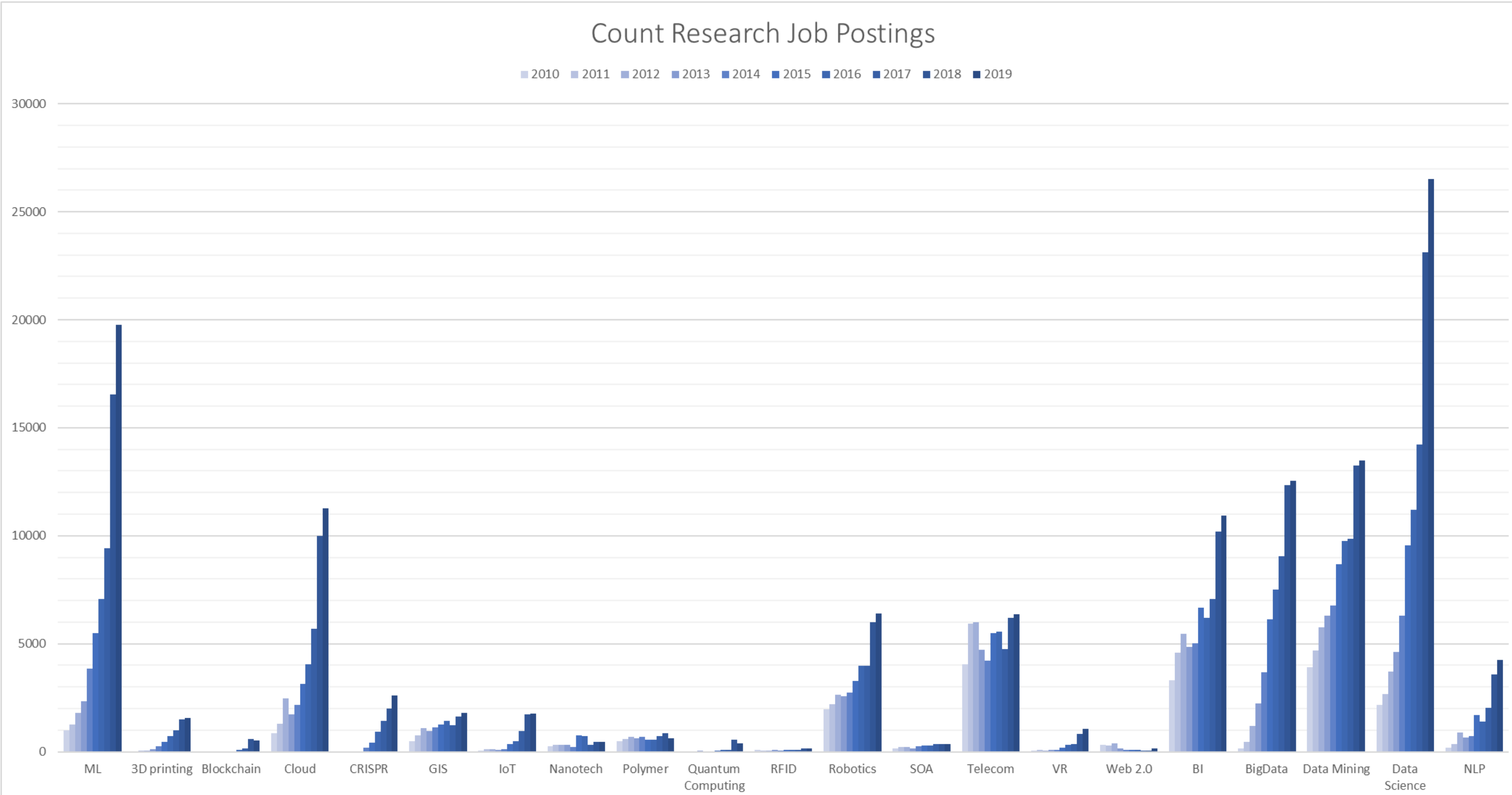
Note: Unit of observation is the technology-year. Dependent variables are based on patent data 2010-2019. Independent variables are based on job posting data 2010-2019. All columns show Poisson regressions with year fixed-effects and robust standard errors clustered at the year level. In columns 1, 3, 5 and 7 the dependent variable is count of patents in thousands. In columns 2, 4, 6 and 8 the dependent variable is count of co-occurring patent classes. *significant at 10%, **significant at 5%, ***significant at 1%

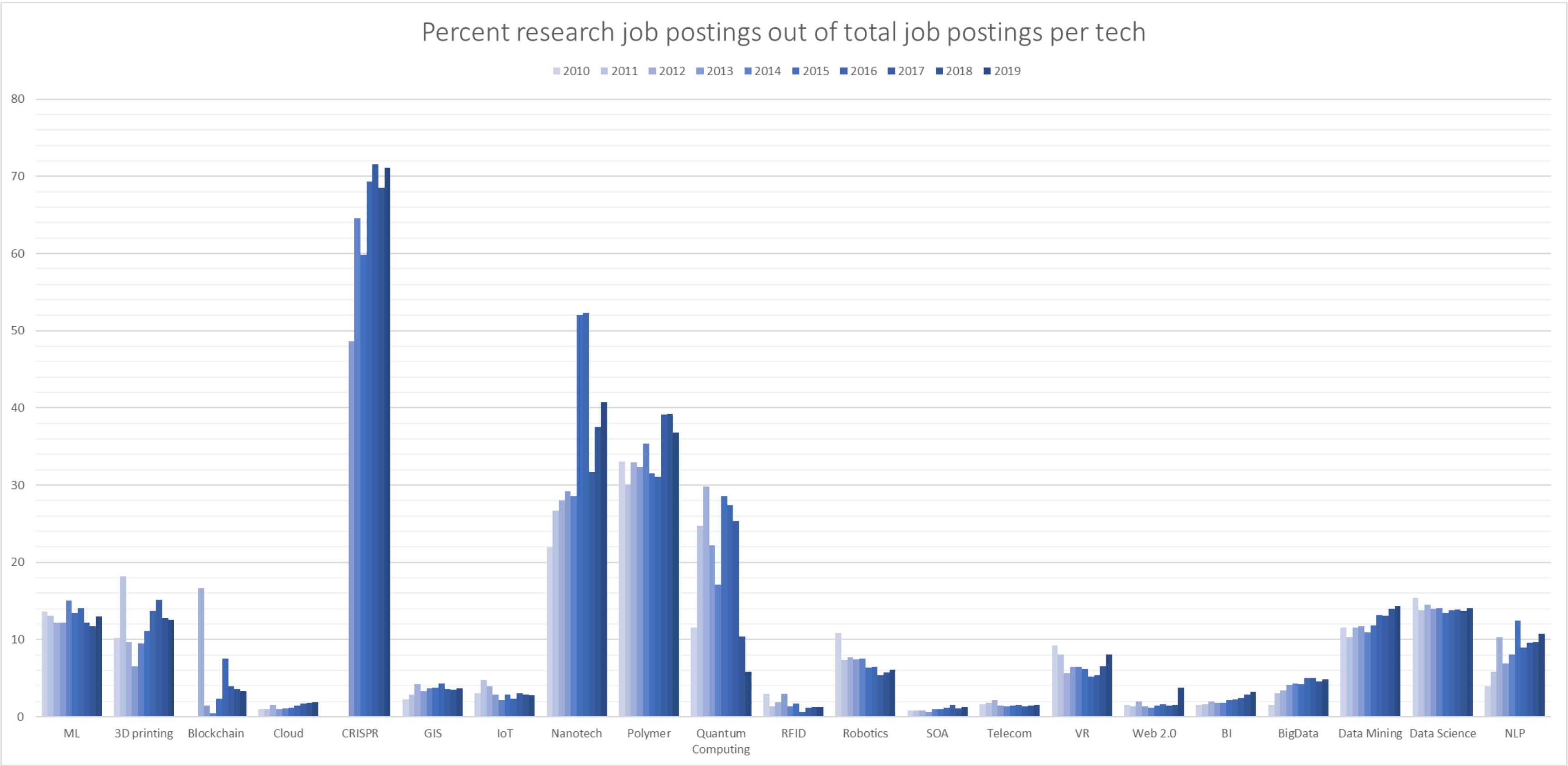
Online Appendix 4: Time-series data for the three GPT criteria (2010-2019)

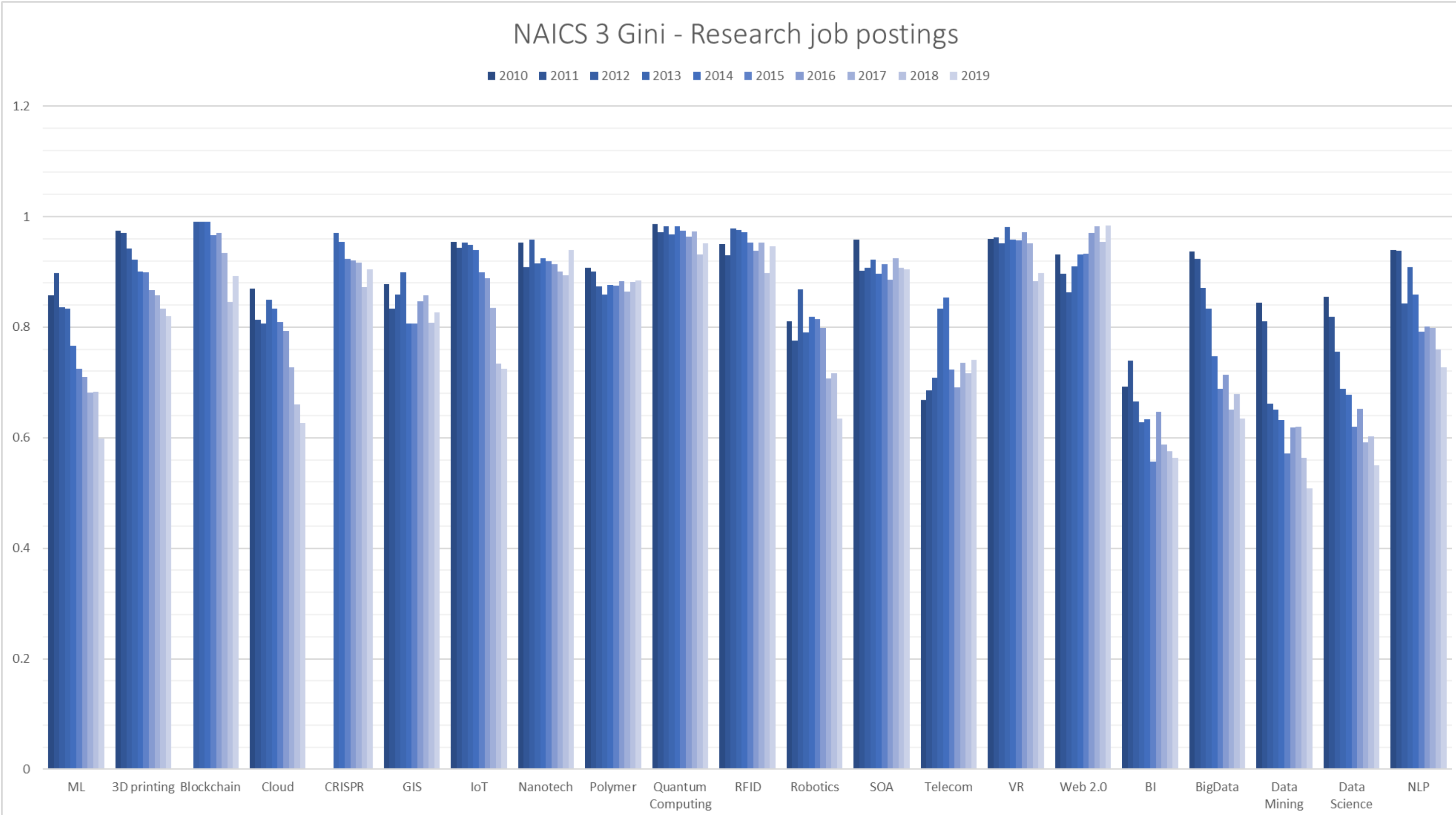



Count All Job Postings
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
700000
600000
500000
400000
300000
200000
100000
0
ML
3D printing
Blockchain
Cloud
CRISPR
GIS
IoT
Nanotech
Polymer
Quantum Computing
RFID
Robotics
SOA
Telecom
VR
Web 2.0
BI
BigData
Data Mining
Data Science
NLP


Percent All Job Postings
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2.5
2
1.5
1
0.5
0
ML
3D printing
Blockchain
Cloud
CRISPR
GIS
IoT
Nanotech
Polymer
Quantum Computing
RFID
Robotics
SOA
Telecom
VR
Web 2.0
BI
BigData
Data Mining
Data Science
NLP


Count Research Job Postings
2010 2011 2012 2013 2014 2015 2016 2017 2018 2019
30000
25000
20000
15000
10000
5000
0
ML
3D printing
Blockchain
Cloud
CRISPR
GIS
IoT
Nanotech
Polymer
Quantum Computing
RFID
Robotics
SOA
Telecom
VR
Web 2.0
BI
BigData
Data Mining
Data Science
NLP


Percent research job postings out of total job postings per tech
2010 2011 2012 2013 2014 2015 2016 2017 2018 2019
80
70
60
50
40
30
20
10
0
ML
3D printing
Blockchain
Cloud
CRISPR
GIS
IoT
Nanotech
Polymer
Quantum Computing
RFID
Robotics
SOA
Telecom
VR
Web 2.0
BI
BigData
Data Mining
Data Science
NLP


NAICS 3 Gini - Research job postings
2010 2011 2012 2013 2014 2015 2016 2017 2018 2019
1.2
1
0.8
0.6
0.4
0.2
0
ML
3D printing
Blockchain
Cloud
CRISPR
GIS
IoT
Nanotech
Polymer
Quantum Computing
RFID
Robotics
SOA
Telecom
VR
Web 2.0
BI
BigData
Data Mining
Data Science
NLP