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
University of Toronto	Burning Glass Technologies	University of Southern California
avi.goldfarb@utoronto.ca	btaska@burning-glass.com	teodorid@marshall.usc.edu

Abstract

There has been a great deal of speculation that machine learning might be a general purpose technology. Many other emerging technologies receive similar speculation. However, general purpose technologies are typically identified with the benefit of many years of hindsight. For managers deciding on technology strategy, this classification will come too late. In this paper, we provide an approach to assessing the likelihood that a technology is general purpose before it has widely diffused, so that the classification can be used to inform technology strategy decisions. Using data from online job postings, we compare several emerging technologies in terms of breadth of industries and the importance and breadth of research roles. Our results show that machine learning and related data science technologies are relatively likely to be general purpose.

Managers cannot develop a sound technology strategy to gain a competitive advantage without an early assessment of whether and how a technology is likely to be transformative. There has been a great deal of attention to the potential of recent advances in artificial intelligence (AI) in academia, industry, and the popular press. This excitement is primarily driven by a particular subfield of AI called machine learning (ML), a prediction technology.[1] Given that prediction tasks are pervasive in business and society, it has been widely speculated that ML is a specific type of transformative technology called a general purpose technology (GPT) (e.g. Brynjolfsson, Rock, and Syverson 2019; Cockburn, Henderson, and Stern 2019; Trajtenberg 2019).

ML is not the only emerging technology that has received considerable GPT hype. A variety of recent innovations are also claimed to represent GPTs, including 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). A better assessment of which of these are likely to be a GPT is important because it affects managerial strategy with respect to innovation, organizational change, and investment. Gaining a competitive advantage from GPTs requires investment in internal innovation capabilities and collaboration with academia and other organizations at the technology frontier, as well as an ability to sustain financial investments with long-return horizons. However, resolving this speculation using existing empirical methods for identifying GPTs requires many years of hindsight. By that point, the managerial relevance of knowing whether ML, blockchain, cloud computing, and other technologies are GPTs is limited. In this paper, we provide a methodology to assess GPT likelihood while the GPT is diffusing.

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

[1] For this reason, we will use "ML" to represent the technology underlying the recent excitement in artificial intelligence, as in Agrawal, Gans, and Goldfarb (2018).

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.[2] In other words, accruing GPT benefits requires 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). As a consequence, the benefits of a GPT are large but occur with a long lag (Bresnahan, Brynjolfsson, and Hitt 2002). They also typically require 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. Therefore, it is useful to have an early sense of whether an emerging technology is likely a GPT in order for managers to make informed decisions about their organization's technology strategy.

First, recognizing early that a technology is GPT would allow organizations to engage in managing the coordinated innovation process by, for example, investing in developing internal R&D capabilities, engaging in collaboration with academic researchers, forming alliances with cutting edge organizations in the producing industry and with other organizations in applications industries—including competitors (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.

[2] 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.

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. In contrast, if the technology is not a GPT, then the investment horizon is shorter, and the need for organizational change is narrower.

These strategic considerations are only useful if managers can identify which emerging innovations are likely to be GPTs. Bresnahan (2010, p. 764) emphasizes that "a GPT (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 patents and 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) attempt to identify all 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 (2011) examination of the purposeful recombination of genetic material (rDNA), use decades of data on patent citations. Similarly, Petralia (2020) examines decades of patent data to identify GPTs. By this point, it is too late to be managerially useful.

Our objective is to identify emerging technologies that are relatively likely to be general purpose. We leverage Tambe and Hitt's (2012a,b) insight that technology diffusion can be measured using labor demand data. In particular, data from online job postings provides an early measure of the diffusion of technology because firms are specific about technology needs in their job postings. 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.

As a GPT evolves, labor demand adjusts to include the type of skills required to develop and adopt the technology, up to a point when the technology is sufficiently pervasive that such skills are no longer explicitly listed in job posting. For example, in the early days of the internet job postings required internet-related skills; whereas, today, knowledge of using email and other internet-related skills are implicit. Thus, it is possible to glean information on the likelihood of a technology being general purpose by comparing patterns of skills in job postings at a point in time when the technology is relatively newly hyped.

We use labor demand data in online job postings to assess whether (1) ML is used in many industries, (2) it is used in many research jobs, and (3) the research jobs are spread across many industry sectors, when compared to other emerging technologies that have received attention in recent years. Our data come from Burning Glass Technologies and include job 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 focus on the 2010 and 2019 data, and examine job postings about 22 emerging technologies. We identify all technologies from the

Gartner “hype cycle” list between its inception in 1995 and 2019 that are also listed as skills in the Burning Glass job posting data. For example, technologies such as tablets or mobile phones are too broad to be sharply matched to job posting skills.

We proceed in two stages. First, we identify closely related technologies by examining the co-occurrence of the technologies in job postings. We find that six data-focused technologies are closely related: ML, business intelligence, big data, data mining, data science, and natural language processing. Our analysis therefore makes clear that discussions of ML as a GPT need to consider a cluster of technologies, similar to Rosenberg’s (1963) emphasis on a cluster of innovations underlying technological change in 19th century machine tools. Of these, we focus on ML because it is the technology that receives the most attention in discussions about general purpose technologies. We present results for the other five in a separate section, and show that this group of data-focused technologies are together likely to be general purpose technologies when compared to the other technologies examined.

Second, we evaluate each technology with respect to the three important aspects of being a GPT. We evaluate widespread use with the Gini coefficient on jobs by 3-digit NAICS. We evaluate use in many research jobs using two measures: the number of research jobs and the fraction of jobs using the technology that are research-focused. We evaluate research jobs across sectors using the Gini coefficient for research jobs by 3-digit NAICS. We interpret research jobs as jobs that help the firm innovate. If a technology is used in research in many industries then it suggests evidence of innovation in applications sectors.

Table 1 presents the final list of technologies and their relative rank in each of these four dimensions in 2019.

Table 1: Relative rank by technology in 2019

Technology	(1) Widespread use (Gini by 3-digit industry)	(2) Many research jobs (count of research jobs)	(3) Disproportionate research jobs (fraction of jobs that are research)	(4) Widespread research use (Gini by 3-digit industry)
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

While there is no formula for weighing these four measures, we interpret our findings to suggest that ML is relatively likely to be a GPT. It is consistently near the top, and it has the most research jobs and those jobs are relatively widespread. Since innovation in using industries is the key distinguishing feature of GPTs, we view the widespread use in research as particularly important. Cloud computing, robotics, and telecommunications are also relatively prevalent in research jobs and widespread research use. Although their fraction of research jobs is relatively low, we cannot exclude the possibility that these technologies are potential GPTs. Our results suggest that most of the other technologies listed are unlikely to be on the path to becoming GPTs in their current form.[3] Of course, several of these technologies—such as CRISPR and service-oriented-

[3] A limitation of our method 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.

architecture—should not be expected to be GPTs. We view it as comforting that our framework provides results consistent with such strong priors.

We also examine data from 2010. In that year, only telecommunications was relatively widespread, with a large number of research jobs in a wide range of industries. In other words, looking at the 2010 data, telecommunications looked like a possible GPT (though less so than ML did in 2019) and none of the other technologies looked like potential GPTs. The patterns from 2011 through 2018 reveal little new information about the degree to which the technologies we examine display characteristics of GPTs; they change roughly monotonically over time. For this reason, we only show 2010 and 2019 in the paper, and include data on the other years in the appendix.

We benchmark our results against a patent-based quantitative method for evaluating GPT likelihood develop by Petralia's (2020). Other patent-based methods for evaluating GPTs rely on citation data over a long period of time. The Petralia (2020) technique focuses on the quantity and growth of patents to measure widespread use, and the quantity of different patent classes in which the technology is used to measure use in application industries. We find that our 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 values of the patent-based ones and the strength of the prediction is highest for technologies that are most emerging.

Overall, we conclude that the cluster of technologies with ML is relatively likely to be a GPT. This cluster of technologies is clearly widely used, capable of ongoing technical improvement, and widely used for innovation in applications sectors. Companies that adopt ML need to anticipate the innovation externalities that are needed by collaborating with other organizations and understanding that any substantial benefits depend on further innovations over a long time horizon. Some of these benefits may therefore accrue to others, whether intermediaries who specialize in key complements or consumers.

1. Data

1.1 Description of the data

Our 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 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 devise a methodology to assess the GPT likelihood of emergent hyped technologies. Gartner has documented the set of emergent hyped 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 to a total of 22 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. These technologies can be viewed as a combination of other technologies that we can identify in the Burning Glass skills.[4] Nevertheless, the 22 technologies we examine should be seen as a non-exhaustive list of technologies that have advanced substantially this century.

We do not restrict our list to technologies that have been hypothesized to be GPTs. We include all hyped technologies that we can map to the Burning Glass job posting data knowing that some of the technologies are clearly not GPTs. We support our approach by examining technologies

[4] A small number of technologies are mentioned in the early Gartner data, and do exist as Burning Glass skills, but in a different form. For example, Java appears in a few of the 90s' Gartner hype cycles. In the 2010-2019 Burning Glass data, only later Java technologies are included as skills in job postings. In our view, these skills would not capture the reason Java appeared in the early Gartner hype cycles.

where there are strong priors about the low GPT likelihood, and demonstrating that our approach suggests that these technologies are indeed unlikely to be GPTs.

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.[5] For example, we define ML if the posting has at least one skill which is in the skill cluster of "machine learning." Online 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 jobs as distinct from non-research jobs. In our view, research jobs are disproportionately likely to drive innovation. We define a job 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 jobs in the data. While a handful of these research jobs relate to background research and have little to do with corporate "research & development" that generates innovation, on balance we believe Burning Glass' research clusters bias toward false negatives. In other words, there are likely few jobs misclassified as research jobs. The skill clustering approach ensures that jobs 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 conservative approach means several of the jobs classified as non-research might be reasonably seen as research jobs. While this means we probably undercount research jobs, we see no reason why this affects the comparison between

[5] Some job postings list skills from multiple technologies. We classify these job postings as requiring both skills.

technologies. In other words, we expect any potential undercounting of research jobs to apply to all technologies in similar ways.

1.2 Identifying closely related technologies

There is no requirement in the Gartner process for the technologies to be mutually exclusive. Our analysis, 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 these clear lines, 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 tend to appear in the same job postings, we argue that they are likely to represent the same underlying tools.[6] Table 2a presents the overlap in job postings in 2010; table 2b presents the same overlap in 2019. Each number represents the fraction of jobs that mention the technology in the row that also mention the technology in the column. For example, Table 2a 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.

There is strong overlap between ML, business intelligence, big data, data mining, data science, and natural language processing. 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 jobs 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 jobs mention ML, 54% of NLP jobs mention ML, and 27% of data mining jobs mention data science. Clearly, these technologies are closely related.

[6] 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.

Of these six technologies, we focus on ML because it is the technology most hypothesized to be a GPT. Later, we show the results for the other five technologies and demonstrate that these technologies display similar patterns and together are relatively likely to represent a GPT. Put differently, and consistent with prior work on other major technological changes (e.g. Rosenberg 1963), our results suggest that it is not appropriate to consider ML as a GPT, independent of a cluster of related technologies.

The only other technologies that are tightly and symmetrically connected are cloud computing and virtual machines. We focus on cloud and drop virtual machines because cloud is much more prevalent. While other technologies are sometimes connected, the connections are not symmetric as with 13% of RFID jobs mentioning telecommunications in 2010, but only 0.1% of telecommunications jobs mention RFID.

Table 2a: Overlap of technology in job postings in 2010

	ML	BI	Big Data	Data Mining	Data Science	NLP	Cloud	Telecom	VM	GIS	Quantum	Robotic	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	2.6	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	1.3	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	9.2	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	0.9	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	0.8	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	1.4	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	31.1	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	2.9	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
VM	0.2	3.7	1.1	0.4	0.1	0.1	34.6	9.0	100	0.3	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.8	0.9	0.1
GIS	0.3	5.9	0.1	1.7	0.9	0.1	1.6	10.3	1.2	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	1.1	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	0.5	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	0.4	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	1.2	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	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	2.8	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	0.0	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	0.0	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	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	3.2	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	3.4	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	1.9	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

Table 2b: Overlap of technology in job postings in 2019

	ML	BI	Big Data	Data Mining	Data Science	NLP	Cloud	Telecom	VM	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	4.2	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	2.2	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	10.7	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	1.6	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 Science	36.2	16.1	29.3	13.3	100	7.6	13.5	1.9	3.3	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	2.8	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	17.0	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	5.0	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
VM	2.9	3.3	12.5	0.7	2.8	0.5	45.3	9.4	100	0.3	0.2	0.4	0.0	2.0	0.0	0.1	0.0	0.0	0.3	0.1	0.8	0.1
GIS	2.0	3.5	2.5	2.2	3.8	0.3	3.9	6.2	1.5	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	5.1	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	0.8	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	0.9	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	6.9	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	0.1	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	1.4	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	0.1	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	0.0	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	4.3	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	5.7	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	6.6	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	1.3	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

2. Analysis and Results

2.1 Widespread use

We evaluate evidence for widespread use by calculating the Gini coefficients by three-digit NAICS industry. 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.[7] Thus, 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. We emphasize the values in 2019, and show the values in 2010 as well as data on the total number of job postings that mention the technologies.

Table 3 presents the data. Column 1 shows the Gini coefficients in 2019. Telecommunications and robotics are most widespread, followed by cloud computing, service-oriented architecture, and ML. The technologies that are not widespread are Web 2.0, virtual reality, polymer science, nanotechnology, and especially CRISPR and quantum computing.

Column 2 shows the same values for 2010. ML, robotics, 3D printing, and internet of things 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

[7] 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).

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.

Table 3: 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

2.2 Potential for innovation – Many research jobs

The definition of a GPT also requires that the technology is "capable of ongoing technical improvement." There are a variety of ways to define such improvement. All technologies studied continue to have related patents filed and papers published.

From the job posting data, we are able to assess whether these technologies are used in research jobs. This is a *necessary condition* for further innovation in the technology. If there is no hiring for the technology skill cluster in research jobs, it is unlikely the technology satisfies the requirement of “capable of ongoing technical improvement.” It is important to recognize that this is not a sufficient condition. 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.

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. These job postings include all research-related jobs that mention the technology skill cluster.

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

ML is near the top in both metrics. 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.

The data from 2010 provide further insight. Even as the number of jobs in ML and cloud grew substantially, the proportion of those jobs that are research jobs remained relatively flat. 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.

Table 4: 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

2.3 Many application sectors – Research jobs across sectors

As noted earlier, GPTs are particularly important technologies because of a positive feedback loop between innovations in the producing sectors and innovation in the applications sectors. In other words, GPTs are distinct because of innovation in the application industry that is complementary to the original innovation in the producing industry (Bresnahan and Trajtenberg 1995). 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. The GPT leads to additional innovation outside of the producing industry.

Table 5 therefore examines whether the technologies we study are used for research in a range of application sectors by repeating the analysis in section 2.1, restricted to research job postings. The set of technologies that were widespread across research jobs in 2019 includes cloud computing, robotics, internet-of-things, telecommunications, and especially ML. We interpret this to suggest that these five technologies are relatively likely to be GPTs, as defined by enabling innovation in applications sectors. Of all of the technologies examined, only telecommunications was widespread in many research jobs in 2010. 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.

Table 5: Enabling innovation in many 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

2.4 Data science, ML, and other technologies

In section 1.2, we identified six technologies that appeared in many of the same job postings and that all relate to the use of data: ML, business intelligence, big data, data mining, data science, and NLP. In Table 6, 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 jobs. We interpret this to suggest that ML is representative of a suite of technologies that together 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. We highlight ML because of the existing literature and the popular discussion of AI as a GPT; however, Table 6 suggests an important nuance. It is this cluster of technologies that together are likely to represent a GPT.

Table 6a: 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 6b: 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 6c: Enabling innovation in many 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

2.5 Comparison to other methods

We benchmark our results against Petralia's (2020) method for identifying the relative likelihood that technologies are GPTs using patent data. This method focuses on the quantity and growth of patents to measure widespread use, and the quantity of different patent classes in which the technology is used to measure use in application industries.

Petralia's method has two key strengths for benchmarking. First, it is quantitative, allowing us to compare our estimates to his directly. This is in contract to the early literature on GPTs, which used qualitative arguments to identify technologies as GPTs. Bresnahan and Trajtenberg (1995) identify examples of technologies that are widespread and involved 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 qualitative arguments do not provide a useful benchmark against which to evaluate our method. Second, it requires only a few years of data. The other existing patent-based quantitative methods for evaluating GPTs emphasize patent citations over a long period of time. Bresnahan (2010, p. 780) emphasizes that citations are useful because they "tell us of technological links between different inventions". Citation data, however, requires a great deal of time to pass in order to be quantitatively useful for measuring GPTs. Moser and Nicholas (2004) measure how patents from the 1920s were cited between 1976 and 2002. Hall and Trajtenberg (2006) and Feldman and Yoon (2011) also examine citations over a decades-long timespan.

Petralia's approach evaluates the relative "GP-ness" of different technologies (p. 2). The paper examines electricity and ICT patent classes from 1993 to 2014 to identify the particular classes that are relatively likely to be GPTs. The empirical results presented focus on the 2005-2010 time period. The GP-ness of technologies is based on three types of measures:

1) Wide scope for improvement: The number of patents in each patent class and the growth rate in the patent class.

2) Potential use in a wide variety of products and processes: A count of the number of other electricity and ICT that are referenced in the focal class patents.
3) Strong complementarity with existing and new technologies: A count of the number of other 3-digit patent classes that are listed in the focal class patents.

Of the technologies we examine, only one appears in Petralia's analysis: Telecommunications. He identifies telecommunications as relatively likely to be a GPT in the 2005-2010 data. This is consistent with our result from 2010 that telecommunications is relatively likely to be a GPT.

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 applied his method, using updated patent data (grant dates from 2000-2019, using www.patentsview.org) and our 16 categories of technologies. This required some minor changes to his approach, since most of our technology categories are not well-represented by 3-digit patent classes. Therefore, we used keywords to define the relevant patents. We drew the keywords from the Burning Glass skills that we used to identify the relevant technologies in our focal Burning Glass job posting data. This generated list of patents for each technology. From these patents, it is straightforward to calculate the wide scope for improvement and strong complementarity measures. In particular, we calculate three values for each technology category in each year:

i) Widespread use, as measured by a count of granted patents per technology per year.
ii) Widespread use, as measured by the seven-year growth rate in granted patents for each technology in each year.[8]
iii) Complementarity with other technologies, as measured by the count of 3-digit patent classes in which the technology keywords appear.

Given the need to use keywords rather than patent classes for most of our technologies, we could not determine how to generate a useful variant of his measure of potential use in a wide variety

[8] Petralia does not specify the appropriate time period for the growth rate. We use 7 years because it captures a long enough time period to reduce substantial short-term fluctuations. In the appendix, we show robustness to using 6 or 8 years.

of products and processes. Petralia uses both patent classes and electricity and ICT keywords to generate the potential use measure.

We compare these three measures from Petralia to our four measures. Table 6 shows that our measures are highly correlated with Petralia's in the expected direction, with the exception of disproportionate research jobs. Columns (1)-(3) show that the Gini coefficient on all job postings across industries is negatively correlated with more patents, a faster growth rate of patents, and more co-occurring patent classes. Since a lower Gini coefficient means more widespread, this suggests a strong correlation between our measures and Petralia's. Columns (4)-(6) show that more research jobs are positively correlated with Petralia's three measures. Columns (10)-(12) show that the Gini coefficient on research job postings is negatively correlated with Petralia's three measures. The exception to the consistency between our measures is the fraction of jobs in research, shown in columns (7)-(9). As documented above, with the exception of machine learning, the categories with very high fraction of jobs in research tend to be relatively weak in the other dimensions for GPTs in the jobs data. This seems to drive the results. As we show below, this category is, however, weakly predictive of future values of the Petralia measures.

Next, we test the degree to which our measures predict Petralia's. If our approach can indeed identify GPT likelihood earlier than a patent-based approach, our measures should predict the patent-based ones. The effects should be driven by the most emerging technologies. At any point in time, the set of hyped technologies tested using our method might include more established technologies, such as telecommunications. In those cases, we have no reason to expect that job posting measures (significantly) precede patent-based measures. In Table 7, we test the prediction power of our measures for all technologies, and in Table 8 we show that, as expected, our measures have the strongest prediction power for technologies that are newly emergent.

Table 7 shows that our measures weakly predict Petralia's measures five years later.[9] These are fixed effects regressions that use Petralia's measures as the dependent variable. The predictors are Petralia's measures and our measures, lagged five years. Columns (1) to (3) show that the

[9] In the online appendix, we show somewhat more significant results if we use different specifications, for example 4 or 6 year lags.

Gini coefficient on widespread use in hiring across industries is a strong predictor of all three of Petralia's patent-based measures five years later. Columns (4) to (6) show a weak positive correlation between the number of research jobs and Petralia's measures. Notably, our measure is at least as predictive as Petralia's lagged measure. Columns (7) to (9) show that the lagged fraction of jobs in research has no significant relationship with Petralia's measures, though the correlation is now in the expected direction for patent growth rate and for the count of co-occurring patent classes. Columns (10) to (12) show a negative but insignificant correlation between the lagged Gini for research jobs and Petralia's measures.

Table 8 repeats Table 7, but it separates out newly emerging technologies as of 2010 from more mature technologies. Several of our categories had fewer than 20 patents in 2010. Table 8 shows that for these categories, our method is especially predictive of Petralia's measures five years later. The coefficient signs are as expected for the emergent technologies in all twelve columns. Furthermore, most results are statistically significant and larger than the values for the less emergent technologies. In this way, Table 8 highlights a key strength of our method. We are able to identify the relative likelihood that a technology is likely to be a GPT even before many patents for that technology have been granted. It provides an earlier view of likely GPTs than Petralia.

In summary, this subsection has shown that our four measures of whether an emerging technology is likely to be a GPT are strongly correlated with the measures developed in Petralia (2020). Furthermore, our measures predict future values of Petralia's measures and this prediction is particularly strong for technology categories with few patents at the beginning of the sample period.

Table 6: Correlation between our job posting measures and Petralia's patent-based measures.

2010-2019	(1) Count of patents	(2) Growth Rate	(3) Co-occurring patent classes	(4) Count of patents	(5) Growth Rate	(6) Co-occurring patent classes	(7) Count of patents	(8) Growth Rate	(9) Co-occurring patent classes	(10) Count of patents	(11) Growth Rate	(12) Co-occurring patent classes
Widespread use (Gini by 3-digit industry)	-3.565*** (0.032)	-1.567** (0.719)	-1.911*** (0.071)									
Many research jobs (count of research jobs in hundreds)				0.012*** (0.000)	0.178*** (0.054)	0.006*** (0.000)						
Disproportionate research jobs (fraction of research jobs)							-7.816*** (0.103)	-1.458** (0.697)	-3.001*** (0.125)			
Widespread research use (Gini by 3-digit industry)										-4.223*** (0.038)	-2.973*** (0.944)	-2.158*** (0.089)
LL	-35,093.9		-4,699.24	-35,752.9		-4,625.84	-35,109.8		-4,849.20	-35,556.8		-4,775.55
R-squared	0.150	0.030	0.070	0.134	0.066	0.085	0.149	0.027	0.040	0.138	0.060	0.055
Observations	160	158	160	160	153	160	160	158	160	160	158	160

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. Columns 1, 4, 7, and 10 show Poisson regressions with count of patents as the dependent variable. Columns 2, 5, 8, and 11 show linear regressions with log(patent growth rate relative to seven years earlier+1) as the dependent variable. Columns 3, 6, 9, and 12 show Poisson regressions with a count of co-occurring patent classes as the dependent variable. *significant at 10%, **significant at 5%, ***significant at 1%

Table 7: Correlation between our job posting measures and Petralia's patent-based measures five years later

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Count of patents	Growth Rate	Co-occurring patent classes	Count of patents	Growth Rate	Co-occurring patent classes	Count of patents	Growth Rate	Co-occurring patent classes	Count of patents	Growth Rate	Co-occurring patent classes
Widespread use lagged 5 years (Gini by 3-digit industry)	-6.780*** (2.496)	-6.815*** (2.189)	-3.382*** (1.180)									
Many research jobs lagged 5 years (count of research jobs in hundreds)				0.016 (0.012)	0.175 (0.166)	0.010** (0.005)						
Disproportionate research jobs lagged 5 years (fraction of research jobs)							-0.165 (3.396)	0.703 (0.752)	1.212 (0.808)			
Widespread research use lagged 5 years (Gini by 3-digit industry)										-4.027 (2.505)	-0.616 (3.109)	-1.491 (1.286)
Dependent variable, lagged 5 years	0.001 (0.000)	-0.136 (0.141)	0.009*** (0.003)	0.000 (0.000)	-0.185 (0.181)	0.008* (0.005)	-0.000 (0.000)	-0.120 (0.170)	0.008* (0.005)	0.001 (0.001)	-0.127 (0.176)	0.008* (0.004)
LL	-4,079.1		-553.11	-4,289.7		-588.82	-4,686.4		-606.90	-4,163.1		-587.81
R-squared		0.014			0.034			0.036			0.056	
Observations	111	108	111	112	102	112	112	109	112	111	108	111

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. Columns 1, 4, 7, and 10 show fixed effect Poisson regressions with count of patents as the dependent variable. Columns 2, 5, 8, and 11 show fixed effect linear regressions with log(patent growth rate relative to seven years earlier+1) as the dependent variable. Columns 3, 6, 9, and 12 show fixed effect Poisson regressions with a count of co-occurring patent classes as the dependent variable. Standard errors clustered by technology category. *significant at 10%, **significant at 5%, ***significant at 1%

Table 8: Comparison of predictions for emerging and other technologies.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Count of patents	Growth Rate	Co-occurring patent classes	Count of patents	Growth Rate	Co-occurring patent classes	Count of patents	Growth Rate	Co-occurring patent classes	Count of patents	Growth Rate	Co-occurring patent classes
Widespread use lagged 5 years (Gini by 3-digit industry)												
Very emergent	-8.348**	-6.793**	-5.900***									
	(4.221)	(3.103)	(1.781)									
Relativity less emergent	-5.927**	-6.834**	-1.480									
	(2.824)	(3.077)	(0.933)									
Many research jobs lagged 5 years (count of research jobs in hundreds)												
Very emergent				0.030**	0.274*	0.040**						
				(0.014)	(0.134)	(0.018)						
Relativity less emergent				0.015	-0.225	0.009						
				(0.013)	(0.535)	(0.005)						
Disproportionate research jobs lagged 5 years (fraction of research jobs)												
Very emergent							0.733	0.843	1.539**			
							(1.080)	(0.657)	(0.612)			
Relativity less emergent							-2.513	-9.362	-1.723			
							(11.211)	(13.72)	(4.326)			
Widespread research use lagged 5 years (Gini by 3-digit industry)												
Very emergent										-17.88***	-11.95*	-13.55***
										(5.412)	(5.688)	(2.006)
Relativity less emergent										-2.660	1.395	-0.471
										(2.825)	(3.547)	(0.952)
DV lag	0.000	-0.136	0.007*	0.000	-0.196	0.007	-0.000	-0.131	0.008	0.000	-0.241	0.004
	(0.000)	(0.143)	(0.004)	(0.000)	(0.192)	(0.004)	(0.000)	(0.165)	(0.005)	(0.001)	(0.203)	(0.003)
LL	-4,061.2		-536.72	-4,271.1		-576.82	-4,668.3		-602.70	-3,710.7		-491.32
R-squared		0.013			0.073			0.001			0.102	
Observations	111	108	111	112	102	112	112	109	112	111	108	111

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. Columns 1, 4, 7, and 10 show fixed effect Poisson regressions with count of patents as the dependent variable. Columns 2, 5, 8, and 11 show fixed effect linear regressions with log(patent growth rate relative to seven years earlier+1) as the dependent variable. Columns 3, 6, 9, and 12 show fixed effect Poisson regressions with a count of co-occurring patent classes as the dependent variable. Standard errors clustered by technology category. *significant at 10%, **significant at 5%, ***significant at 1%

3. Conclusions

When new technologies appear, it is not unusual to find claims that these technologies are GPTs. Our results suggest that a suite of data-focused technologies—often represented as ML—are relatively likely to be a GPT. Cloud computing, robotics, and telecommunications also display some characteristics of GPTs.

GPTs are different from other technologies. If a technology is a GPT, there is a coordination problem between producing and application industries. The large productivity gains from GPTs occur when there is a 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 benefit may require several years to appear. Some of these benefits may accrue to intermediaries or end-users rather to the innovators.

Our results also show several technologies are unlikely to be GPTs. For some of these technologies, this is unsurprising. We view the result that RFID, CRISPR, Web 2.0, and service-oriented architecture are 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 as GPTs seem unlikely; notably for 3D printing, nanotechnology, the internet of things, and blockchain.

An important limitation of this analysis is that our measurement focuses on a particular time window in GPT diffusion. It will not capture GPTs that have already widely diffused and no longer have innovation in the applications sectors. It will also not capture technologies in very early stages. For example, quantum computing may someday become a GPT but the technology is not yet mature. In light of this limitation, an action-focused interpretation of our results is that it is too early for managers in using industries to expect a payoff from investment in quantum computing related innovations. In contrast, for the suite of technologies represented by ML, the managers that invest in complementary innovations will be those that reap the largest gains.

References

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

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 and 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. 2011. 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.

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

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

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

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

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.

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,	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,	64,335	2,842,778	2,907,113

			Machine Learning, Writing, Scikit-learn, Statistical Methods, Natural Language Processing, Experiments, SPSS, R, Pattern Recognition, Research, Critical Thinking, SQL, D3.js, SAP BusinessObjects, Qlikview, WEKA	Apache Hadoop, Microsoft Office, 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	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

			Management, Polymer Science, Numerical analysis, Technical Training				
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,	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

			Risk Management, Meeting Deadlines, Fixed Income, Writing, Pattern Recognition, Data Science, Statistical Methods, Equities, 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,	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,	53,241	3,364,855	3,418,096

			Oracle, Telecommunications, Java, Writing, Experimental Design, JavaScript, PERL Scripting Language, Communication Skills, SAP BusinessObjects, C++, Tableau, Microsoft Powerpoint, Verbal / Oral Communication, Microsoft Word, Statistical Analysis, Teradata	Energetic, Problem Solving, Repair, Cisco, Network Switches, Domain Name System (DNS), E-Commerce, It Support, Team Building, Hardware and Software Configuration, Planning			
Virtual Machines (VM)	2008	At least one skill in BG defined skill cluster "Virtual Machines (VM)"	ID: 38480201828 Title: Senior Data Engineer Role Employer: (not available) Degree-level: PhD Skills: R, UNIX, Quantitative Research, Computer Literacy, Amazon Redshift, Java, Apache Hadoop, Scala, Python, Apache Hive, AWS Elastic Compute Cloud (EC2), Planning, AWS Simple Storage Service (S3), Data Science, Bash, AWS Redshift	ID: 38472246790 Title: Unix/Linux Systems Administrator Employer: Raytheon Degree-level: Bachelor's Skills: Shell Scripting, Predictive / Preventative Maintenance, UNIX, System Administration, Preventive Maintenance, VMware, Python, Systems Engineering, Solaris, Linux, Troubleshooting, System Operation, Domain Name System (DNS), Amazon Web Services (AWS), Amazon Web Services (AWS), Engineering Support, PERL Scripting Language, NetApp, Oracle, Microsoft Active Directory, Citrix, Citrix, Lightweight Directory Access Protocol (LDAP), Debugging	15,712	1,614,223	1,629,935
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	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

			Analysis, Natural Sciences, Case Management, Public administration, SAS, Planning, Microsoft Access, Report Writing, Writing, Research Design				
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 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	ID: 38444276675 Title: PLC Programmer Employer: Diedre Moire Corporation Degree-level: (not available) Skills: Variable Frequency Drives (VFDs), Programmable Logic Controller (PLC) Programming, 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#	35,705	514,053	549,758
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,	ID: 38452828314 Title: High Vacuum Technician Employer: Texstars Llc Degree-level: Bachelor's Skills: Detail-Oriented, Manufacturing Processes, Quality Management, Plumbing, Repair, Robotics,	4,101	7,279	11,380

			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	Purchasing, Technical Support, Electronic Schematics, Programmable Logic Controller (PLC) Programming, Preventive Maintenance, Equipment Repair, Predictive / Preventative Maintenance, Quality Assurance and Control, Nanotechnology, Windows Programming			
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, 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	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, Product Development, Planning, Software as a Service (SaaS), Software as a Service (SaaS), Product Marketing	5,835	198,806	204,641
CRISPR (DNA logic and/or editing)	2005	At least one skill in BG defined skill clusters	ID: 38444688851	ID: 38414235752	7,549	3,339	10,888

		”CRISPR” or “CRISPR-DM”	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	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			
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 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	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

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 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	ID: 38491786895 Title: Process Engineer Employer: Cps Degree-level: Bachelor’s Skills: Machinery, Mechanical Engineering, Polymer Science, Process Engineering, Chemical Engineering	6,425	12,354	18,779

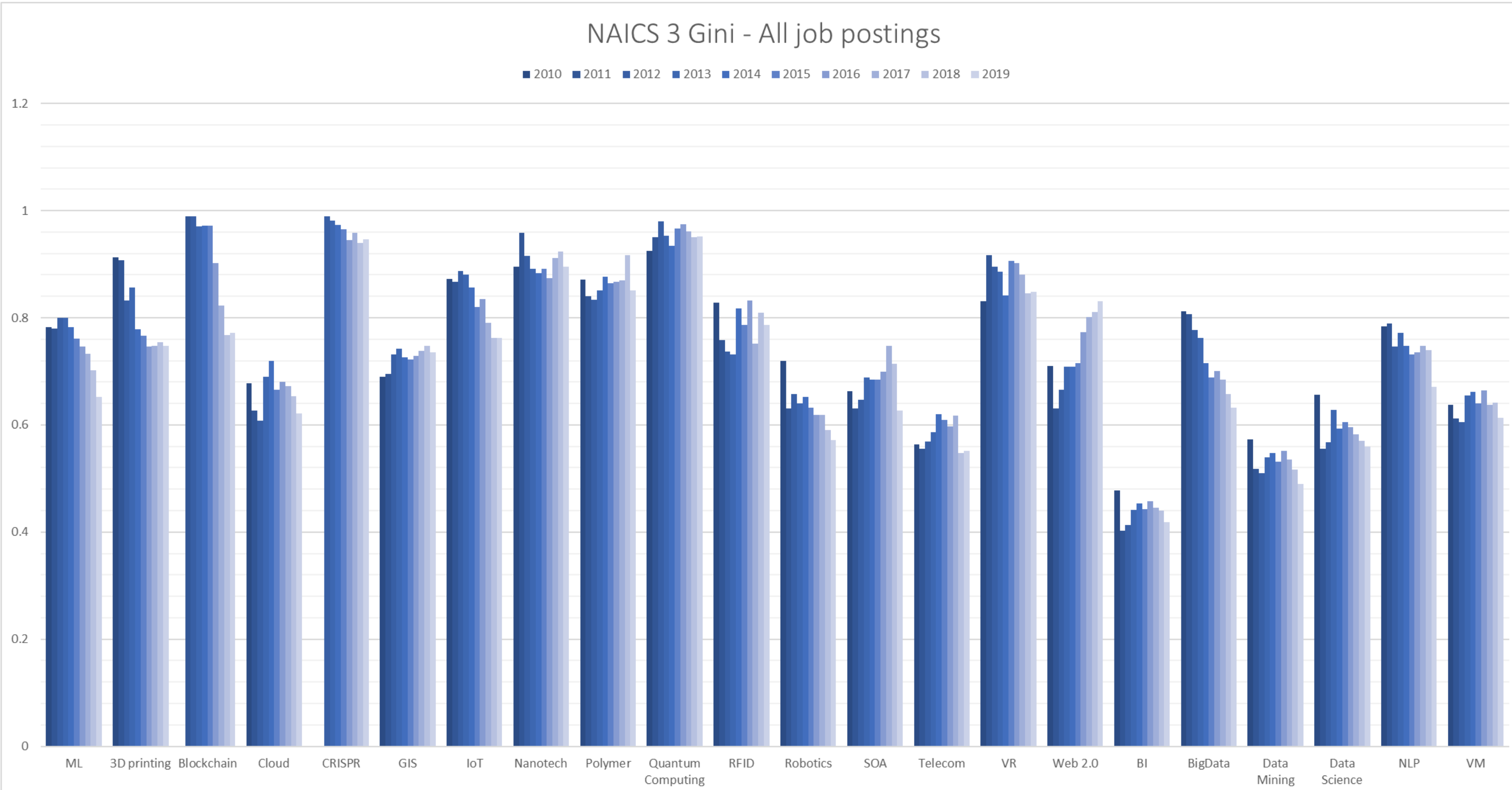
Blockchain	2016	At least one BG defined skill “Blockchain” or “Bitcoin”	ID: 38493674160 Title: Blockchain Researcher Employer: Anchorage Degree-level: (not available) Skills: Structured Methods, Economics, Research Reports, Cryptography, Algebra, Blockchain, Creativity, Onboarding, Research, Teamwork / 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 Engineering, Technology Transfer, Written Communication, Writing, System Design, Planning, Web Servers, Routers, PERL Scripting Language, Security Vulnerability & Penetration Testing, Vulnerability assessment	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 Data Center, Troubleshooting, Routing Optimization	1,653	101,416	103,069
Service-Oriented Architecture (SOA)	2004	At least one BG defined skill “Service-	ID: 38540054216 Title: Emerging Technologies Lead Systems/Software Engineer	ID: 38529201989 Title: Senior Software Developer .Net Developer	2,586	252,861	255,447

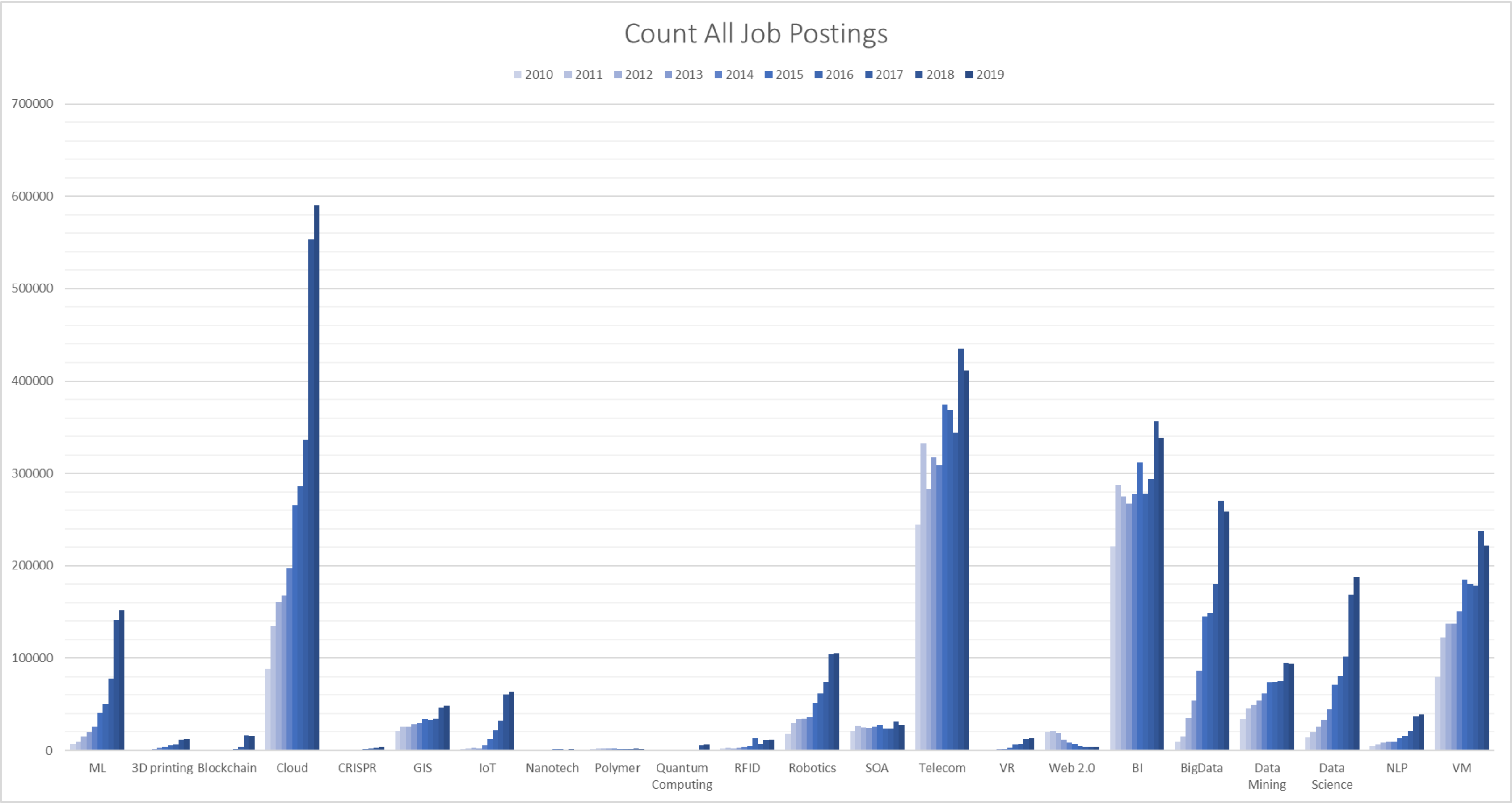
		Oriented Architecture (SOA)”	Employer: MITRE Corporation Degree-level: PhD Skills: Prototype Design Development, Written Communication, AJAX, Systems Engineering, Python, Software Engineering, C++, Service-Oriented Architecture (SOA), Java, Unified Modeling Language (UML), Object-Oriented Programming, 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	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			
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 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,	ID: 38472278381 Title: Secure Mobile Systems Engineer, Senior Employer: Booz Allen Hamilton Inc. Degree-level: Master’s 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 /	853	62,788	63,641

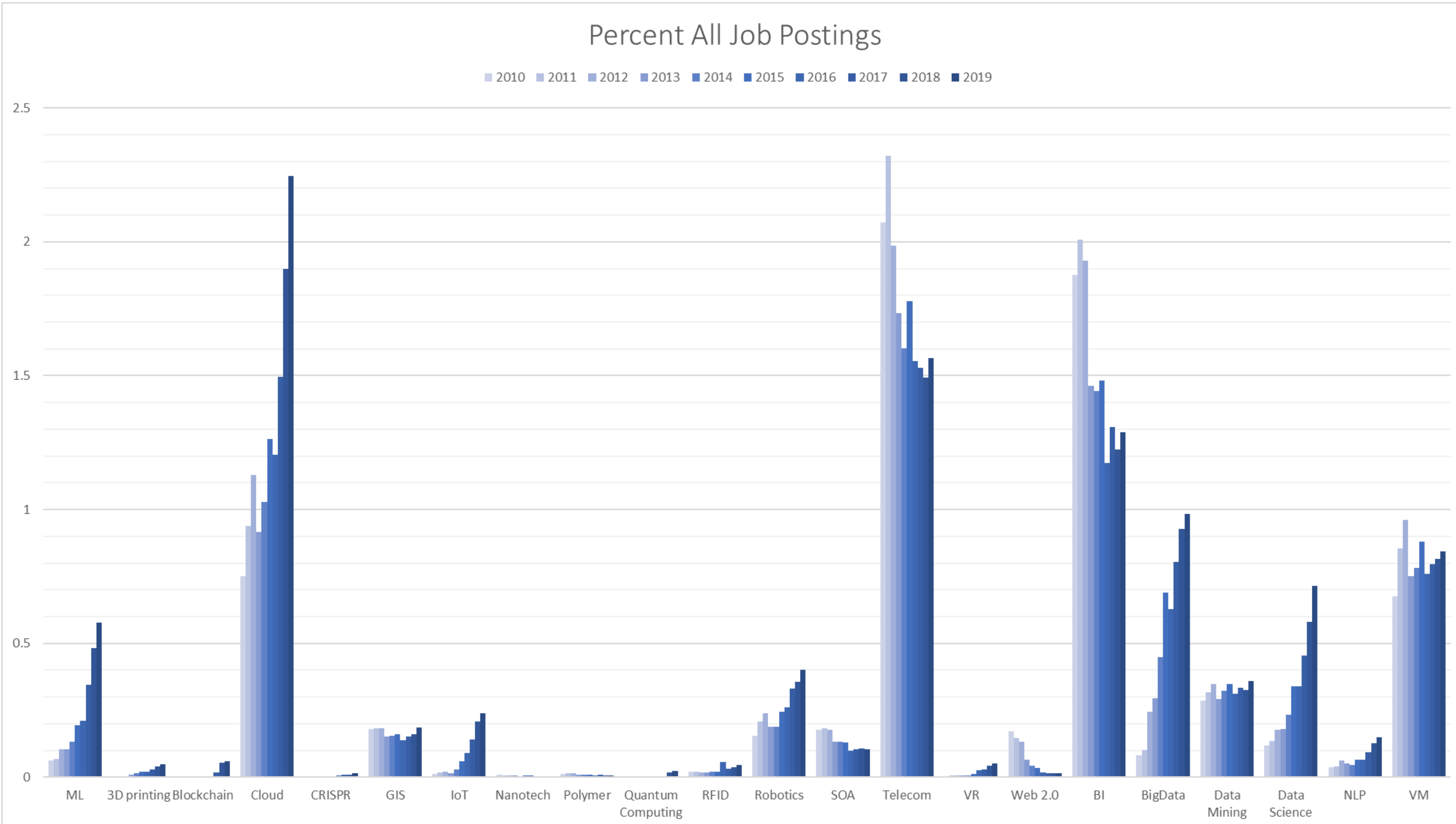
			Telematics, Graphics Processing Units (GPU)	Internet Protocol (TCP / IP), Certification & Accreditation, Configuration Management, Objective C, Linux, Information Systems, Bash, Microsoft C#, Building Effective Relationships, Software Customizations, Ruby, Software Development, Chef Infrastructure Automation, Hardware and Software Configuration, Systems Management			

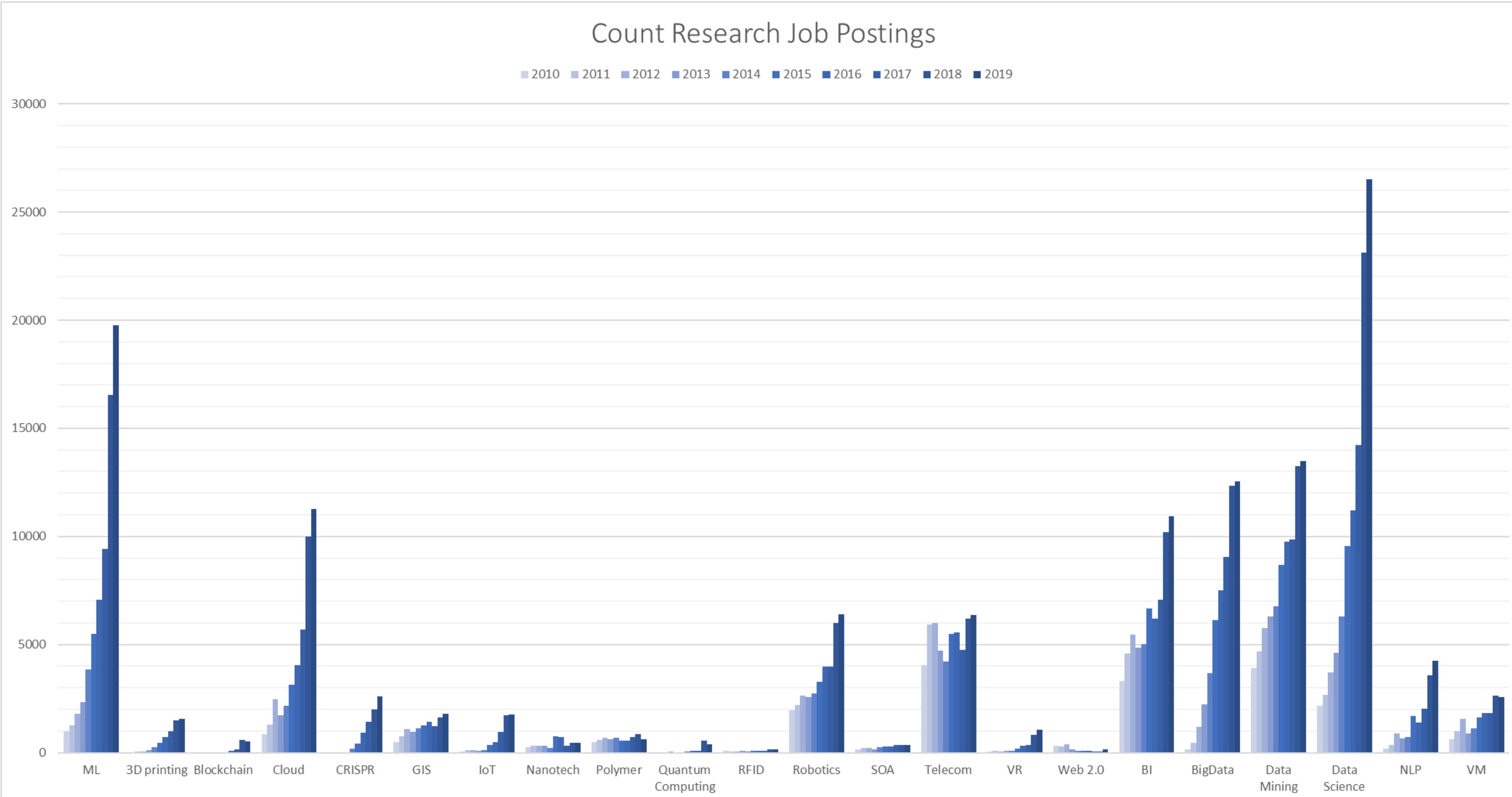
Note: Research jobs defined as such if at least one skill in BG defined skill clusters labeled as "…Research…" and referencing scholarly-type research (i.e. "Research Methodology", "Laboratory Research", "Medical Research" and "Clinical Research"

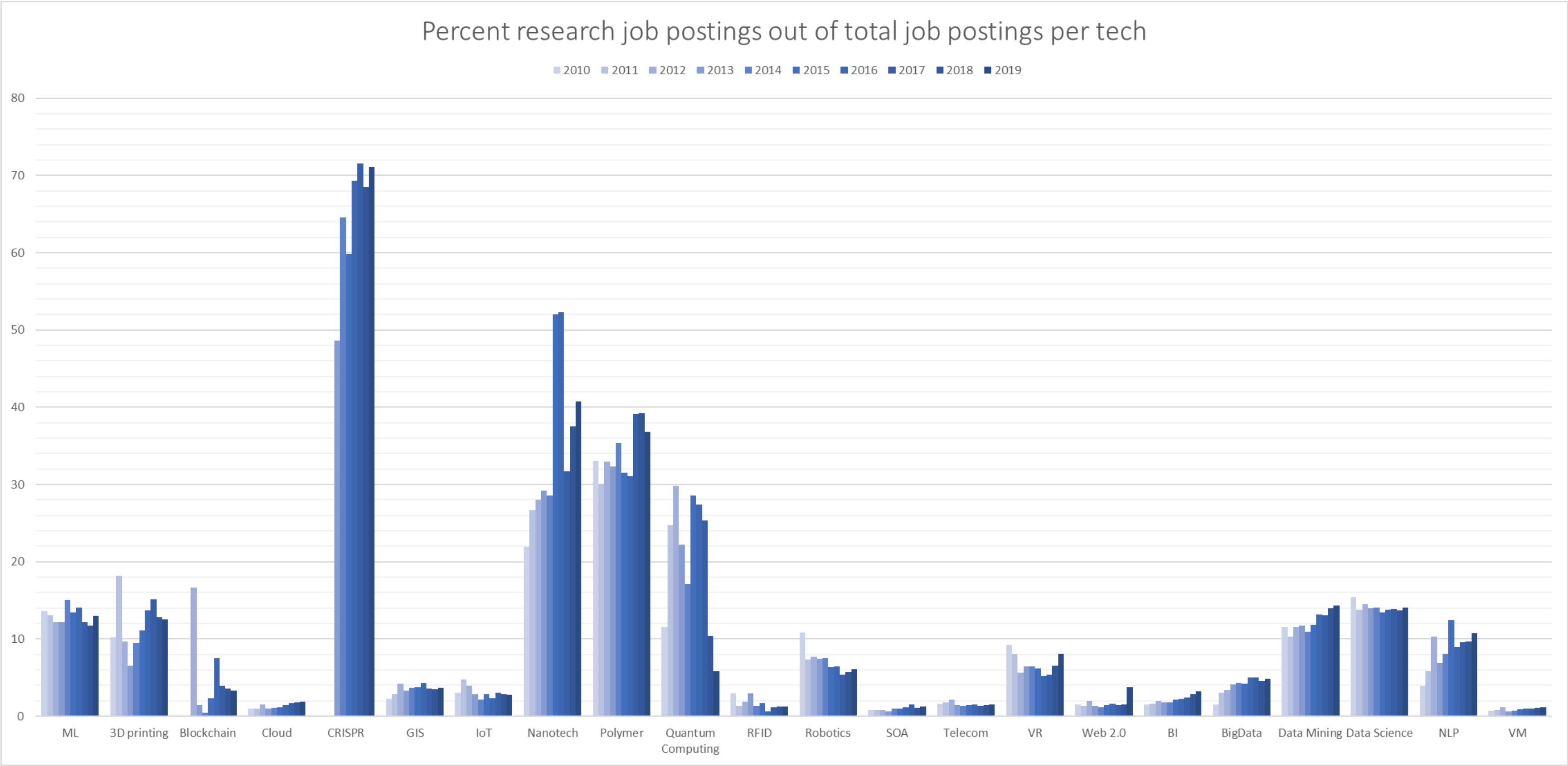
Online Appendix 2: Time-series data for the three GPT criteria (2010-2019)

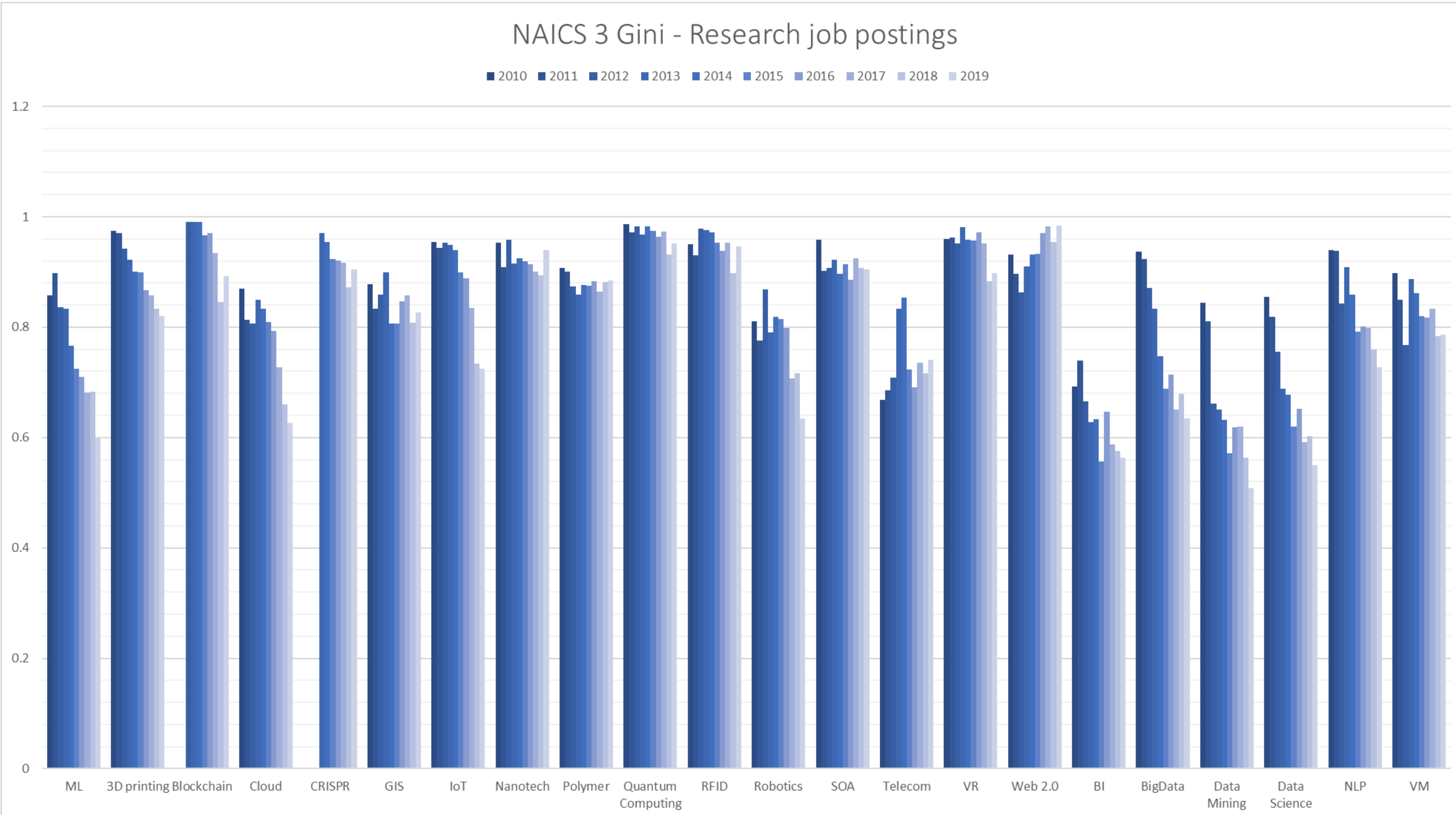



Count All Job Postings
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
700000
600000
500000
400000
300000
200000
100000
0
ML
3D printing
Blockchain
Cloud
CRISPR
GIS
IoT
Nanotech
Polymer
Quantum Computing
RFID
Robotics
SOA
Telecom
VR
Web 2.0
BI
BigData
Data Mining
Data Science
NLP
VM


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
VM


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
VM


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
VM



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
VM

Online Appendix 3: Additional descriptive statistics - Job postings by NAICS 2-digit industry sector (2010 and 2019)

Table A3.1. 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
VM	15	72	99	127	7970	485	1102	412	6950	3051	220	22006	142	4482	1217	1421	134	2585	307	759

Table A3.2. 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
VM	72	256	358	584	17556	683	4760	1528	14157	12525	749	48229	373	23877	3439	3769	500	3131	623	3998

Table A3.3. 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
VM	0	0	3	0	103	1	1	3	91	3	0	196	2	26	9	17	0	18	1	6

Table A3.4. 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
VM	0	3	1	2	253	3	57	10	165	177	6	692	1	254	92	62	2	36	8	71