Algorithmic Bilinguals

Prasanna (Sonny) Tambe*
The Wharton School, U. Pennsylvania

July 20, 2023

Abstract

Using workforce data from US firms, this study tests the hypothesis that generating value from algorithms requires employing domain experts who can also effectively interact with data and algorithms. This decentralization of technical human capital stands in contrast to earlier generations of business technologies for which the complementary skills were primarily embodied in IT specialists, and is due to the task complementarities associated with integrating decision-making algorithms into a production framework. Using two different data sets, I show that 1) employers have been shifting hiring towards requiring greater technical expertise from domain experts, 2) technical human capital in frontier firms has become more dispersed across occupations, and 3) the market assigns higher value to firms' investments in algorithms when they have also made these complementary workforce changes, indicating the presence of valuable intangible assets that can yield future productivity benefits. Implications for training, education, and algorithmic decision-making are discussed.

Keywords: human capital, algorithms, IT intangibles, future of work, IT complements, digital literacy

1 Introduction

The potential impact of algorithmic decision-making technologies on organizations is a topic of rapidly growing academic interest (Rock, 2019; Wu et al., 2019; Agrawal et al., 2018; Zolas et al., 2021). Much of the research in this area focuses on the labor reallocation effects of automation and AI-based technologies (Acemoğlu and Restrepo, 2016; Autor and Salomons, 2018; Brynjolfsson et al., 2018; Eloundou et al., 2023), with some of this work demonstrating that these technologies are not simply labor displacing (Agrawal et al., 2019; Gregory et al., 2022). New technologies also are likely to generate new jobs and new types of jobs (Bessen, 2019; Autor et al., 2022), and scholars have begun exploring how humans can be most effective when working alongside algorithms (Faraj et al., 2018; Cowgill, 2018; Beane, 2019; Agrawal et al., 2019; Lebovitz et al., 2022). These perspectives include questions about how the workforce should be re-skilled to meet the changing needs for human capital from industry.

This paper develops new theory and evidence that a complement to the use of these new technologies is the coupling of domain knowledge and expertise with algorithms within many jobs (Collins, 2004), requiring "algorithmic bilinguals" to fill these positions. Domain knowledge is expertise accumulated in a business domain such as customer retention, manufacturing quality, or healthcare delivery outcomes. This paper focuses on technical expertise related to algorithmic technologies, such as data science, machine learning, and AI tools, which enable firms to convert streams of data into strategic decisions in the pursuit of business goals. Prior work suggests that both domain knowledge and technical expertise are important when placing predictive technologies into a production context, particularly in sensitive contexts like law or medicine, where the payoff function for a decision is difficult to define or where the tolerance for machine-based prediction error is low (Kleinberg et al., 2018; Choudhury et al., 2020).

This migration of some level of algorithmic familiarity to jobs requiring domain knowledge stands in contrast to one in which firms' technical expertise is principally confined to specialized information technology (IT) workers. The explanation for these changes builds on the literature related to specialization, job design, and productivity (Smith, 1776; Becker and Murphy, 1992; Teodoridis, 2017) and is based in the notion that information task complementarities raise the returns to bundling these skills together, particularly in uncertain environments (Lindbeck and Snower, 2000; Postrel, 2002; Mao et al., 2019). This argument generates hypotheses related to how firms can be expected to adjust their workforce when they use algorithms for decision-making, and how workforce adjustments can generate market returns for algorithm-intensive firms.

The paper uses workforce databases to test these hypotheses in the context of empirical shifts that have been occurring over the last decade in US firms. The first data source captures a "near-universe" of job listings issued by US firms and has been used in prior work on the changing skill requirements of jobs (Deming and Kahn, 2018; Acemoglu et al., 2022) and to track the spread of new technologies (Goldfarb et al., 2023). The second data source is a fourteen-year panel of how workers with different technical skills move across different occupations in different firms over

time. We combine these data with administrative data on the knowledge content of occupations from the Bureau of Labor Statistics O*NET database and with employers' financial data from the Compustat-Capital IQ database. Neither of these workforce databases can capture the depth of technical expertise required for jobs, but how technical skill plays a role at the extensive margin in the job search and hiring process may itself be informative.

There are three main findings. First, using the job listings data, I show that algorithmic skill gradually spread across occupations from 2010 to 2016; these skills became more widely dispersed across occupations than those related to other business technologies. By 2016, only one-third of this human capital was embedded in IT jobs. This stands in contrast to skills such as network administration or database management, which has remained highly concentrated in IT jobs. Moreover, skills related to algorithms were most likely to migrate to occupations requiring domain knowledge, which is consistent with the argument that informational task complementarities drive these bundling choices.

Second, these trends are consistent with employer workforce changes over a recent fourteen year sample period (2008-2022). In high-value firms, skills related to algorithms have become dispersed across occupations, which as in the job listing data, stands in contrast to other technologies which have not experienced similar changes. From 2008 to 2022, the steady migration of these skills led to a 5% reduction in the fraction of these skills that reside in IT occupations. By contrast, there was no change in this figure for most technical skills.

Third, these workforce adjustments are generating productive, intangible assets for firms. This finding is in the spirit of prior work that has investigated how employers adjust their workforce in response to earlier computing technologies (Black and Lynch, 2001; Bresnahan et al., 2002). Financial markets assign higher value to public firms with investments in algorithms when they make complementary workforce adjustments, which suggests that firms derive the largest benefits from these technological investments when their professionals in marketing, service delivery, R&D, product management, and other disciplines have the skills to effectively integrate these technologies into business decisions. These findings are persistent to including employer fixed-effects, which can account for time-invariant differences in employers. Similar patterns on market value are not observed for other technologies, suggesting that the returns to organizing human capital this way are specific to algorithmic technologies.

This study contributes to the academic literature in two areas. First, with its focus on firms, it builds on a literature identifying organizational and human capital complements to information technology (Bresnahan et al., 2002; Black and Lynch, 2001; Caroli and Van Reenen, 2001; Bartel et al., 2007; Bloom et al., 2012). These analyses have been rooted in a perspective based in IT as a technology that performs "routine" tasks, but the application of algorithmic technologies to contexts where decision rules are not easily codified has reopened the discussion on how IT affects labor force needs (Brynjolfsson et al., 2018). This paper contributes to emerging work that examines management practices that complement investments in predictive algorithms (Brynjolfsson et al., 2021; Zolas et al., 2021; Dixon et al., 2021; Xue et al., 2022).

Second, it contributes to a literature on how the emergence of new AI and automation technologies will shape the future of work, an increasingly important area of research as new technologies subsume many of the tasks done by humans while simultaneously generating new areas of demand for human labor (Agrawal et al., 2019). Most prior work on technical skills has focused on the IT workforce (Ang et al., 2002; Levina and Xin, 2007; Mithas and Krishnan, 2008; Wiesche et al., 2019; Tambe et al., 2020), but there has been limited work on the implications of technical human capital for broader workforce outcomes (Deming and Noray (2018) is one exception). The absence of work in this area is important given the growing demand from students and workers from all disciplines for "coding" and other technical skills. These findings, therefore, contribute to our understanding of how the connection between humans and algorithms will shape the demand for skills as employers continue to embrace algorithmic decision-making.

2 Theory and Hypothesis Development

2.1 Background literature

2.1.1 Information technology and the workforce

A large literature analyzes the effects of IT adoption on firms' demand for labor and the impact of IT on changes to the skill content of the workforce. The spread of computing technologies over the last six decades has led to "skill-biased technical change", which can be defined as a relative increase in the demand for college-educated workers (Berman et al., 1994; Bresnahan et al., 2002; Bartel et al., 2007). IT-based production methods complement education because by automating routine tasks, they raise the relative value of front-line workers who solve problems creatively and make effective decisions (Autor et al., 2003). When authority is allocated to front-line decision makers, this system of changes can yield higher productivity (Bresnahan et al., 2002), making IT investments particularly valuable in turbulent environments where the value of decisions depends on rapidly changing external conditions (Mendelson and Pillai, 1998; Pavlou and El Sawy, 2006; Tambe et al., 2012). For example, hotel desk agents can rapidly adapt to changing customer preferences and factory floor workers trained in problem-solving can fix manufacturing problems as they arise.

These workforce changes have also been linked to higher value for firms that adopt these practices. For instance, Brynjolfsson et al. (2002) show that public firms that invest in computing technologies and decentralized work practices are more highly valued in the market. Black and Lynch (2001) provide evidence that decentralized decision-making, when combined with high-powered incentives, are associated with more productive firms. Other studies in this literature most often develop measures of workplace organization based on survey instruments administered to large samples of employers (Caroli and Van Reenen, 2001; Bresnahan et al., 2002; Bloom et al., 2012) and they connect these measures through a production framework (or hedonic market value regression) to firms' revenue (market value). This body of literature has provided compelling evidence that generating value from computing technologies requires employers to restructure their human capital

and to change the methods they use to manage these workers.

Employers can change their human capital by changing their mix of occupations or by changing the mix of skills within occupations, which is the focus of this paper. The literature on job design dates back to Adam Smith's treatise on specialization (Smith, 1776). An advantage of specialization is that workers become more productive with tasks through repetition (Rosen, 1983). There is a trade-off, however, between a) the productivity benefits that arise from specialization and b) the costs of coordinating activities across workers (Becker and Murphy, 1992; Hart and Moore, 2005), where coordination costs are the costs of information exchange when executing interdependent activities across workers. The manager's problem is to balance the productivity gains that arise from specialization against the coordination costs incurred by agents who have to exchange complex information.[1]

Information technologies, by shifting the information content of jobs, have been associated with a change in job design, and specifically with a move towards multi-task work. Lindbeck and Snower (2000) argue that informational task complementarities in knowledge-rich jobs have been driving a move away from specialization towards "holistic" work. Multi-task work organization is productive because workers' productivity in a task is interdependent with their levels of activities in other tasks. In work contexts such as these, bundling tasks to avoid the need for information exchange between employees can be beneficial (Postrel, 2002). Part of the benefit of an educated workforce is that educated workers can more effectively adapt to a multi-task job design such as this one.

The trade-offs between specialization and multi-task are closely related to the knowledge domain and the costs of acquiring new knowledge. Using academic publication data, Teodoridis (2017) shows that a decrease in the cost of acquisition of new technical knowledge changes the optimal mix of expertise when constructing diverse knowledge teams. By doing so, they can derive the greatest productivity benefits from informational complementarities that arise when engaged in production, especially when response time must be rapid. Dessein and Santos (2006) relate environmental uncertainty and adaptation to job specialization, and argue that improvements in communication technology result in more adaptive organizations with less specialized employees. This literature suggests that the costs of acquiring new knowledge and the interdependent nature of tasks and human capital acquisition have implications for how jobs are structured.

2.1.2 The emergence of data-driven technologies

The existing literature is predicated on technologies that excel at performing routine tasks (Levy and Murnane, 1996). Modern predictive algorithms, however, are distinguished in their ability to make predictions even when the relationship between inputs and outputs is not easily specified (Brynjolfsson et al., 2018; Agrawal et al., 2018). Just as importantly, data-driven technologies are

---

[1]Here we focus on this trade off, although scholars have of course also examined the role of other factors such as incentives, output measurement, and incomplete contracts in influencing job design (Holmstrom and Milgrom, 1991; Baker and Hubbard, 2003).

often designed to directly facilitate decisions. These changes raise the question of whether the workforce changes that complement personal computing – employee discretion and high-powered incentives, for instance – remain central to a work context where algorithms are intensively used for decision-making.

Decision-support technologies are not new. They date back at least to the 1970's (Shim et al., 2002) and provide information at the point of decision. With new data science and AI applications, an important question relates to the interdependence of data and decisions. The modern data science process can be typified by *CRISP-DM* (Chapman et al., 2000).[2] A typical workflow has six essential phases: Business understanding, Data understanding, Data preparation, Modeling, Evaluation, and Deployment. Challenges with application of data science techniques, however, have motivated a large body of literature on the challenges associated with successful application of domain knowledge to different stages of the data science process (Mao et al., 2019; Choudhury et al., 2020; Park et al., 2021).

There has been growing recognition in data science of the importance of workers who can synthesize technical skill and domain expertise. The most visible example is the famously tight labor market for "data scientists" who, by definition, combine technical and statistical skills with domain expertise (Davenport and Patil, 2012; Provost and Fawcett, 2013). The importance of domain expertise for effective data science has been discussed online[3], in industry panels[4], and increasingly, in the business press (Oostendorp, 2019). Beyond data scientists, there has been growing recognition that "unicorns", who couple domain expertise with technical skill, are becoming essential in many algorithmic decision-making contexts (Jha and Topol, 2016).

The educational community has also started to respond to these changes. For instance, the notion that data-driven employers increasingly demand "bilingual" workers (i.e. individuals who have both technical skills and subject matter expertise) was underscored by an announcement from MIT on their investment in a new College for Artificial Intelligence.

> The goal of the college, said L. Rafael Reif, the president of M.I.T., is to "educate the bilinguals of the future." He defines bilinguals as people in fields like biology, chemistry, politics, history and linguistics who are also skilled in the techniques of modern computing that can be applied to them. Additionally, it is expected that the "bilingual" graduates who emerge from this new College — combining competence in computing and in other fields — will be of enormous value to employers. [5]

2.2   Hypotheses

The background literature above suggests that the application of data science and AI in a production context, by introducing new challenges related to coordinating domain knowledge with effective data modeling, analysis, and application, amplify the productivity benefits that arise when hiring

---

[2]Annual polls indicate that CRISP-DM is the most widely used of data science processes. For example, see https://www.datascience-pm.com/crisp-dm-2/.

[3]For example, see Is domain knowledge necessary for a data scientist? Accessed on March 11, 2019.

[4]A video of one such industry panel is captured here: https://youtu.be/qKcUsIqoSHE.

[5]*New York Times*, Oct 15, 2018. MIT Plans College for Artificial Intelligence, Backed by $1 Billion.

employees that can synthesize both types of knowledge. For instance, in healthcare applications, a user of an algorithm must understand how to model complex medical data, how to interpret diagnostic output, and how to assess the relative costs of misdiagnosis.

For machine learning, the most high-profile algorithmic tool in the current business media environment, there are many tasks that could require subject matter expertise. In the CRISP-DM model referenced above, almost of these steps require significant context. Therefore, users of these technologies must understand the tradeoffs required when choosing which data to include in a model, what problems might arise with training data, how to construct features, what models strike the right balance, and how to assign value to the costs of prediction errors (Kleinberg et al., 2018; Cowgill, 2018). Moreover, data analysis is an inherently iterative process which involves a substantial amount of learning-by-doing, so it has been argued that exploration, experimentation, and learning in the data science process favors generalists over specialists (Colson, 2019). Based on these observations, we argue that skills related to the use of algorithms are more likely to be bundled with domain knowledge. The first hypothesis tested in this paper is:

> **H1**: *Skills related to the use of algorithms are more widely dispersed across occupations than skills related to other information technologies.*

Coordination costs may be particularly high when the domain is complex. For instance in domains like medicine or engineering, even the process of creating features from the raw data can require a great deal of subject matter expertise. In pharmaceutical industries, there has been some recent empirical evidence of the importance of embedding the relevant human capital in downstream occupations to achieve successful innovation outcomes (Wu et al., 2019). Lebovitz et al. (2022) describes the challenges arising in a hospital setting when interpreting the output accuracy of machine learning tools, and Jha and Topol (2016) argues for the importance of medical practitioners acquiring the skills required to understand the output of predictive algorithms. The second hypothesis tested in the paper is:

> **H2**: *Skills related to the use of algorithms are more likely to be embedded in non-technical occupations when they require significant levels of domain knowledge.*

Firms pursue these workforce adjustments because they are a complement to the use of algorithms in production. As with other economic complements to IT investment, the benefits of investing in these workforce complements can be expected to raise the value of investing in the technologies themselves (Black and Lynch, 2001; Bresnahan et al., 2002). In the case of algorithms, prior work has demonstrated that training in domain knowledge benefits the quality of the output produced by software developers (Cowgill et al., 2020). Empirically, we should observe that firms that co-invest in these two factors of production – algorithmic technologies and the employment of workers with domain knowledge who also have expertise in algorithms – should be assigned higher values by the financial markets. These higher market values reflect the valuable intangible assets that can be expected to eventually yield productive benefits. The final hypothesis tested in this paper is:

H3: *Financial markets assign higher value to firms that use algorithmic technologies when algorithmic literacy is dispersed across occupations at the firm.*

These hypotheses form the core of the analysis below, and in the next sections, we discuss the data, definitions, and approaches used to test these hypotheses.

3   Data and measures

3.1   Data sources

3.1.1   Job listing data

When employers seek job applicants, they often post job details online on their corporate web sites or on job boards. These listings generally contain information that identifies the employer and the job title, the geographic location of the open position, the skills and education sought from candidates, the wages offered, and other fields relevant to the job search process. This study measures when specific skills begin to appear in these online job ads and how skills co-occur in these listings. Data on job listings have been used in a number of papers studying the changing skill requirements of the IT workforce (Todd et al., 1995; Slaughter and Ang, 1996; Gallivan et al., 2002; Lee and Han, 2008). This study uses data from Lightcast, a labor market analytics firm that 1) uses software to crawl a "near-universe" of online job postings and 2) uses natural language processing to parse skills and other job information.[6]

This data provider uses proprietary software to collect and standardize data from over 17,000 job boards and corporate web sites, and these listings data are processed to ensure that a job listing is not counted multiple times if an employer posts it several places on the web. The processed data include posting date, job location (metropolitan area), employer name, job title, educational degrees required for the position, any certifications required for the position, and the skill expectations required for each job. A growing number of studies have been published that use this data source to study labor markets (Hershbein and Kahn, 2018; Deming and Kahn, 2018; Modestino et al., 2019), includng how AI related skills spread across jobs and industries (Acemoglu et al., 2022; Goldfarb et al., 2023).

The job title for each listing in the database is associated by the data provider with a standardized occupational code (BLS O*NET codes) and the employer for each listing was tagged to a North American Industry Classification Systems (NAICS) industry. Each job opening is associated with a list of skills, such as *Python*, *Random Forest*, *Chemistry*, *Supply Chain*, *Accounting*, *Data Science*, *Teamwork*, or *Communication*. The skill data in the listings are standardized using a skill dictionary created and maintained by Lightcast. Of course, the skill data should not be strictly interpreted as "requirements". Employers may omit skills from a listing, some skills may be assumed

---

[6]Until June of 2022, Lightcast was known as "Burning Glass Technologies", and is referred to as such in much of the prior work that has used this data set. In this paper, for consistency, we use the name Lightcast throughout, including when referencing the use of these data in prior papers.

but not listed, and there may be successful candidates who do not have some of the skills in a listing. Nonetheless, employers are thoughtful about the skills they place in listings because the inclusion of any skill can attract or repel the wrong type of applicant. Therefore, these listings are likely to provide useful information about aggregate trends in the market.

The provenance of these data also raises questions about the sampling frame. Prior work has provided useful information on the sampling properties of the data. See, for example, Appendix A of Deming and Kahn (2018) who make comparisons of these Lightcast data with administratively collected data sources. Key findings from their comparison are that these job listing data are over-represented in computer and mathematical occupations, as well as in management, health care, business, and financial occupations. They represent IT workers particularly well. They are a less robust indicator for labor in blue-collar occupations.

3.1.2 Corporate workforce data

The workforce data used in this study were collected through a partnership with a workforce intelligence company called Revelio Labs.[7] These databases are constructed from a variety of data sources including online career profiles and federal databases.[8] The Revelio data are similar in informational content to that posted on online professional networks such as LinkedIn, and cover a large fraction of white-collar work in the US. The data cover public US firms, and many private firms as well, but only public firms are used in this study so that these workforce data can be connected with financial market data, which are only available for public firms.

These workforce data are processed to develop measures of annual firm-occupation-skill employment activity dating from 2008 through 2022. This panel captures skills related to different technologies embedded across occupations in a firm's workforce. For instance, the data report how many employees in each occupation have "machine learning" skills in each year. Moreover, the data contain CUSIP identifier codes for employers, and so these workforce data can be joined with external firm-level financial databases such as the Compustat-Capital IQ data (described below). Together, these data and external identifiers provide scope for an analysis of how the distribution of technical human capital at the firm is associated with firm value.

As with the job listing data, questions arise about the sampling frame about the corporate workforce data. These data cover a large fraction of the US workforce; the provider purports to collect information on most US workers. It is useful, however, to conduct comparisons of these data with administrative data sets. These comparisons are provided in Appendix A (appendix to be completed).

---

[7]See https://www.reveliolabs.com/

[8]Scholars have compellingly argued that the lack of firm-level data on workforce skills is a constraint for understanding how firms are adjusting to technological change (Frank et al., 2019; Raj and Seamans, 2018).

3.1.3 Additional data sources

To create job expertise measures, job listing data are connected with the Occupational Information Network (O*NET) content model published by the Bureau of Labor Statistics.[9] The O*NET database has been widely used in academic research,[10] is government administered, collected by surveying occupational experts, and provides information on employment, wages, and the work content of US jobs. The O*NET taxonomy reports work requirements including the knowledge needed for different occupations. A diagram of the O*NET data relationships is shown in Appendix B. The O*NET data is periodically revised to reflect the changing structure of the US workforce. Although it was revised in 2019, I use the version of the taxonomy from before this revision in order to match the O*NET codes in the Lightcast data, which in our version of that database were also based on the taxonomy structure before the O*NET revision took place. Finally, some of the analyses use financial data from Compustat-Capital IQ which was collected through WRDS data services.

3.2 Construction of key measures

3.2.1 Algorithmic expertise

The key unit of analysis for the paper is worker skills. A challenge when analyzing large volumes of skills data is always the development of classifications that can provide meaning to groups of skills.[11] This absence of standardized taxonomies on the skill content of jobs is reflected in the existing academic literature, where there have been notable efforts to develop meaningful taxonomies around IT skills (Lee et al., 1995; Niederman et al., 2016). Empirical papers that study large quantities of archival, digitally collected skill data have also used manually constructed mappings of skills to conceptual measures (Deming and Kahn, 2018). Even foundational papers in the economics literature in this area have required the authors to use their own discretion (or those of colleagues or experts) to identify which skills in a database are most relevant to their phenomenon of interest (Autor et al., 2003). The small number of papers in the emerging literature on the impact of AI technologies have also generated taxonomies based on their own judgment (Brynjolfsson et al., 2018).

This paper takes a simpler approach, which is to rely on the categorizations provided by the data providers. Both data providers have engaged in extensive efforts to use data-mining to group skills together into different business technology areas, such as "data science", "AI", or "Big data". For this analysis, we use the taxonomies generated by these providers to assess whether listings or profiles have skills related to a particular business technology group. We group data science and AI skills together and refer to them as "algorithmic" tools and we conduct comparisons

---

[9] See https://www.onetonline.org.

[10] Notable examples include (Autor et al., 2003).

[11] Indeed, precisely because of the growing interest in the "future of work", the construction of taxonomies that can make sense of emerging sources of skills data and inform career development pathways is an active and ongoing area of research among businesses and information agencies. For example, see recent efforts by Nesta in the UK or Lightcast.

between these technologies and other business technologies that do not fall into this category. We focus on data science and AI because, as discussed above, they are the key inputs to the recent wave of technologies that directly make decisions, and our theoretical arguments are based on the coordination costs that arise when these automated decision-makers are directly integrated into a production context. Appendix details which specific skills fall into each of the focal categories used in this analysis.

3.2.2 Domain knowledge

We encode jobs according to the domain knowledge they require. Domain knowledge is "knowledge of a specific, specialised discipline or field". The O*NET program curates a list of all of the possible knowledge domains with which US jobs may require workers to be familiar.[12] Knowledge domains were extracted from the "Knowledge" O*NET table,[13] which delineates "organized sets of principles and facts applying in general domains."[14] It then assigns occupational titles to these knowledge requirements. For example, the Accountant occupation receives high scores in this database for the level of knowledge they require in the "Economics and Accounting" domain.

I use this dictionary of occupational knowledge domains to encode the domain knowledge required for job listings in the following way. Job listings are encoded as requiring domain knowledge (binary) if the skill data required for the job listing includes one or more of the terms identified in the O*NET dictionary of knowledge domains (e.g., Mathematics, Transportation). One limitation of these data and this approach is that it cannot identify differences in the intensive margin of use for the different knowledge domains.

It is useful to contrast this approach with one in which job listings are identified as requiring domain knowledge based solely on their job titles. Since the O*NET database assigns knowledge requirements to job titles, an alternative approach assigns job listings to knowledge content based on the job title and O*NET measures for that title, omitting the skill content data for construction of this measure. It is important to note that this implies that job listings for the same title can differ in the knowledge they require. One listing for "Financial Manager" may require knowledge of "Accounting" while another may not. This is important because it allows for a stronger test of the argument that knowledge, rather than occupation, is driving the bundling of skills studied in this paper.

---

[12]See https://www.onetonline.org/find/descriptor/browse/Knowledge/.

[13]See https://www.onetonline.org/find/descriptor/browse/2.C.

[14]The domain categories identified in the O*NET knowledge set are Administration and Management, Biology, Building and Construction, Chemistry, Clerical, Communications and Media, Customer and Personal Service, Design, Economics and Accounting, Education and Training, English Language, Fine Arts, Food Production, Foreign Language, Geography, History and Archeology, Law and Government, Mathematics, Mechanical, Medicine and Dentistry, Personnel and Human Resources, Philosophy and Theology, Physics, Production and Processing, Psychology, Public Safety and Security, Sales and Marketing, Sociology and Anthropology, Therapy and Counseling, and Transportation. From this list, Computers and Electronics, Engineering and Technology, Telecommunications, and Mathematics were removed because they overlap with the notion of algorithmic expertise.

3.2.3 Other job attributes

Beyond technology skills and occupational domain knowledge, some analyses include indicators of a listing containing skills related to *cognitive*, *social*, *character*, and *management* job attributes. These job attributes was based on prior work using the same data source and were constructed using the methods reported in that paper (Deming and Kahn, 2018).[15] More information on the mapping of skills to job attributes is available in Appendix B.

3.2.4 Financial and industry variables

Financial variables are accessed through the Compustat-Capital IQ database (S&P Global Market Intelligence). The workforce data provide CUSIP identifiers for public companies in the sample which can be used to connect them with financial variables in the Capital IQ data. From the Capital IQ data, I construct measures of employers' market value, employment, industry, PPE (property, plant, and equipment), and other assets. Industry variables are extracted at the three-digit NAICS (North American Industry Classification System) level. Using approaches from the literature that uses similar methods to study the market impact of investment in human and organizational complements to technology (Brynjolfsson et al., 2002), market value is computed as the value of the company's common stock plus its preferred stock plus total debt, and assets are computed as total assets minus PP&E.

4 Results

4.1 Descriptive evidence on changes to the skill content of jobs

We use these measures to evaluate changes in the labor market associated with investment in algorithmic technologies. Before testing the hypotheses discussed earlier in the paper, I provide some descriptive findings from the workforce data.

Figure 2 illustrates growth in the incidence of algorithmic technologies appearing in listings within the sample period spanning 2010 to 2016. The listings seeking these skills climbs from 2,000 per month in 2010 to approximately 16,000 per month in 2016. Each stacked bar is divided into three regions, representing the distribution of these openings over three broad occupational categories that were manually assigned to the job families already provided in the data: 1) IT jobs, 2) Data science and business intelligence jobs, and 3) other jobs. Each of these categories is responsible for one third of the total job listings in each month that require familiarity with these algorithmic technologies. Second, the shading of each stacked bar is in proportion to the domain knowledge required by

---

[15]Deming and Kahn (2018) construct these job attribute measures based on whether a listing has a skill related to the attribute. These skills, as reported in Table 1 of that paper, are: *cognitive* [problem solving, research, analytical, critical thinking, math, statistics], *social* [communication, teamwork, collaboration, negotiation, presentation], *character* [organized, detail oriented, multitasking, time management, meeting deadlines, energetic], and *management* [project management, supervisory, leadership, management (not project), mentoring, staff]. Deming and Kahn (2018) also include *writing*, *customer service*, *financial*, *computer*, and *software* job attributes in their analysis but those attribute families are not included in this analysis.

workers in that occupational family. IT jobs require the least domain knowledge and non IT and non data science jobs require the most. During this period, growth in demand for these skills is driven by growth in all three of these categories.

Figure 3 illustrates the Gini coefficient across different types of technical expertise across the entire sample of listings, including some of skills most closely related to predictive decision-making, AI, and analytics.[16] These figures indicate that algorithmic expertise is less concentrated by occupation than other forms of technical expertise, in support of Hypothesis 1. Skills related to predictive analytics, data science, and analysis are more dispersed to other occupations than database or software skills and only slightly less so then skills related to the Microsoft Office Suite, which are commonly used by workers in many occupations. This evidence from the job listing data is consistent with the claim that there are gains for employers to bundling expertise related to algorithmic tools in occupations where the appropriate subject matter expertise can already be found.

Figure 4a illustrates where different technical skills fall on a chart where the x-axis indicates the level of knowledge of the application domain required in job openings that require the skill, and the y-axis indicates the fraction of openings with the skill that are for workers in IT professions. In the bottom left quadrant of the chart are skills that are principally concentrated in traditional IT workforce jobs, such as programming and database management skills. Towards the top right quadrant of the chart are those skills that have spread to occupations outside of the IT workforce and that tend to appear in jobs that also require some domain knowledge. This top right quadrant, in addition to business applications such as enterprise systems and office applications such as spreadsheets and documents, includes most of the algorithmic skills that are the focal point of this analysis. In short, in their occupational footprint, skills related to the use of algorithmic tools share more in common with enterprise and office applications than with programming and database skills.

Figure 4b is similar to Figure 4a except that along both axes, it plots the *changes* that occurred from 2010 to 2016 along the same two dimensions – the fraction of listings for the skill that require domain expertise and the fraction of listings that are IT workforce jobs. The dotted horizontal and vertical lines on the chart indicate where there is no change in the incidence of a skill from 2010 to 2016. This chart illustrates that these skills have been moving out to new domains over the six years spanned by the panel of listings. A particularly interesting comparison is between algorithmic skills and Swift, a popular programming language for iOS applications that appeared in 2014 and shares a common level of maturity with many algorithmic skills. As Swift has matured, it has moved to non-IT jobs, but the combination of technical skill and domain expertise appears to be restricted to data analysis and classification skills such as Mahout, SAS, and Data and Cluster Analysis, which is consistent with the argument that familiarity with algorithmic tools complements domain expertise, potentially by minimizing coordination costs during the decision making process.

---

[16]The Gini coefficient in this context represents the concentration of skills. A higher (lower) Gini coefficient indicates skills are more (less) concentrated by occupation.

4.2 Tests of changes in firms' hiring patterns

I now turn towards using these measures to test the hypotheses developed earlier in the paper. First, I test for patterns of co-occurence in job listings between technology skills and domain knowledge. We investigate if domain knowledge is more likely to accompany the use of algorithmic technologies in job listings than it is to accompany the use of other business technologies in the data pipeline: namely data management and data cleaning technologies. The specific form of the logistic regression used to evaluate these correlations is the following:

$$DOM_i = ALG_i + DATA\ MANA_i + DATA\ COLL_i + \gamma_i + \epsilon_i \tag{1}$$

The unit of observation i is the job listing. The dependent variable is a binary indicator of whether a listing i requires the applicant to have domain knowledge (DOM). The measures on the right-hand side include whether the listing requires technical skills in any of the three data technology categories. It also includes a vector of control variables (γ) that includes job title, industry, and a measure of the log of the total number of skills in the job ad. In column (1) of Table 1, we can observe correlations between domain knowledge and skills related to algorithmic technologies. There is no evidence of a similar relationship between domain knowledge and the other data technologies. These empirical tests include job-title fixed effects, so these correlations indicate that algorithmic technologies are entering jobs that require domain knowledge in a way that has not occurred with other data-related technologies. This pattern of results provides support for **Hypothesis 2**.

For comparison, columns (2) through (5) of Table 1 report correlations with other job attributes where the construction of these attributes is described earlier in the paper: *social*, *character*, *cognitive*, and *management*. There are positive correlations between the use of algorithmic technologies and cognitive job attributes and weak correlations with social job attributes, and negative correlations with management or character-related job attributes.

4.3 Changes in the workforce composition of public US firms

The evidence from job listings, discussed above, documents the growing co-occurence of subject matter expertise and algorithmic tools, relative to other business technologies. Job listings have the advantage that they (i) indicate employer preferences and (ii) can be adjusted immediately, which means they serve as a leading indicator of labor market changes. However, whether these listings indicate hard requirements or the "wish list" of employers, or whether vacancies with this combination of skills get filled requires analysis of other data sources. Therefore, we now turn to data sources that document how the skill composition of firms has been changing over the last decade.

Figure 5 illustrates how technical skills are dispersed across occupations within public, US firms. In this figure, all changes in skill trend lines are depicted relative to their base rates in 2008. We can see from this figure that there is wide variation in how skills are spread between IT and non-IT occupations. "Infrastructural" technologies, like investments in networks and the Internet became

increasingly specialized as companies scaled up technical employment from the 2008 baseline. When all technical skills are visualized as a group, the levels of concentration are relatively flat throughout the sample. Skills related to algorithmic tools, however, trend markedly downwards. Relative to their 2008 baseline, these skills are spreading to other occupations, and are therefore, less concentrated in the IT workforce. This is consistent with the evidence provided in the previous section on changes in hiring requirements indicated by the job listings.

4.4   Complementarities between technology and human capital

Table 2 embeds these workforce measures, along with measures of investment in different technology classes into a regression framework that tests if the market assigns value to firms that are able to join these technology and workforce complements. We focus on market value, rather than productivity, for two reasons. First, firms take time to adjust new technologies to their production context, and most of the scholarly evidence to date suggests that firms are not yet realizing value from many of their investments in AI and analytics. Secondly, examining market value has the benefit that the value of workforce investments captured in rising market value can be interpreted as intangible assets, which are valuable to the firm and should have implications for future productivity. This argument and empirical approach is similar to that used in prior work to test whether workforce complements build new intangibles and raise the returns to broader investments in information technology (Brynjolfsson et al., 2002). The OLS regression that is tested is:

$$Log(MV)_{it} = Log(Assets)_{it} + TECH_{it} + WORK_{it} + (TECH_{it} \times WORK_{it}) + \gamma_{it} + \epsilon_{it} \qquad (2)$$

In this model, i indexes the firm and t indexes years, $TECH_{it}$ is an indicator of investment in different classes of technology measured using the stock of skills a firm has in the area, $WORK_{it}$ is how dispersed the skill is across occupations at the firm, and γ_{it} is a vector of fixed-effects including year, industry at the three-digit NAICS level, and depending on the specification, employer fixed-effects. A positive and significant coefficient estimate on the interaction term indicates that the market assigns value to firms that co-invest in both the technology and the complementary workforce characteristics.

Columns (1) through (4) in Table 2 have industry and year fixed-effects but do not include employer fixed-effects. In all columns, the coefficient on $Log(Assets)$ is significant and explains most of the variation in market value, as would be expected. Column (1) focuses on technologies related to algorithms. In this column, the coefficient estimates on algorithmic investment and the interaction term between these investments and the dispersion of algorithmic expertise are separately positive and significant. This pattern of estimates suggests that firms that invest in algorithmic decision-making and are adjusting the skill content of their workforce in the way described in this paper are building valuable intangibles (t=1.90). Notably, we do not see a similar pattern of results for column (2), which substitutes infrastructural technologies for data science and algorithm-based technologies, which is consistent with the argument that the informational complementarities that

15

arise when using algorithms drive these workforce changes. There is no reason to expect similar types of bundling of technical skill and subject matter expertise for infrastructure technologies.

Columns (5) through (8) replicate the specifications from (1) through (4) except that these regressions include firm fixed-effects. When including firm fixed-effects, any firm-level unobservables that are being observed by the technology investment variables should be absorbed by these firm effects. Indeed, when including these fixed-effects, many of the positive coefficients on technical investment disappear, which suggests that some of the positive coefficient on the investment measures reflect firm-level heterogeneity. Notably, however, the positive coefficient estimate on the interaction term for algorithmic technologies is robust to including firm fixed-effects (t=2.60). The pattern of results in this table indicates that firms that invest in algorithms unlock the value of these investments when they are able to disperse the human capital related to algorithms throughout the firm. We do not observe similar results for other technologies, where coordinating technical and subject matter expertise may not be so important. Collectively, these results provide evidence for **Hypothesis 3**.

When considered together, the analyses discussed above indicate that in the last decade, (i) employers adjusted their hiring practices in order to require subject matter experts to have expertise in algorithms, (ii) human capital related to algorithmic tools began to spread to non-technical occupations, and (iii) employers that made investments in algorithmic technologies and jointly made the appropriate workforce adjustments realized higher market values, indicating that the presence of valuable intangible assets in these firms. Together, these three pieces of corroborating evidence support the argument that a greater level of technical skill in a firm's subject matter experts is a valuable complement to its use of algorithmic tools.

5   Conclusions

Using novel, firm-level data on the workforce, this paper provides corroborating evidence from two different sources that i) expertise related to algorithms is broadly dispersed across occupations, ii) that this is due to informational complementarities that arise between technical and subject matter expertise, and iii) that the market assigns higher value to firms that make these workforce investments in tandem with investments in algorithmic tools.

A change in the skills required from workers in data-driven firms has potentially important implications. Many institutions that have not traditionally been focused on producing technical human capital, such as business schools, have observed a surge in interest in demand for courses teaching data, analytics, and AI technologies (Eisenmann, 2013; Lohr, 2017; Guetta and Griffel, 2021; Becker, 2023). This study suggests that these changes may be an appropriate response to a labor market in which successful applicants are increasingly ones who can combine subject matter expertise with the expertise required to interact with algorithmic decision-making. Within firms, these findings point to the rising importance of providing some technical training to workers in occupations outside those that fall under the umbrella of the traditional IT workforce.

These findings also have implications for the adoption of algorithmic technologies. Adoption of

these technologies, and particularly AI, has been shown to be difficult and uneven. This concentration in investment has been linked to the rising importance of superstar firms in the US and abroad (Autor et al., 2020). Our evidence suggests that the human capital stock of leading, data-driven firms may differ substantially from that of firms that are lagging in this domain. If complementary investments to the adoption of algorithmic decision-making tools require substantial workforce changes, it suggests considerable adjustment costs for firms seeking to adopt algorithmic practices. High adjustment costs imply higher levels of concentration for investment in AI and algorithms, and competitive rents for firms that have successfully installed the right human capital complements.

For managers too, these findings have important implications. Perhaps most interesting is that technical expertise has economic attributes that differentiate it from other types of human capital. Markets for technical skills are known to derive significant productivity benefits from geographic agglomeration (Saxenian, 1996; Fallick et al., 2006). Moreover, rapid skill depreciation changes the economics of professions in which technical human capital plays an important role, and this has implications for topics like gender diversity and skilled immigration that routinely attract scrutiny from policy makers. If a growing number of jobs requires expertise with technology and algorithms, it has implications for the structure of labor markets for these professions. If new technical skills spread to new occupations, it may introduce challenges normally reserved for IT workforce management to the management of these occupations.

There is likely to be considerable scope for future work in this area. By most accounts, we are at the beginning of a very large wave of investment in technologies that aid the conversion of data to decisions, and research about this phenomenon and its impact on the workforce is in its infancy. There is yet much that must be learned about how to design organizations such that humans can effectively work together with algorithms, and therefore, what workforce investments should be made to complement production contexts where algorithmic decision-making plays an important role. Firms' information capabilities will continue to evolve and algorithms will become easier to deploy as better software and tools become available, which will lower the costs of adoption and accelerate the diffusion of these technologies into new firms and jobs.

Several limitations of this study are worth noting. The data analyzed here provide limited visibility into the degree and nature of the expertise required by workers, and our analysis is limited to the narrow question of which skills are bundled into jobs. We cannot observe whether subject matter experts require deep expertise with the technology, or the interactional expertise required to engage with developers and builders of these tools. There are also many broader questions that firms are facing about how best to organize workers to complement algorithmic decisions, such as how to restructure decision pipelines and where the firm should place oversight of algorithmic decisions.

Moreover, this paper, like most related research that has come before it, takes a static view. At this early stage of adoption, there remains little evidence that the use and adoption of these technologies drives performance at the firm-level or has broad labor market consequences (Acemoglu et al., 2022). Stronger causal evidence of the impact of these workforce changes on performance will likely require allowing firms more time to adapt to this new mode of production. Additionally, new

technologies for data collection, analysis, prediction, and visualization will offer improved capabilities to generate insights. For instance, the continued evolution of "no-code" tools can lower barriers to data analysis, further altering where data science is done in the organization. As this boundary pushes forward, it will continue to change markets for these skills, raising new questions about how employers can best integrate algorithms into the workflow.

References

Acemoglu, D., Autor, D., Hazell, J. and Restrepo, P. (2022), 'Artificial intelligence and jobs: Evidence from online vacancies', *Journal of Labor Economics* **40**(S1), S293–S340.

Acemoğlu, D. and Restrepo, P. (2016), 'The race between machines and humans: Implications for growth, factor shares and jobs'.

Agrawal, A., Gans, J. and Goldfarb, A. (2018), *Prediction machines: The simple economics of artificial intelligence*, Harvard Business Press.

Agrawal, A., Gans, J. S. and Goldfarb, A. (2019), 'Artificial intelligence: the ambiguous labor market impact of automating prediction', *Journal of Economic Perspectives* **33**(2), 31–50.

Ang, S., Slaughter, S. and Yee Ng, K. (2002), 'Human capital and institutional determinants of information technology compensation: Modeling multilevel and cross-level interactions', *Management Science* **48**(11), 1427–1445.

Autor, D., Chin, C., Salomons, A. M. and Seegmiller, B. (2022), New frontiers: The origins and content of new work, 1940–2018, Technical report, National Bureau of Economic Research.

Autor, D., Dorn, D., Katz, L. F., Patterson, C. and Van Reenen, J. (2020), 'The fall of the labor share and the rise of superstar firms', *The Quarterly Journal of Economics* **135**(2), 645–709.

Autor, D., Levy, F. and Murnane, R. J. (2003), 'The skill content of recent technological change: An empirical exploration.', *Quarterly Journal of Economics* **118**(4).

Autor, D. and Salomons, A. (2018), 'Is automation labor share-displacing? productivity growth, employment, and the labor share', *Brookings Papers on Economic Activity* p. 1.

Baker, G. P. and Hubbard, T. N. (2003), 'Make versus buy in trucking: asset ownership, job design, and information', *American Economic Review* **93**(3), 551–572.

Bartel, A., Ichniowski, C. and Shaw, K. (2007), 'How does information technology affect productivity? Plant-level comparisons of product innovation, process improvement, and worker skills', *The Quarterly Journal of Economics* **122**(4), 1721–1758.

Beane, M. (2019), 'Shadow learning: Building robotic surgical skill when approved means fail', *Administrative Science Quarterly* **64**(1), 87–123.

Becker, G. S. and Murphy, K. M. (1992), 'The division of labor, coordination costs, and knowledge', *The Quarterly Journal of Economics* **107**(4), 1137–1160.

Becker, S. (2023), 'How MBA programs are embracing data in their curricula', *Fortune* .

Berman, E., Bound, J. and Griliches, Z. (1994), 'Changes in the demand for skilled labor within US manufacturing: Evidence from the annual survey of manufactures', *The Quarterly Journal of Economics* **109**(2), 367–397.

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

Black, S. E. and Lynch, L. M. (2001), 'How to compete: The impact of workplace practices and information technology on productivity', *Review of Economics and Statistics* **83**(3), 434–445.

Bloom, N., Sadun, R. and Reenen, J. V. (2012), 'Americans do IT better: US multinationals and the productivity miracle', *American Economic Review* **102**(1), 167–201.

Bresnahan, T. F., Brynjolfsson, E. and Hitt, L. M. (2002), 'Information technology, workplace organization, and the demand for skilled labor: Firm-level evidence.', *The Quarterly Journal of Economics* **117**(1).

Brynjolfsson, E., Hitt, L. M. and Yang, S. (2002), 'Intangible assets: Computers and organizational capital', *Brookings papers on economic activity* **2002**(1), 137–181.

Brynjolfsson, E., Jin, W. and McElheran, K. (2021), 'The power of prediction: Predictive analytics, workplace complements, and business performance', *Business Economics* **56**(4), 217–239.

Brynjolfsson, E., Mitchell, T. and Rock, D. (2018), What can machines learn and what does it mean for occupations and industries?, *in* 'AEA Papers and Proceedings'.

Caroli, E. and Van Reenen, J. (2001), 'Skill-biased organizational change? Evidence from a panel of British and French establishments', *Quarterly Journal of Economics* pp. 1449–1492.

Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., Wirth, R. et al. (2000), 'Crisp-dm 1.0: Step-by-step data mining guide', *SPSS inc* **9**(13), 1–73.

Choudhury, P., Starr, E. and Agarwal, R. (2020), 'Machine learning and human capital complementarities: Experimental evidence on bias mitigation', *Strategic Management Journal* **41**(8), 1381–1411.

Collins, H. (2004), 'Interactional expertise as a third kind of knowledge', *Phenomenology and the Cognitive Sciences* **3**(2), 125–143.

Colson, E. (2019), 'Why data science teams need generalists, not specialists', *Harvard Business Review* .

Cowgill, B. (2018), 'Bias and productivity in humans and algorithms: Theory and evidence from resume screening', *Columbia Business School, Columbia University* **29**.

Cowgill, B., Dell'Acqua, F., Deng, S., Hsu, D., Verma, N. and Chaintreau, A. (2020), Biased programmers? or biased data? A field experiment in operationalizing AI ethics, *in* 'Proceedings of the 21st ACM Conference on Economics and Computation', pp. 679–681.

Davenport, T. H. and Patil, D. (2012), 'Data scientist: The sexiest job of the 21st century', *Harvard business review* **90**(5), 70–76.

Deming, D. J. and Noray, K. L. (2018), 'Stem careers and technological change'.

Deming, D. and Kahn, L. B. (2018), 'Skill requirements across firms and labor markets: Evidence from job postings for professionals', *Journal of Labor Economics* **36**(S1), S337–S369.

Dessein, W. and Santos, T. (2006), 'Adaptive organizations', *Journal of Political Economy* **114**(5), 956–995.

Dixon, J., Hong, B. and Wu, L. (2021), 'The robot revolution: Managerial and employment consequences for firms', *Management Science* **67**(9), 5586–5605.

Eisenmann, T. (2013), 'Should MBAs learn to code?', *Harvard Business Review* .

Eloundou, T., Manning, S., Mishkin, P. and Rock, D. (2023), 'Gpts are gpts: An early look at the labor market impact potential of large language models', *arXiv preprint arXiv:2303.10130* .

Fallick, B., Fleischman, C. A. and Rebitzer, J. B. (2006), 'Job-hopping in Silicon Valley: Some evidence concerning the microfoundations of a high-technology cluster', *The Review of Economics and Statistics* **88**(3), 472–481.

Faraj, S., Pachidi, S. and Sayegh, K. (2018), 'Working and organizing in the age of the learning algorithm', *Information and Organization* **28**(1), 62–70.

Frank, M. R., Autor, D., Bessen, J. E., Brynjolfsson, E., Cebrian, M., Deming, D. J., Feldman, M., Groh, M., Lobo, J., Moro, E. et al. (2019), 'Toward understanding the impact of artificial intelligence on labor', *Proceedings of the National Academy of Sciences* **116**(14), 6531–6539.

Gallivan, M., Truex III, D. P. and Kvasny, L. (2002), An analysis of the changing demand patterns for information technology professionals, *in* 'Proceedings of the 2002 ACM SIGCPR conference on Computer personnel research', ACM, pp. 1–13.

Goldfarb, A., Taska, B. and Teodoridis, F. (2023), 'Could machine learning be a general purpose technology? a comparison of emerging technologies using data from online job postings', *Research Policy* **52**(1), 104653.

Gregory, T., Salomons, A. and Zierahn, U. (2022), 'Racing with or against the machine? evidence on the role of trade in europe', *Journal of the European Economic Association* **20**(2), 869–906.

Guetta, D. and Griffel, M. (2021), 'Python: The new MBA must have', *ORMS Today* .

Hart, O. and Moore, J. (2005), 'On the design of hierarchies: Coordination versus specialization', *Journal of political Economy* **113**(4), 675–702.

Hershbein, B. and Kahn, L. B. (2018), 'Do recessions accelerate routine-biased technological change? Evidence from vacancy postings', *American Economic Review* **108**(7), 1737–72.

Holmstrom, B. and Milgrom, P. (1991), 'Multitask principal-agent analyses: Incentive contracts, asset ownership, and job design', *JL Econ. & Org.* **7**, 24.

Jha, S. and Topol, E. J. (2016), 'Adapting to artificial intelligence: Radiologists and pathologists as information specialists', *Jama* **316**(22), 2353–2354.

Kleinberg, J., Lakkaraju, H., Leskovec, J., Ludwig, J. and Mullainathan, S. (2018), 'Human decisions and machine predictions', *The Quarterly Journal of Economics* **133**(1), 237–293.

Lebovitz, S., Lifshitz-Assaf, H. and Levina, N. (2022), 'To engage or not to engage with AI for critical judgments: How professionals deal with opacity when using AI for medical diagnosis', *Organization Science* **33**(1), 126–148.

Lee, C. K. and Han, H. (2008), 'Analysis of skills requirement for entry-level programmer/analysts in Fortune 500 corporations', *Journal of Information Systems Education* **19**(1), 17.

Lee, D. M., Trauth, E. M. and Farwell, D. (1995), 'Critical skills and knowledge requirements of IS professionals: A joint academic/industry investigation', *MIS quarterly* pp. 313–340.

Levina, N. and Xin, M. (2007), 'Research note-comparing IT workers' compensation across country contexts: Demographic, human capital, and institutional factors', *Information Systems Research* **18**(2), 193–210.

Levy, F. and Murnane, R. J. (1996), 'With what skills are computers a complement?', *The American Economic Review* **86**(2), 258–262.

Lindbeck, A. and Snower, D. J. (2000), 'Multitask learning and the reorganization of work: From Tayloristic to Holistic organization', *Journal of labor economics* **18**(3), 353–376.

Lohr, S. (2017), 'Where non-techies can get with the programming', *THE NEW YORK TIMES, April* **4**.

Mao, Y., Wang, D., Muller, M., Varshney, K. R., Baldini, I., Dugan, C. and Mojsilović, A. (2019), 'How data scientistswork together with domain experts in scientific collaborations: To find the right answer or to ask the right question?', *Proceedings of the ACM on Human-Computer Interaction* **3**(GROUP), 1–23.

Mendelson, H. and Pillai, R. R. (1998), 'Clockspeed and informational response: Evidence from the information technology industry', *Information Systems Research* **9**(4), 415–433.

Mithas, S. and Krishnan, M. S. (2008), 'Human capital and institutional effects in the compensation of information technology professionals in the united states', *Management Science* **54**(3), 415–428.

Modestino, A. S., Shoag, D. and Ballance, J. (2019), 'Upskilling: do employers demand greater skill when workers are plentiful?', *Review of Economics and Statistics* pp. 1–46.

Niederman, F., Ferratt, T. W. and Trauth, E. M. (2016), 'On the co-evolution of information technology and information systems personnel', *ACM SIGMIS Database: the DATABASE for Advances in Information Systems* **47**(1), 29–50.

Oostendorp, N. (2019), 'Radical change is coming to data science jobs'.

Park, S., Wang, A. Y., Kawas, B., Liao, Q. V., Piorkowski, D. and Danilevsky, M. (2021), Facilitating knowledge sharing from domain experts to data scientists for building nlp models, *in* '26th International Conference on Intelligent User Interfaces', pp. 585–596.

Pavlou, P. A. and El Sawy, O. A. (2006), 'From IT leveraging competence to competitive advantage in turbulent environments: The case of new product development', *Information systems research* **17**(3), 198–227.

Postrel, S. (2002), 'Islands of shared knowledge: Specialization and mutual understanding in problem-solving teams', *Organization science* **13**(3), 303–320.

Provost, F. and Fawcett, T. (2013), 'Data science and its relationship to big data and data-driven decision making', *Big data* **1**(1), 51–59.

Raj, M. and Seamans, R. (2018), Artificial intelligence, labor, productivity, and the need for firm-level data, *in* 'The economics of artificial intelligence: An agenda', University of Chicago Press, pp. 553–565.

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

Rosen, S. (1983), 'Specialization and human capital', *Journal of Labor Economics* **1**(1), 43–49.

Saxenian, A. (1996), *Regional advantage*, Harvard University Press.

Shim, J. P., Warkentin, M., Courtney, J. F., Power, D. J., Sharda, R. and Carlsson, C. (2002), 'Past, present, and future of decision support technology', *Decision support systems* **33**(2), 111–126.

Slaughter, S. and Ang, S. (1996), 'Employment outsourcing in information systems', *Communications of the ACM* **39**(7), 47–54.

Smith, A. (1776), *From The Wealth of Nations*.

Tambe, P., Hitt, L. M. and Brynjolfsson, E. (2012), 'The extroverted firm: How external information practices affect innovation and productivity', *Management Science* **58**(5), 843–859.

Tambe, P., Ye, X. and Cappelli, P. (2020), 'Paying to program? Engineering brand and high-tech wages', *Management Science* .

Teodoridis, F. (2017), 'Understanding team knowledge production: The interrelated roles of technology and expertise', *Management Science* **64**(8), 3625–3648.

Todd, P. A., McKeen, J. D. and Gallupe, R. B. (1995), 'The evolution of IS job skills: A content analysis of IS job advertisements from 1970 to 1990', *MIS Quarterly* pp. 1–27.

Wiesche, M., Joseph, D., Thatcher, J., Gu, B. and Krcmar, H. (2019), 'Research curation: IT workforce', *MIS Quarterly* .

Wu, L., Lou, B. and Hitt, L. (2019), 'Data analytics supports decentralized innovation', *Management Science* **65**(10), 4863–4877.

Xue, M., Cao, X., Feng, X., Gu, B. and Zhang, Y. (2022), 'Is college education less necessary with AI? Evidence from firm-level labor structure changes', *Journal of Management Information Systems* **39**(3), 865–905.

Zolas, N., Kroff, Z., Brynjolfsson, E., McElheran, K., Beede, D. N., Buffington, C., Goldschlag, N., Foster, L. and Dinlersoz, E. (2021), Advanced technologies adoption and use by us firms: Evidence from the annual business survey, Technical report, National Bureau of Economic Research.

Figure 1: A job listing requiring both domain expertise and algorithmic expertise

Figure notes: This figure is a sample listing for a job requiring familiarity with algorithmic tools (highlighted in yellow) and subject matter expertise, related in this example to marine biology (highlighted in orange). This listing and screenshot were extracted from the website www.indeed.com.

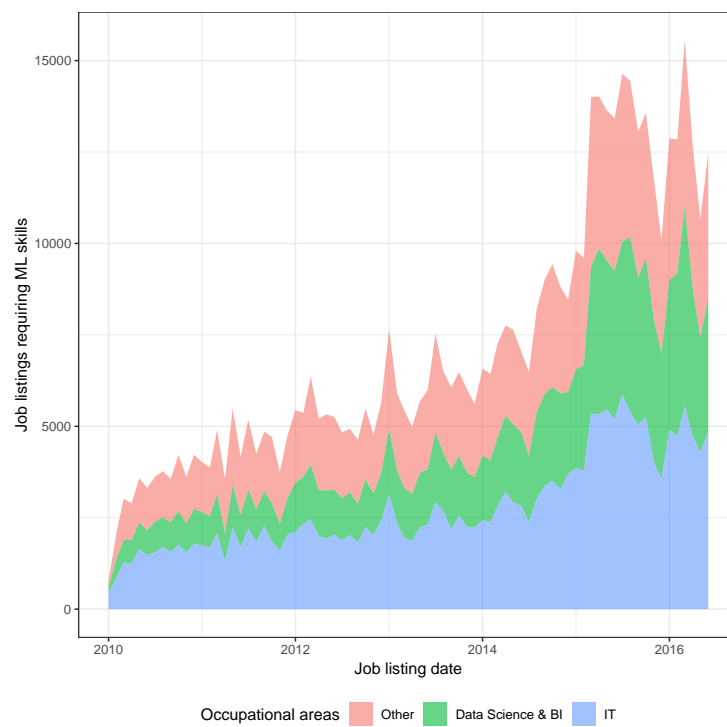Figure 2: Growth in number of job listings requiring algorithmic expertise

Figure notes: This chart illustrates growth in number of job ads requiring expertise with algorithmic tools from 2010 to 2016. The different colors indicate the occupational categories in which these listings appear: IT, Data science and business intelligence jobs, and all other occupations.

Figure 3: Distribution of different types of technical expertise in job listings
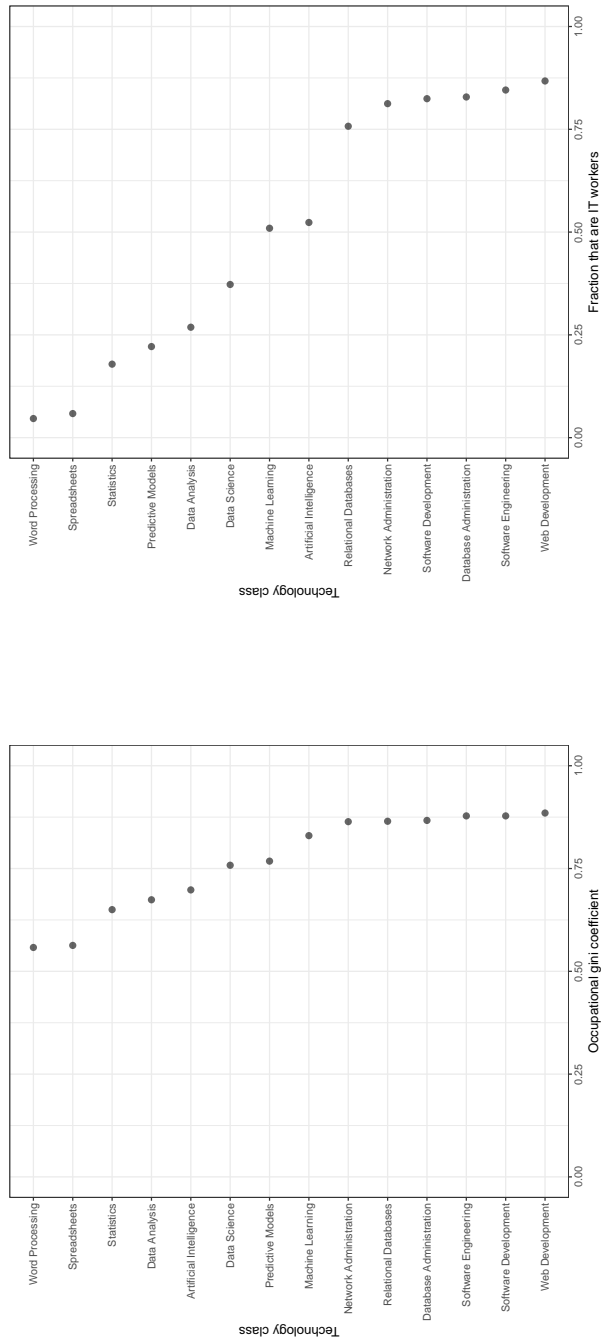
(a) Occupational concentration of skills

(b) Fraction with skill who are IT workers

Figure notes: The figure on the left illustrates the concentration of how common technology skills appear across occupational listings. Higher Gini coefficients (points towards the left) imply that the skill is more widely dispersed across occupations. The figure on the right plots the fraction of occurrences where the technical skill appears outside of an IT job. Points to the left imply that the skill is more commonly observed outside IT jobs.

Figure 4: Bundling of technical and subject matter expertise in job listings
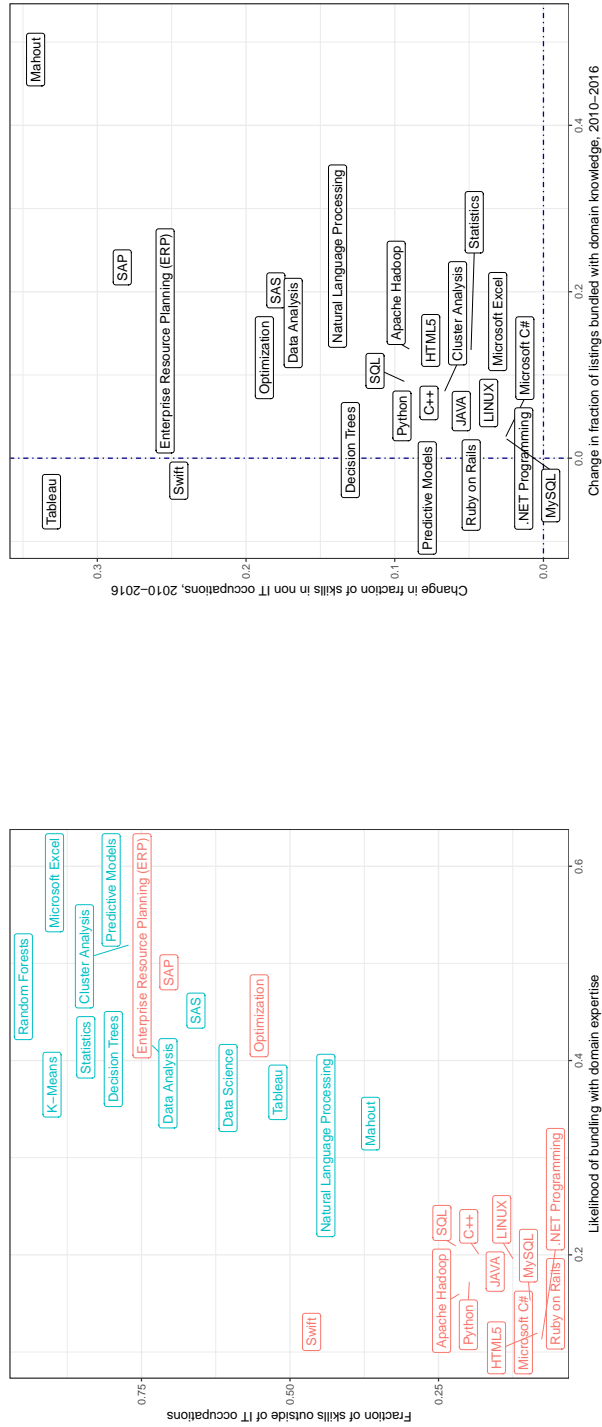
(a) Different technologies and occupational knowledge

(b) Changes over time, 2010-2016

Figure notes: The figure on the left illustrates the concentration of how common technology skills appear across occupational listings. Higher Gini coefficients (points towards the left) imply that the skill is more widely dispersed across occupations. The figure on the right plots the fraction of occurrences where the technical skill appears outside of an IT job. Points to the left imply that the skill is more commonly observed outside IT jobs.

26

Table 1: Logistic regressions of algorithmic tools on domain expertise and other job skill attributes

	Dependent variable:				
	Domain	Social	Character	Cognitive	Management
	(1)	(2)	(3)	(4)	(5)
Algorithmic tools	0.213**	0.119*	−0.391***	0.212***	−1.084***
	(0.091)	(0.067)	(0.094)	(0.066)	(0.143)
Data management	−0.414***	−0.157***	−0.229***	−0.264***	−0.680***
	(0.078)	(0.056)	(0.074)	(0.057)	(0.085)
Data cleaning & collection	0.002	0.050	0.088	−0.070*	−0.024
	(0.065)	(0.041)	(0.057)	(0.042)	(0.066)
Log(No. of Skills)	1.559***	1.552***	1.903***	1.774***	2.237***
	(0.056)	(0.036)	(0.055)	(0.038)	(0.064)
	(0.185)	(0.112)	(0.176)	(0.120)	(0.204)
Job Title FE	Yes	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes	Yes
Observations	24,387	24,387	24,387	24,387	24,387
Log Likelihood	−6,929.778	−13,096.350	−7,572.031	−12,642.710	−5,966.757
Akaike Inf. Crit.	14,271.560	26,604.700	15,556.060	25,697.420	12,345.510

Table notes: This table reports results from the logit regression $domain_i = ALG_i + MANAGE_i + COLLECT_i + \phi_i + \epsilon_i$. It estimates conditional correlations between algorithmic and skill based job attributes. The variable ALG indicates whether the job ad includes at least one skill related to an algorithmic tool and $Log(No. of skills)$ is the log of the total number of skills in the job ad. The dependent variable indicates whether or not a job listing requires knowledge of an application domain, social skills, character, cognitive skills, and people management skills, respectively. All regressions include job title and industry fixed-effects. Standard errors are shown in parentheses, with *, **, and *** denoting significance at the 10%, 5%, and 1% level, respectively.

Figure 5: Changes in the locus of technical skills in firms, 2008-2022
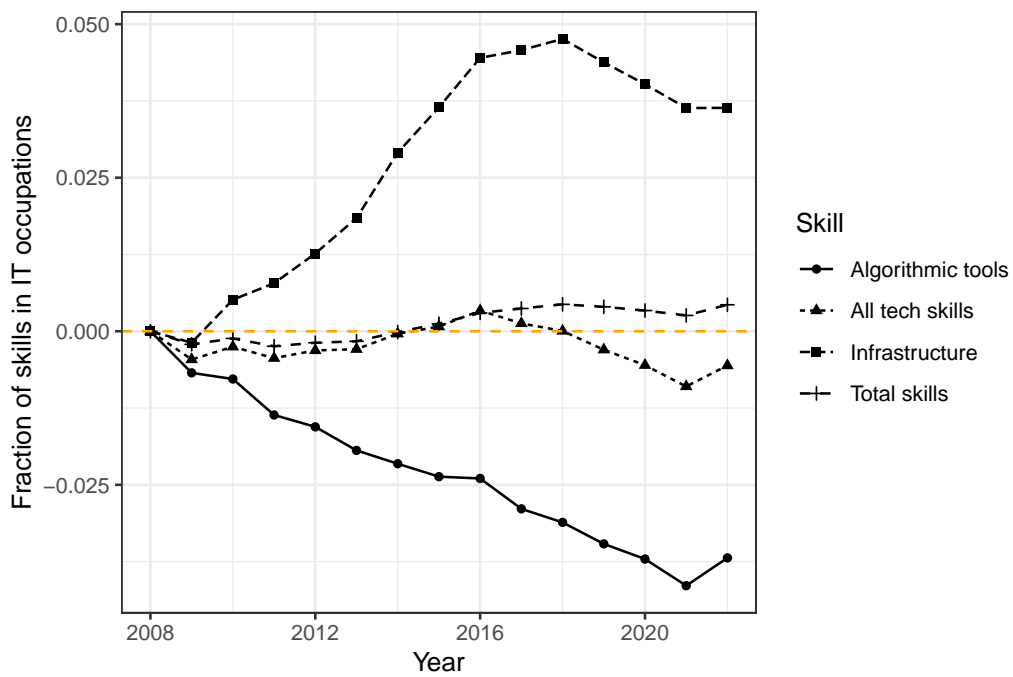

Figure notes: This figure illustrates the movement of different skill categories out of IT occupations from 2008 to 2022 in a large sample of public US firms. For each trend line, differences are reported from their 2008 base value.

Table 2: Relationship between market value, technology investment, and skill dispersion in the firm

Model:	Alg (1)	AI (2)	Network (3)	All (4)	Alg (5)	AI (6)	Network (7)	All (8)
Variables								
Dispersion	-0.1666	-0.1543	-0.0572	-0.3462	-0.1829	-0.0978	-0.2166	-0.2562
	(0.1665)	(0.1138)	(0.2029)	(0.2903)	(0.1118)	(0.0851)	(0.1481)	(0.2311)
Log(Investment) × Dispersion	0.1659*	0.1573**	0.0826	0.0988	0.1354**	0.0509	0.0642	0.0603
	(0.0896)	(0.0784)	(0.0732)	(0.0610)	(0.0531)	(0.0479)	(0.0412)	(0.0489)
Log(Investment)	0.0632***	0.0634**	0.0422*	0.0498**	0.0413	0.0735**	0.0450**	0.0400
	(0.0196)	(0.0256)	(0.0217)	(0.0196)	(0.0277)	(0.0283)	(0.0226)	(0.0274)
Log(Assets)	0.8524***	0.8474***	0.8693***	0.8594***	0.7685***	0.7667***	0.7760***	0.7686***
	(0.0255)	(0.0251)	(0.0254)	(0.0260)	(0.0314)	(0.0272)	(0.0280)	(0.0319)
Fixed-effects								
Firm					Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Fit statistics								
Observations	27,291	18,866	25,800	27,291	27,278	18,853	25,787	27,278
R²	0.81721	0.83056	0.81287	0.81614	0.94203	0.94919	0.94172	0.94198
Within R²	0.77300	0.78099	0.76630	0.77168	0.26210	0.27782	0.25911	0.26145

Clustered (naics4) standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Figure notes: This table reports regressions of how the skill composition of the workforce are related to the firm's market value. The regression model being tested is $Log(Market\ Value)_{it} = Log(Assets)_{it} + Log(Inv)_{it} + Log(Skill\ Dispersion)_{it} + \epsilon_{it}$ where observations are at the level of the firm-year. For each column, the *Investment* and *Dispersion* measures relate to the technology in the column header. Columns (1) through (4) include fixed-effects year and industry (at the NAICS 3 level). Columns (5) through (8) repeat the regressions in (1) through (4), but they add firm fixed-effects.
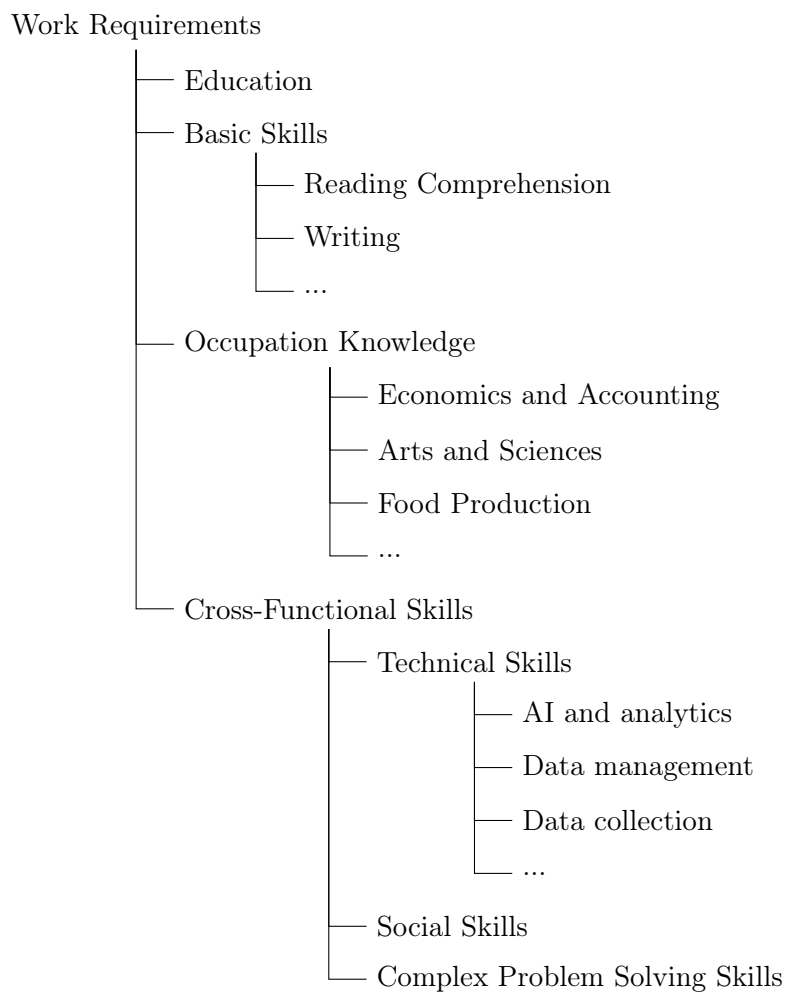
A Revelio workforce data

Pointer to Revelio site comparisons

A.1 O*NET Comparison

A.2 Firm employment comparison

B   O*NET taxonomy of worker skill requirements

Work Requirements
- Education
- Basic Skills
  - Reading Comprehension
  - Writing
  - ...
- Occupation Knowledge
  - Economics and Accounting
  - Arts and Sciences
  - Food Production
  - ...
- Cross-Functional Skills
  - Technical Skills
    - AI and analytics
    - Data management
    - Data collection
    - ...
  - Social Skills
  - Complex Problem Solving Skills

Appendix notes: This tree is an illustration of the taxonomy of "Worker Requirements" laid out in the Bureau of Labor Statistics O*NET database. This information can be found at https://www.onetcenter.org/content.html#cm2.

C   Categorization into skill groups

Major data pipeline area categories

AI, DATA SCIENCE, AND ANALYTICS	Machine Learning, Mahout, Predictive Analytics, Predictive Models, Support Vector Machines, Neural Networks, K-Means, Decision Trees, Artificial Intelligence, Predictive Modeling, Random Forests, Data Mining, Deep Learning, Language Processing, Cluster Analysis
DATA MANAGEMENT	Big Data, Apache Hadoop, NoSQL, MongoDB, Apache Hive, Splunk MapReduce, PIG, Cassandra, SOLR, Sqoop, SQL, MySQL, Structured query language, database management, database administration, data cleaning, data extraction, database querying
DATA COLLECTION	Objective C, Swift, HTML5, Javascript, iOS, CSS, Cisco, Network Engineering, Network Administration, Computer Networking, Network Support, Network Concepts and Terminology, Data Communications, Network Installation, Wireless Local Area Network (LAN), Network Management System, Network Infrastructure

Developer portfolio categories

DBA	SQL, Microsoft SQL, SAP, ERP
LAMP	Java, Linux, Apache, MySQL, PHP
FRONT END	Javascript, Ruby on Rails, PHP, HTML5, CSS

Deming and Kahn (2018) categories

COGNITIVE	Problem Solving, Research, Analytical, Critical Thinking, Math, Statistics
SOCIAL	Communication, Teamwork, Collaboration, Negotiation, Presentation
CHARACTER	Organized, Detail Oriented, Multi tasking, Time Management, Meeting Deadlines, Energetic
WRITING	Writing
CUSTOMER SERVICE	Customer, Sales, Client, Patient
PROJECT MANAGEMENT	Project Management
PEOPLE MANAGEMENT	Supervisory, Leadership, Management (not project), Mentoring, Staff
FINANCIAL	Budgeting, Accounting, Finance, Cost
COMPUTER GENERAL	Computer, Spreadsheets, Microsoft Excel, Microsoft Powerpoint
SOFTWARE	Java, SQL, Python