European Commission

JRC Technical Report

Digital skills for all?

From computer literacy to AI skills in online job advertisements

JRC Working Papers Series on Labour, Education and Technology 2022/07

Matteo Sostero and Songül Tolan

LABOUR

EDUCATION

TECHNOLOGY

Joint Research Centre

Contact Information
Name: Matteo Sostero
Address: European Commission, Joint Research Centre (Brussels, Belgium)
Email: matteo.sostero@ec.europa.eu
Tel.:

EU Science Hub
https://joint-research-centre.ec.europa.eu

JRC130291

Seville: European Commission, 2022

How to cite this report: Sostero, M., and Tolan, S., *Digital skills for all? From computer literacy to AI skills in online job advertisements* JRC Working Papers on Labour, Education and Technology 2022/07, European Commission, Seville, 2022, JRC130291.

Digital skills for all?

From computer literacy to AI skills in online job advertisements

Matteo Sostero, Songül Tolan
(European Commission – Joint Research Centre)

Abstract

The digital transition of the economy is widely expected to change the nature of work. This may happen both through creating new digital job profiles, and by digitising existing jobs. As these changes unfold, new digital skills may be needed at the workplace. We track the trends in demand for digital skills across occupations, using data from over 60 million online job advertisements in the United Kingdom over 2012-2020, the longest-running dataset of this type in Europe. Although online job advertisements tend to understate the prevalence of basic digital skills (like computer literacy or office software) compared to representative surveys, they are particularly precise in tracking skills related to emerging digital technologies. We classify over 13,000 different skills required by employers in the data into clusters, through a community-detection algorithm based on the co-occurrence of skills in job advertisements. Among the many different clusters that emerge, we identify several that relate to advanced digital skills in emerging domains, including Artificial Intelligence (AI). These advanced digital skills, despite forming distinct clusters, are evolving and slowly becoming more interconnected with the rest. We also find that digital skills are at the core of some "non-digital" domains, like the administrative and clerical cluster. Advanced digital skills also pay a notable wage premium: skills in the AI & Big Data cluster are associated with about 10.8% higher offered wages, compared to similar advertisements. For skills in the Advanced ICT cluster, the wage premium is about 15.9% and for ICT Support the premium is about 6.3%. Overall, online job advertisements provide a unique view into the emergence of distinct skill profiles, which can ultimately result in new occupations.

Keywords: Digital Transformation, Future of Work, Digital Skills, Artificial Intelligence

Joint Research Centre reference number: JRC130291

Related publications and reports: Sostero, M. and Fernández-Macías, E., *The Professional Lens: What Online Job Advertisements Can Say About Occupational Task Profiles*, JRC Working Papers on Labour, Education and Technology 2021-13, European Commission, Seville, 2021, JRC125917 `https://joint-research-centre.ec.europa.eu/publications/professional-lens-what-online-job-advertisements-can-say-about-occupational-task-profiles_en`

Acknowledgments: The authors would like to thank Enrique Fernández-Macías for his guidance and support in this project; Montserrat López-Cobo, Sofia Samoili and Riccardo Righi of the JRC AI-Watch project for sharing the list of keywords to identify artificial intelligence skills; Konstantinos Pouliakas of Cedefop for a preview into the second edition of the European Skills and Jobs Survey. We are thankful to Jiří Braňka of Cedefop and Bruno Carballa Smichowski for their comments and suggestions. We are grateful to Alfredo Peña Palma at the JRC for setting up the necessary database infrastructure and providing guidance in operating it.

Contents

Executive summary

The **digital transition** of the economy is widely expected to change the nature of work. This may happen both through **creating new job profiles**, and by **digitising existing jobs**. As these changes unfold, **new digital skills** may be needed at the workplace.

A key challenge for European skills policy is to understand exactly what digital skills are needed, and for which jobs. The **European Skills Agenda** (2020) states that "Skills [are] crucial for long-term and sustainable growth, productivity and innovation [...] and competitiveness". The fundamental dimensions and levels of digital competence required by citizens have been conceptualised by the JRC in the Digital Competence Framework for Citizens (**DigComp**), see Vuorikari et al. (2022). However, digital technology tends to evolve relatively fast, especially at the frontier of the information and communication technology (ICT) domain. Employers look for recruits who are proficient in very specific digital technology, and these demands for digital skills occur at granular and fast-changing level that foundational competence frameworks are not intended to address. As a result, the European Skills Agenda also instructs to collect Skills Intelligence: new forms of labour market data on jobs and skills – including online job advertisements (OJA).

This paper addresses the problem of identifying and classifying the many different skills required by employers, including the advanced digital skills in emerging domains like **Artificial Intelligence** (AI). We provide new evidence from a rich data source of over 60 million online job advertisements from the United Kingdom, the longest time series of its kind in Europe, over the period 2012–2020. Previous research has shown that this source of data, while not completely representative of the labour market, focuses particularly on in-demand professional occupations, and provides a particularly rich vocabulary for tools of work like software.[1] Therefore, we use this data to:

1. Measure the demand for different levels of digital skills, using definitions comparable with those of traditional survey sources like the European Skills and Jobs survey.
2. Explore how digital skills relate to each other, how they form distinct job profiles and how they relate to other (non-digital) skills profiles, to test whether digital skills are the preserve of specialised occupations, or if they are becoming mainstream.

We find that **the digital transformation is happening at different speeds, and on vastly different scales, between basic and advanced digital skills**. Applying data-driven network methods to online job advertisements allow to uncover the relation between groups of different skills, and the importance of different kind of digital skills:

- **Basic digital skills like office software are now at the core of a transversal skill set associated with the administrative and clerical domain.** Indeed, these skills are so common that often they are not always explicitly mentioned in job advertisements. As a result, the share of advertisements mentioning basic computer literacy (only 2.9% of all advertisement in 2019) or office software skills (11.3% of all advertisements) actually *understates* the share of jobs that require basic digital skills, compared to traditional survey sources like European Skills and Jobs survey, where 75% of respondents in the UK reported needing Basic/Moderate digital skills, including Office suite software in 2014.).
- **Digital skills are also found in skill clusters that primarily relate to "non-digital" domains**. In addition to the administrative and clerical domain, digital graphic design and website design are important parts of the *Marketing & Design* skill cluster.

[1] Sostero, M. and Fernández-Macías, E., *The Professional Lens: What Online Job Advertisements Can Say about Occupational Task Profiles*, Seville: European Commission, 2021, JRC125917

- **Advanced digital skills also form separate, specialist domains, which may slowly be mainstreaming**. Our analysis uncovers that, among the 13,446 distinct skills included in the data, there are at least three distinct clusters of advanced digital skills, (1) *Advanced ICT*, which includes over 700 skills, including many related to software development, (2) *ICT technical support*, with over 500 skills, and (3) *Big Data and Artificial Intelligence*, with over 350 skills. **We find tentative evidence that these specialised skills are gradually being mainstreamed:** although these clusters remain recognisably distinct among the universe of skills, over time they are gradually becoming more interconnected with the rest. This may reflect the consolidation process of specialist ICT profiles.
- **Advanced digital skills pay a wage premium**. Job advertisements that list any of the skills in the three digital skills clusters are associated with significant wage premia, even when controlling for job title, employer or industry. Job advertisements that list **skills in the AI & Big Data cluster are associated with about 10.8% higher offered wages** than job advertisements that do not list these skills, even when controlling for occupation. **For skills in the Advanced ICT cluster the wage premium is about 15.9% and for ICT Support the premium is about 6.3%**, when controlling for city, year and occupation.
- **The demand for specialist ICT job titles seems to remain relatively small and stable, though the variety of job titles makes it difficult to measure.** The share of advertisements with job titles similar to "Software developer" or "Software engineer" hovers around 1% of all advertisements, or around 56,000 in 2019. That for "IT support" is around 0.3% of advertisements, or 27,000. The recently-minted title of "Data Scientist" was used in around to 11,500 advertisements (0.12% of the total) in 2017, before declining to 0.1%.
- **The skills required for specialised ICT job profiles are not yet consolidated, and evolve relatively fast**. The skill clusters relating to advanced digital skills, while remaining clearly distinguishable, changed between 2012 and 2020, by gaining new skills, while others fell out of favour. Tellingly, the most common skills required for specific job titles like "Data Scientist" change relatively more: some skills become dominant (notably the *Python* programming language for advanced ICT and *Machine Learning* and *Data Science* as a domain), while others become less prominent. Likewise, **the skill set for professional Artificial Intelligence development is still consolidating.** The machine learning paradigm is on the rise, but still in its relative infancy, as shown by the rate at which its skills (and the tools they represent) are changing.

Overall, online job advertisements provide a unique view into the process of defining the skills required for emerging occupational profiles, which complements, but cannot replace traditional occupational surveys. **Although specialised skill profiles and emerging job titles seem to show substantial ebb and flow in their (digital) skills, better-established occupations do not show evidence of being radically transformed in short periods of time**. Future research will explore these insights further by examining the complex relation between skill clusters, job titles, and occupational categories.

This exercise is intended as a complement, not a substitute to the development of competence frameworks like DigComp, which is intended to be technology-neutral and addresses a sets of needs broader than those of employers, including those of citizens and educators. Nevertheless, the trends highlighted in this paper – including the mainstreaming of certain digital technology and the fast-changing boundaries of Artificial Intelligence – can tentatively inform the development of digital competence frameworks.

1 Introduction

Recent years have seen a steep increase in the digitisation of the economy and in the global economic importance of artificial intelligence (AI). This is changing the way we work and the skills needed to perform our work (Acemoglu and Restrepo, 2019), which is likely to cause a substantial impact on labour markets (Fernández-Macías, 2018; Goos et al., 2019; Shoham et al., 2018; Tolan et al., 2021). At the same time, there is a persistent concern among policy-makers that, over two decades after the advent of the internet, many people lack the basic digital skills needed for a complete digital transformation of the economy (European Commission, 2020). The general lack of digital skills has been posited as one of the reasons why AI has yet to boost productivity and economic growth (Brynjolfsson et al., 2021). Further, the co-existence of digital illiteracy with fast development of digital technology and AI requiring specialised skills, raises the possibility of a widening digital divide of the workforce, in Europe and beyond.

In response to these developments, policymakers have made the development of skills and competences – especially digital ones – the core of EU employment and education policy, as can be seen in the European Commission's Digital Education Action Plan.[2]. Another challenge for policymakers has been to understand the demand for digital skills and the required skills in digitally-intensive (including AI-related) occupations, which informs individuals in their career choice and addresses recurring concerns about so-called "skill mismatches".

These problems require solutions at several levels of generality. To ensure that citizens are equipped with the digital competence, the DigComp framework 2.2 identified the key components of digital competence in 5 areas: Information and data literacy; Communication and collaboration; Digital content creation; Safety; and Problem solving. (Vuorikari et al., 2022) This framework, which applies to a multitude of domains in which citizens interact with digital technology, is neutral by design about the specific software package used, focusing instead on what can people can do with technology. In the job market, by contrast, employers are often concerned with recruiting people who are skilled in the exact technologies used, often spelling out in very precise details the specific software package or programming language that they work with. When asking employers what digital skills they expect people to have, the answer can thus be very specific (at a level that frameworks like DigComp are not intended to address), but also to change relatively fast over time (whereas the competences described in DigComp have a longer horizon).

In this paper, we illustrate how the digital transformation is changing different job profiles through two measurement exercises from job advertisements. First, we assess the trends in demand across different occupations for four different levels of digital skills: (1) basic computer literacy, (2) Office Suite, (3) specialised software, and (4) artificial intelligence. Second, we use the networked structure of job advertisements to uncover clusters of related skills – some of which turn out to be digital – and illustrate changes in popularity of competing digital technologies, the professionalisation of distinct occupational profiles, and the emergence of new digital skills in the workplace, including AI.

Our analysis is based on the Nova UK dataset on online job advertisements in the United Kingdom (hereinafter, Nova UK), which is provided by Burning Glass Technologies (BGT), a private data vendor. Job advertisements can be particularly informative about AI diffusion and the digital transformation of firms as working with AI requires highly specialized human capital (Gofman and Jin, 2020). The data covers over 60 million individual job advertisements in the United Kingdom over the period January 2012 to January 2020 with about 15 million distinct job title strings, mapping to all 369 4-digit UK

[2] See `https://ec.europa.eu/education/sites/default/files/document-library-docs/deap-communication-sept2020_en.pdf`

Standard Occupation Codes (SOC-4) and nearly 700,000 distinct employers, mapping in turn to most UK Standard Industry Codes (SIC) sectors. Most job advertisements are described by a variable number of “skills”, indexed from a large dictionary comprising 13,446 distinct entries.

Previous analysis of this dataset (which is the longest-running one of its kind in Europe) has shown that, compared to representative occupational surveys, it tends to relatively over-represent white-collar professional occupations, and describe them in greater detail. In particular, it describes the tools of work, notably digital technology, in especially great detail, by tracking a large number of software and hardware products products (Sostero and Fernández-Macías, 2021). However, when taking job advertisements at face value, we find that less than 30% of them mention any digital technology at all.

Comparing these estimates with those measured in the European Skills and Jobs Survey, we find that, on the one hand, online job advertisements understate the prevalence of basic digital skills like computer literacy, or the use of office software, but the estimates for more advanced forms of digital skills, including programming (software development) are roughly comparable across sources. We attribute the relative under-representation of the lower end of digital skills in online job advertisements to a combination of the high level of digital literacy of the UK, the fact that job advertisements tend to explicitly mention skill requirements only when these cannot be safely assumed form context, and to the selection effect of the advertisements being posted online themselves. Indeed, some basic digital technologies, like office software, may increasingly be commonplace, as they are integrated into the core skills of many different occupations (see Dillender and Forsythe, 2022) and hence no require a distinctive set of skills that defines a distinct occupation.

The comparative advantage of online job advertisements, then, is not in estimating the prevalence of basic digital skills, but that they provide a rich and evolving vocabulary for in-demand digital technology and skills. By using the terminology of employers, they can provide a timely picture of the technological frontier, even in emerging fields like Artificial Intelligence. Many observers expect digital technologies to evolve relatively rapidly over time, which changes the composition of the exact skills required by occupations (specialised or not) to change accordingly. In some instances, some new professional profiles may also develop around a distinct set of emerging skills, such as “Data Scientist”.

As the frontier of the professional field of Artificial Intelligence evolves, it is challenging to develop consistent definitions to measure its extent. As a result, the scientific literature has proposed different frameworks to measure the diffusion of AI, particularly in terms of how it can affect existing jobs, using data on work tasks (Brynjolfsson et al., 2018), combined with information on AI capabilities (Felten et al., 2018; Tolan et al., 2021) or patent data (Webb, 2020; Montobbio et al., 2021). Online job vacancy data has also been used before to measure AI diffusion and the demand for AI skills across occupations and industries (Squicciarini and Nachtigall, 2021) and to analyse the relationship with employment and wages (Alekseeva et al., 2021; Acemoglu et al., 2020) and industry concentration (Babina et al., 2020). We complement this literature by identifying distinct digital and AI job profiles in a data-driven way and by analysing associated wage premia and the fluctuation of skillsets over time.

To understand the evolving landscape of occupational skills, we classify these many different skills (digital or not) from online job advertisements in a relatively small number of “skill clusters” through a data-driven approach, without pre-defining categories manually, or through any ICT taxonomy. This approach uses the implicit network structure linking over 60 million job advertisements to the 13,000 different skills that they mention, and creates “clusters” from groups of skills that are frequently mentioned together in job advertisements, using network “community detection” methods. Among the different skill clusters that emerge, we detect three that relate to broad domains of digital skills, related respectively to: (1) *Advanced ICT*, (2) *ICT Support*, and (3) *Big Data and AI*, containing between around 1,900 and 360 skills each.

While measuring the demand for individual skills informs us about the diffusion of different technologies in the economy, measuring the demand for skills in combinations of skill-sets provides a better understanding of the full spectrum of digital skills and complementary skills needed to work in a labour market that is undergoing a digital transformation. A network-based approach to measuring skills can inform us about how workers exploit skill complementarities to conduct job transitions and manage career changes (Dawson et al., 2021), how these dynamics contribute to occupational polarization (Alabdulkareem et al., 2018), and how having a set of complementary skills can yield higher earnings for freelancers on online labour markets (Anderson, 2017). Relatedly, using online job advertisements, Samek et al. (2021) perform a community detection algorithm on AI-related job advertisements to identify skills bundles needed to work in AI-related jobs. They find that the programming language "Python" and the skill "Machine Learning" are indispensable when working in AI-related jobs and they detect different profiles of AI-related jobs as well as a growing importance of socio-emotional skills in AI-related jobs. We complement their work by looking at a broader set of digital skills among which AI is one category and by applying a community detection approach to the full, unrestricted dataset. This allows us to identify different digital job profiles in a data-driven way and to identify the skills that distinguish a system administrator, say, from a software developer, or a Data Scientist.

The rest of this paper is structured as follows: Section 2 presents the dataset and provides the operational definition of "skill". Section 3 measures the implied prevalence of different levels of digital skills in online job advertisements, over time and by occupation, and compares them with a representative occupational survey. Section 4 explores the relation between skills using the network structure of online job advertisements. Section 5 tracks the most commonly required skills for selected job titles related to various ICT profiles. Section 6 discusses the findings and limitations, and provides policy messages.

2 Data

The dataset used in this paper comprises over 60 million individual job advertisements from the United Kingdom over the period 2012–2020. The dataset, called Nova UK, is provided by Burning Glass Technologies (BGT), which claims to cover the near-universe of job advertisements posted online. It is the European dataset providing the longest time series of advertisements.

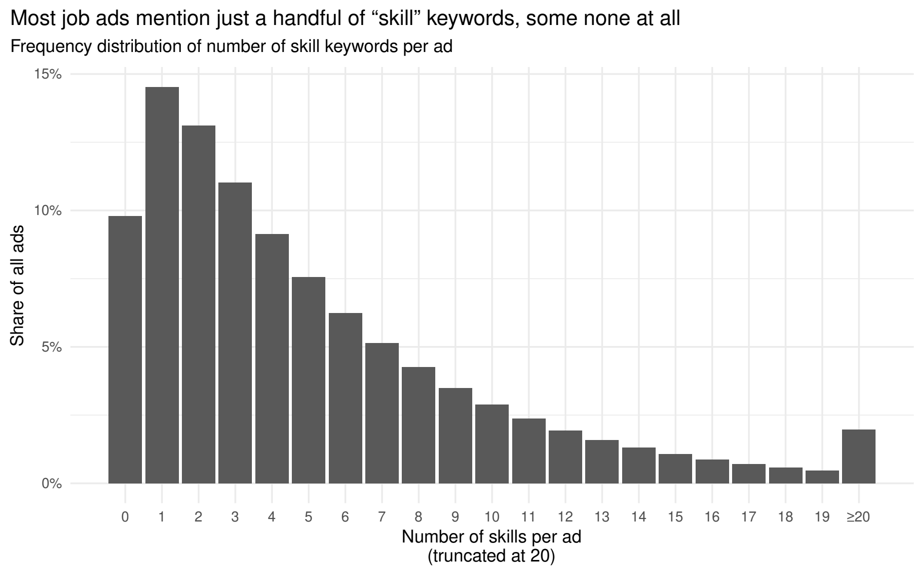
In general, job advertisements can be published on different online portals, in many formats and layouts. When this type of online content is collected, either through scraping or through application program interfaces, it comes in the form of relatively unstructured text data, which needs to be cleaned and standardised for its intended use. This is a delicate task, as it inevitably involves some value judgement on what parts of the raw data are relevant or not, whether it reliably measures a given phenomenon, and to what extent it fits into existing classifications. The Nova UK job vacancy dataset is the final product of these collection, cleaning and standardization efforts by Burning Glass.

Among the variables in the Nova UK dataset, some have been observed directly from the text of the online advertisement, like job titles, employer name, or location – while others were derived or imputed from available information during the cleaning process by Burning Glass – like occupational classification or industrial sector of activity. Table 3 in the Appendix presents an overview and assessment of the data, based on the codebook and documentation available, and from our own analysis and reconstruction. Further, we present summary statistics of the variables relevant for our analysis in Table 4 in the Appendix.

Data on concepts such as "skills", by contrast, can be more subjective to identify, potentially leading to biases. We should also note that there are different theoretical conceptualisations of the term “skill” (Attewell, 1990), and their appropriate use depends on the context (Sostero and Fernández-Macías, 2021). A simple but useful definition is: *a skill is the ability to do a specific task* (Rodrigues et al., 2021). However, the Nova UK dataset seems to use a broader definition, which further comprises broad skill categories, personal attitudes or character traits and knowledge areas such as broad industry knowledge or specific software knowledge (Sostero and Fernández-Macías, 2021). Still, the dataset allows for assessing which skills, in the words of employers, are more commonly required across different occupations.

Most of the 60 million job advertisement in the Nova UK dataset mention just a handful of distinct skills (Figure 1). The median number of skills detected in each advertisement is below five, and around 10% (or nearly six million advertisements) contain no skill information at all; a small minority of advertisements can mention over 20 skills (and over 100 in a minority of those). This heavily skewed distribution most likely reflects a combination of the limited amount of information available in the original text of the job advertisement as originally written, coupled with the technical limitations in the processing of text. Overall, this relative scarcity of information can limit our ability to describe and differentiate jobs, but we trust that the focus on the most salient skills is able to capture their principal characteristics.

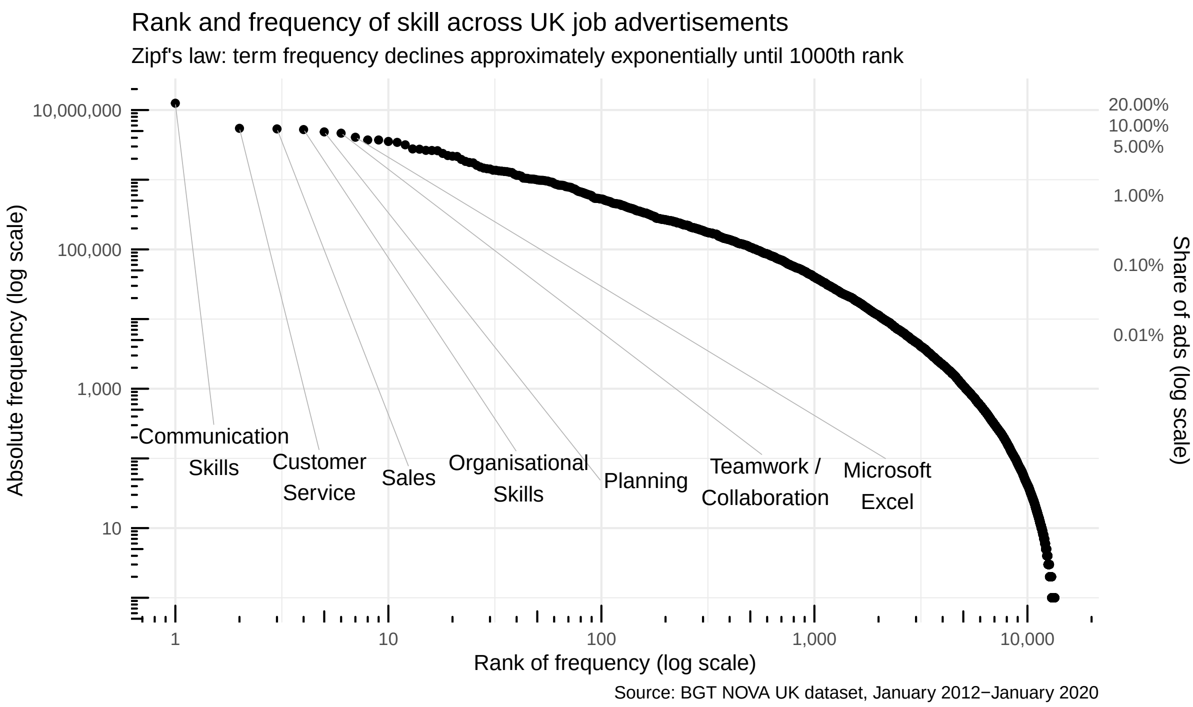
Figure 1: Number of skills per advertisement



The frequency of occurrence of skills across advertisements is revealing as well. Figure 2 plots the frequency of all 13,446 distinct skills on the vertical axis, in terms of their absolute frequency of occurrence (on a logarithmic scale), or the average relative frequency across all adds (in percentages, also on a logarithmic scale). The horizontal axis shows the rank of the skill frequency – from most common to least common – on a logarithmic scale. The figure highlights some of the most frequent skills: “Communication skills”, which appears in over 12 million (left vertical axis) or around 20% (right axis) of all advertisements. The next most common skill is “Customer service” (appearing in around 5.5 million, or 9% of all advertisements) followed by “Sales”, “Organisational Skills”, etc.

The overall pattern describes the frequency distribution of words often observed in natural languages, where frequency is inversely proportional to rank, which appears as a linear relation under a double logarithmic transformation (known as Zipf’s law, see e.g., Fagan and Gençay, 2011). Applied to the frequency of occurrence of skills across job advertisements, we can see that the relation appears to hold approximately – following a straight line – from the most frequent until the 1,000 most frequent skill. After that point, the next skills appear much less frequently, at a decreasing rate. This means that the majority of skills seldom appear in advertisements – approximately those ranking below the 1,000 most frequent, occur in at most 40,000 or 0.065% of advertisements. This “long tail” of relatively infrequent, or “niche” skills, sought by a minority of employers, are nevertheless informative about many different jobs. The ability to track a large number of relatively rare skills is an advantage of the Nova UK dataset, because it allows to track the prevalence of skills relating to emerging technologies, which necessarily begin with adoption by a minority of companies, before spreading across occupations and industries. It is precisely this type of information – notably about digital technologies at different stages of maturity and adoption – that we use to establish the prevalence of digital and artificial intelligence skills in the next section.

Figure 2: Rank and frequency of skills



3 Prevalence of digital and AI skills across occupations

This section examines what the Nova UK dataset of job advertisements can say about the digitisation of the economy. We distinguish four different levels of digital skill, based in part on existing taxonomies, and compare the prevalence of each across occupations and over time. We then contrast these figures with those of the European Skills and Jobs Survey, a representative occupational survey.

3.1 Prevalence of digital skills in online job advertisements

To measure the state of digitisation of the economy, we are interested in knowing how many job advertisements require different kinds of digital skills, understood as skills related to information and communication technology (ICT). The Nova UK dataset is arguably well suited for this task, because digital skills are especially well indexed in the dataset through the *Burning Glass Skill Taxonomy*, a classification that indexes over 1,200 distinct skills (out of over 13,000) with a "Software" flag, which mostly denotes the commercial name of a software package or programming language (such as "Microsoft Excel", "JavaScript", "Microsoft C#") but also broader software-related skills, such as "Software Development", "Software Testing", or "User-interface design". These numbers indicate a large and rich vocabulary to describe digital skills, which we attribute to a combination of three factors. The first factor is the professional dynamics in the ICT sector. Jobs in this sector were among the first to be advertised online, and continue to be well-represented among the set of online job advertisements, relative to the number of people employed in the labour market Sostero and Fernández-Macías (2021). Second, the sector implicitly defines *skill* as "proficiency with a particular software/technology". Given that these tend to be brand-names or otherwise well-codified terms with limited alternative spelling (e.g., "Microsoft Office", "Python", "Java"), often presented as items in bullet-point lists in advertisements, they are relatively easy to identify, index, and disambiguate, compared with terms of art in other trades. The third factor is that Burning Glass Technology, being a digital technology company itself, is well-versed in the trade jargon, and has commercial incentives in tracking the evolution of these skills.

Considering the great variety of digital skills, for the purposes of tracking the prevalence of digital skills across occupations, we divide them into four broad groups, describing increasing levels of digital skill:

1. *Computer Literacy*: is a skill that is recorded as such in the dataset. It is most likely not spelled out verbatim in the original ad, but the result of re-naming or standardisation of equivalent expressions by Burning Glass into its Skills Taxonomy, though the details on this post-coding are not available.

2. *Office suite*: this second group results from our re-classification, covering office software suites – Microsoft Office and its various applications, iWorks, OpenOffice, LibreOffice, and equivalent solutions.

3. *Specialised software*: all other "software" skills – i.e., referring to a name-brand software technology – not included elsewhere.

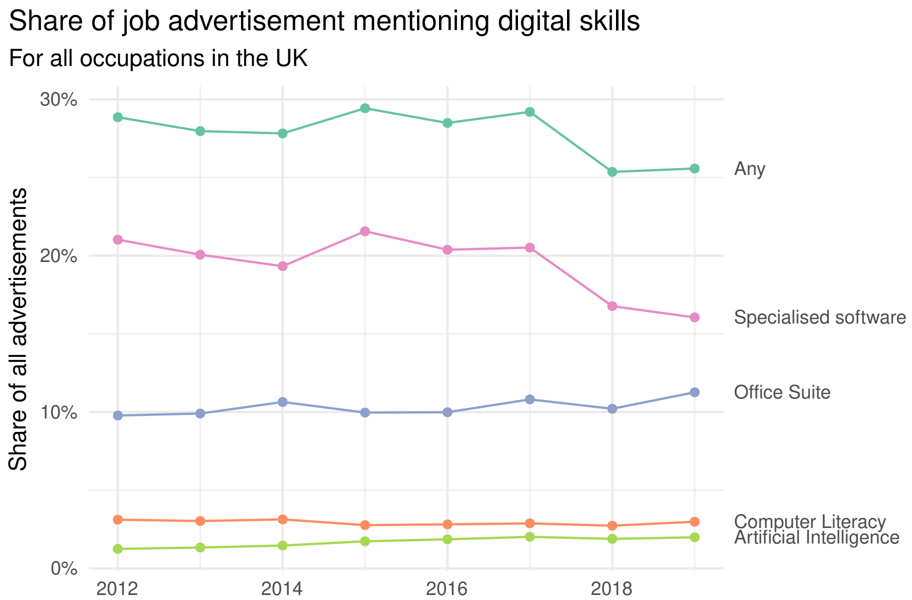
4. *Artificial Intelligence*: a set of digital skills strictly related to artificial intelligence and machine learning, as identified from different sources, including specific sections of the Burning Glass Skill Taxonomy, and a list developed by the JRC AI-Watch observatory[3] at the European Commission Joint Research Centre (Samoili et al., 2020; Righi et al., 2020), and further complemented by desk research. The full list of keywords is reported in Table 8 in the appendix.

In principle, all four categories of skills can co-occur in the same advertisement, for instance in a job requiring both "Microsoft Office" (*Office suite*) and "SAP" (*Specialised software*). In practice, however, we are unlikely to encounter advertisements that mention "Computer literacy" together with the other two categories. Although computer literacy is certainly needed to use any other software, job advertisements tend to explicitly mention only the most advanced skill requirements and omit those that can safely be assumed from context – in this case, "Microsoft Office" implies "Computer literacy", so the latter is often omitted.

With that in mind, Figure 3 show shows how the prevalence of different levels of digital skills evolved over the years. Strikingly, the share of ads mentioning any digital skill at all hovered slightly below 30% from 2012 to 2017. This relatively low prevalence is partly explained by the fact that most advertisements mention few skills of any kind at all (see Figure 1), and digital skills are a minority of all skills. We also observe a slight drop in mention of any digital skill to around 25% in 2018 and 2019.

[3] See `https://knowledge4policy.ec.europa.eu/ai-watch_en`

Figure 3: Digital skill requirements from job advertisements



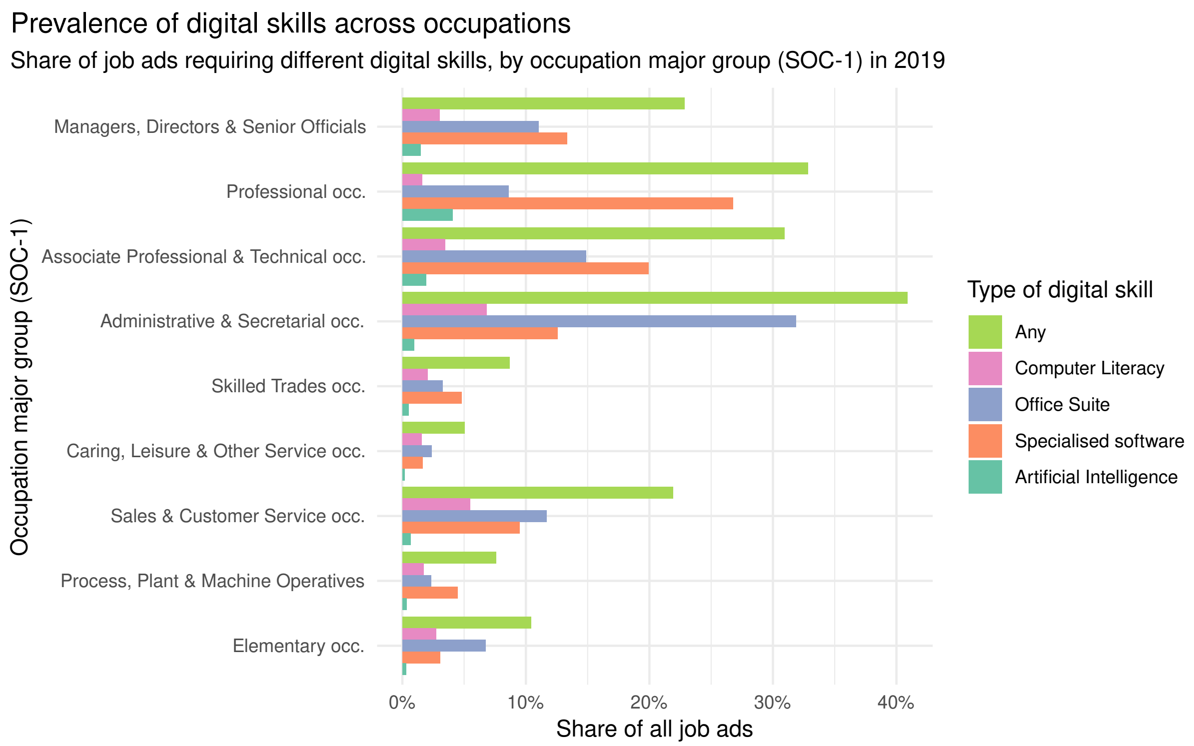
This is mostly attributable to the trend in mentions of specialised software, which declined in the last two years of the dataset, for reasons unclear. The mentions of Office Suite remained relatively constant at around 10%. Perhaps surprisingly, mentions of "computer literacy" are relatively minor, below 5%, and in slight decline over the years. Finally, the share of ads mentioning artificial intelligence skills has been growing slowly but steadily, to around 2% of UK advertisements in 2019.

To better understand the degree of penetration of digital technologies, it is useful to see whether they are widespread across all occupations, or they are the preserve of a few specialised jobs. Taking an occupational perspective allows to gauge how often more basic digital skills like "Computer Literacy" and "Office Suite" are explicitly required, and thus cannot be safely assumed, depending on the context. Figure 4 shows how all four categories of digital skills vary across occupation major groups (the first digit of the Standard Occupation Classification in the UK, which also traces an approximate gradient between so-called "high-skilled" and "low-skilled" occupations).

The most commonly required digital skills relates to Office suite applications in administrative and secretarial occupations, at just above 30% of all ads. That is also the occupational group with the highest share of job ads that mention any digital skill at all (over 40%). Intriguingly, Office software is mentioned less often in so-called "higher-skilled" occupations: around 15% of job ads for associate professional occupations, and around 10% of ads for both professionals and managerial occupations. This lower rate could be partially explained by the fact that, at least in advertisements for professional and associate professionals jobs, specialised software is relatively more common, and this could be enough to imply familiarity with the Office Suite. Nevertheless, the fact remains that mentions of digital technology of any type is lower in most so-called "low-skilled" occupations (the skilled trades; caring, leisure, and other services; process, plant and machine operatives; elementary occupations).

Figure 5 depicts the trends between January 2012 and December 2019 by occupations, where, due to strong differences in levels, we highlight the trends for AI in the lower figure. Overall, the prevalence of digital technologies in job advertisements is growing, particularly after 2018, for the occupation groups *Managers, Associate Professionals, Administrative,* and *Sales* occupations, while it remains constant for *Skilled Trades, Service* and *Elementary* occupations, and interestingly, decreasing for *Process, Plant*

Figure 4: Digital skills by occupation

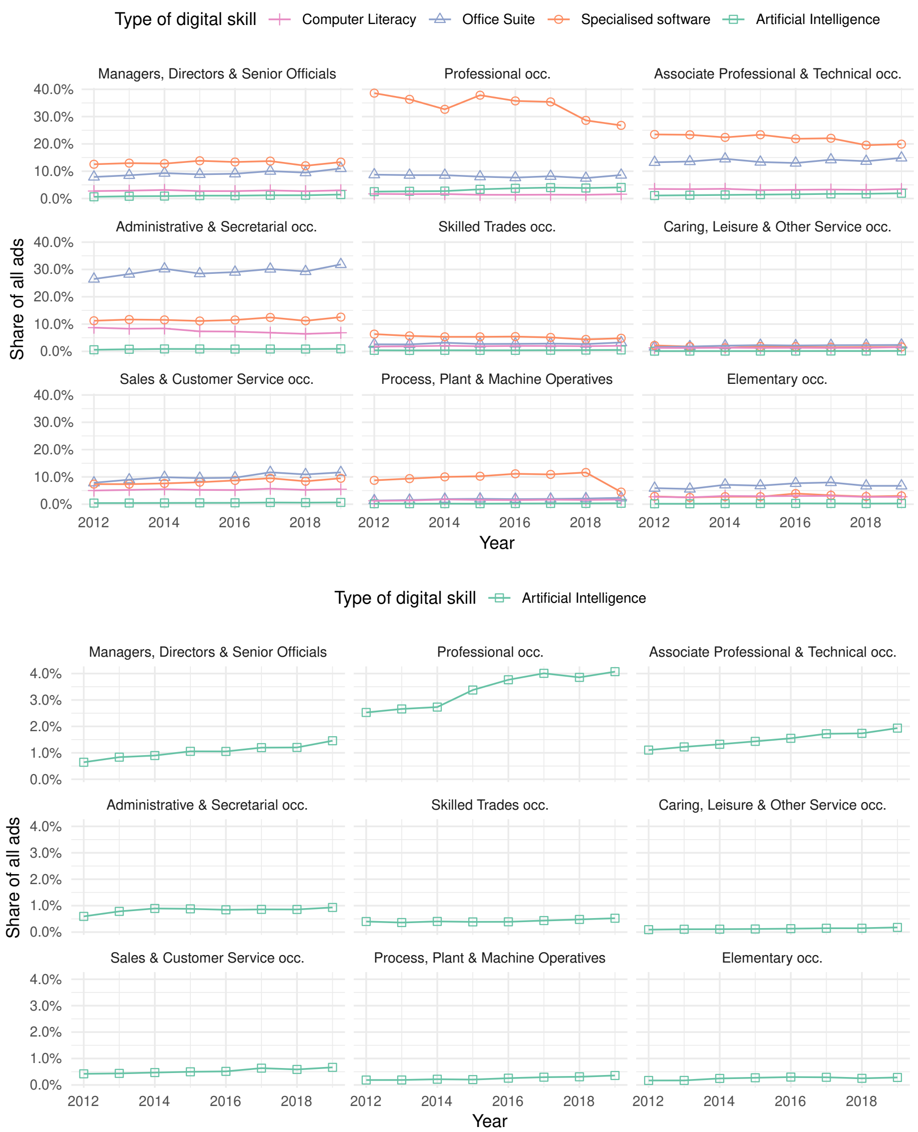


& Machine Operatives and *Professional* occupations.

Yet, compared to the large decline from 39% to 31% and a notable decrease from 24% to 20% in the demand for Specialised software in *Professional* and *Associate Professional* occupations, there is a substantial relative growth in the demand for AI skills, from 2.5% to 4% of all advertisements for *Professionals* and from 1.1% to 2% for *Associate Professionals*. We find a similarly large increase in the demand for AI skills for *Managers*, from 0.6% to 1.5%, and even a slight increase for Administrative occupations (0.6% to 1%) but no notable change for so-called “lower-skill” occupations.

Finally, the figure shows notable growth in the demand for Office Suite skills for *Administrative and Sales* occupations from 26% to 32% and from 8% to 12%, respectively and a slight increase also for *Associate Professionals* from 13% to 15%, while there are no considerable changes in the demand for Office Suite skills for *Professional* occupations or *Managers*. In summary, the changes observed in the demand for digital technologies in some high-skill and administrative and sales occupations are driven by different types of technologies. While high-skill occupations are more affected by growing demand for AI skills, administrative and sales occupations show a larger increase in the demand for Office Suite skills.

Figure 5: Trends in digital skills by occupation



Note: Trends in share of ads mentioning different digital skills, by occupation major group (SOC-1). The bottom panel zooms on the prevalence of Artificial Intelligence skills, which is much more limited than the rest.

3.2 Comparing digital skills prevalence with the European Skills and Jobs Survey

To better contextualise the figures on the prevalence of digital skills across the workforce, it is useful to compare the data from online job advertisements presented in the previous section with evidence from representative surveys, wherever possible. According to the first edition of Cedefop's European Skills and Jobs Survey [4] (Cedefop, 2018, p. 58) in 2014 over 75% of respondents in the UK reported needing basic-to-moderate digital skills in their jobs, as shown in Figure 1 below. These are defined as "using a PC, tablet or mobile device for emailing or internet browsing, using word-processing or creating documents or spreadsheets" which corresponds to the levels of digital skill we call *Computer Literacy* and *Office Suite*, respectively. Another 16% reported needing advanced digital skills, above and beyond basic-to-moderate ones, defined as "developing software, applications or programming, and using computer syntax or statistical analysis packages" (mostly *Specialised software*, in our classification, though the ESJS does distinguish artificial intelligence skills separately). Strikingly, only around 9% of respondents reported needing no ICT skills in their job. The United Kingdom also ranked among the European countries where the combined share of basic-to-moderate and advanced digital skill requirements was the highest – though the range is fairly narrow, with the lowest share of basic-to-moderate digital skills requirement in Cyprus only slightly below 60% (*ibidem*).

Table 1: Digital skills needed at work, United Kingdom in 2014

Skill level	Share (%)
No ICT Skills	9
Basic/Moderate *emailing or internet browsing, word-processing [...] spreadsheets*)	75
Advanced (*software development, [...] statistical analysis packages*)	16

Source: First edition of Cedefop's European Skills and Jobs Survey. Share of valid responses to the question: 'Which of the following best describes the highest level of information communication technology skills required for doing your job?

A similar picture emerges from a preview of the second edition of the ESJS (Table 2). This more recent data-collection exercise, currently ongoing, provides more granular information on the type of digital tools used at work, distinguishing between different kinds of software, which allows us to create levels of digital skills similar to those used in the previous section for online job advertisements. Unfortunately, this edition of the survey excludes the United Kingdom, but data from the previous edition that the UK should have somewhat above-average figures across the board for the use of digital tools.

Table 2: Digital skills needed at work, European average in 2021

Skill level	Share (%)
No ICT skills	12.66
Basic (*browsing, word, spreadsheets and presentations*)	22.69
Specialised (*specialised software and database management*)	42.80
Advanced (*programming, ICT maintenance and development*)	16.16
Not available	5.69

Note: Average for EU 27, Norway and Iceland (excludes UK). Source: Authors' own laboration based on a preview of the second edition of Cedefop's European Skills and Jobs Survey.

The measurements on the prevalence of digital skills from online job advertisement (Nova UK) are not exactly comparable with those from surveys (ESJS). In Nova UK, a single advertisement can invoke a

[4] https://www.cedefop.europa.eu/en/events-and-projects/projects/european-skills-and-jobs-esj-survey

combination of both basic skills and more advanced ones, the different categories are not exclusive; by contrast, in ESJS, the different levels of digital skills are mutually exclusive. Nevertheless, we can make a tentative comparison between the two sources, and observe that, while in the Nova UK OJA dataset as many 70%–75% of ads do not explicitly mention any digital skill at all (as deduced from Figure 5, which shows that across the years only 25–30% of ads mention at least one digital skill) whereas the equivalent figure from the 2014 edition ESJS is around 9% (Table 1). Comparing specific levels of digital skills across sources, in 2014 Nova UK only 3.1% of advertisements mentioned as skills computer literacy and 10.6% for office suite. By contrast, the first edition of the ESJS showed that 75% of respondents in the UK needed at least basic/moderate skills (defined as using a computer for browsing internet, e-mail, word-processing and spreadsheets); the European average for the second edition of ESJS is around 22.7% (see Table 2).

Interestingly, the apparent prevalence of more advanced digital skills is more consistent across sources. The first edition of the ESJS reported that around 16% of respondents in the UK worked in jobs requiring “developing software, applications or programming, and using computer syntax or statistical analysis packages” (Table 1). The corresponding European average for a similar skill level (namely, “programming, ICT maintenance and development”) is also around 16% in the second edition of the ESJS (Table 2). By comparison, the share of advertisements that mention specialised software has hovered around 20% until 2017, before also reaching around 16% in 2019 (see Figure 3). This may in part due to the somewhat more expansive definition of “Specialised software”, compared to the categories of “Advanced digital skills” used in ESJS, but we consider the figures to be roughly in line.

Overall, it seems clear that online job advertisements understate the prevalence of basic digital skills like basic computer literacy, and probably also the use of the Office suite. This likely results from the fact that the more basic skill requirements are implicit in the context where more advanced ones are mentioned. There is also likely some selection effect at play: since online job advertisements are disseminated on the internet, employers may feel it unnecessary to explicitly mention basic digital skill (such as computer literacy, which implies internet browsing) as a requirement. At the opposite end of the spectrum of digital skill, many job portals sourced in the Nova UK database include those for specialist ICT positions, where basic digital skills and Office Suite can also be assumed, but list in greater detail the type of specialised digital skills they are looking for. By contrast, the apparent prevalence of advanced digital skills, including programming (software development) and specialised software in online job advertisements is roughly comparable with the figures from survey sources. This is consistent with the relatively concise nature of job advertisement: in Nova UK, each advertisement lists relatively few individual skills, in part because employers seem to mention only the types of skills that are not implicit from context, such as novel or specialised ones.

The relative advantage of online job advertisements, then, is in the large number specialised skills resulting from the sheer number and variety of advertisements. In particular, they provide comparatively rich information concerning emerging (and in-demand) digital technologies and skills. This feature speaks to the relative benefits of each data source: representative surveys are better at establishing the baseline prevalence of well-conceptualised levels of digital proficiency, which however are bound to vary over time and risk becoming less relevant, as technology evolves. By contrast, online job advertisements do not reliably track the prevalence of more basic digital skills, but offer a rich and constantly evolving vocabulary of specialised digital skills. We make use of this timely and evolving data in the next Section, to describe how the relation between different kinds of skills – particular digital ones – has evolved over time.

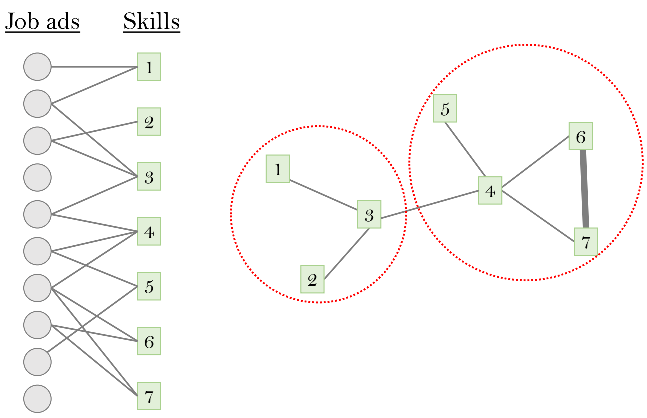
4 Detecting skills clusters from co-occurrence in job ads

This section documents the existence of specific skill clusters from the implicit network structure of online job advertisements, based on how often different skills co-occur in the same advertisement. This data-driven approach, which does not rely on manually pre-defined categories of skills, allows to distinguish which types of (digital) skills form distinct clusters – denoting specialised professional domains — and which are more integrated into a more generic skill clusters, applying to different jobs. Among the different skill clusters that emerge, we observe that several skills relate exclusively to digital technology, but some generic clerical skill clusters also include digital skills. We also estimate the wage premium associated with skills belonging to different clusters, and measure how the clusters themselves have evolved over time.

4.1 Representing job ads in skill networks

In the Nova UK dataset, each individual job advertisement can refer to multiple distinct skills, and any given skill can be referred to across multiple job ads. From this *bipartite relation* between job advertisement and skills, we can derive a relation between the skills themselves: two different skills are linked together if they co-occur, that is, they are both mentioned within the same advertisement. Mapping all the different links across all skills gives rise to a network structure that represents how all the different skills relate to one another. Within this network structure, we can potentially distinguish several distinct skill clusters, in terms of *network community structures*, as shown in Figure 6.

Figure 6: Construction of Skills Network



Job ads and skills (left) form a bipartite relation, where different skills can co-occur in different ads. This produces a network structure among skills (right), from which we can distinguish two distinct clusters (circled). Skills 3 and 4 play a central role in their clusters, partly because they have a relatively high number of links.

Formally, for every unique skill observed in the data ($s \in S$), we denote $c(s_i, s_j)$ as the link (*edge*) between skills s_i and s_j, where $c(s_i, s_j)$ is equal to 1 if s_i and s_j co-occur at least once, and 0 otherwise – i.e., two skills are linked if they are both mentioned in at least one job ad.

Moreover, the link between two skills can be more or less strong based on the frequency with witch they co-occur across job advertisements, meaning that the same couple of skills are can occur more or less often across a number of distinct job advertisements. (In Figure 6 this is represented by a thicker line linking skills 6 and 7, because those two skills co-occur in two separate ads, unlike any other pair of skills.)

The weighting of the links between the skills (or *edge weight*) is based on the notion of skill similarity in Alabdulkareem et al. (2018) and Dawson et al. (2021). For this purpose we denote n_i as the number of occurrences of skill i in all job ads and n_{ij} as the number of co-occurrences of skill i and j in the same job ad. Then, the edge weight between skills i and j (w_{ij}) is the number of their co-occurrences relative to the maximum of total occurrences of skills i and j in the job ads, i.e.:

$$w_{ij} = \frac{n_{ij}}{\mathsf{max}(n_i, n_j)} \tag{1}$$

Thus, w_{ij} is equal to 1, if s_i and s_j always co-occur, equal to 0 if they never co-occur and strictly increasing between 0 and 1 as the number of co-occurrences increases. This ensures that w_{ij} is high when both n_i and n_j are high demand as skills together, i.e., when this skill pair exhibits a high degree of *complementarity* (Alabdulkareem et al., 2018; Dawson et al., 2021).

We measure the overall importance of a skill (s_i) in terms of its *network centrality*. There are multiple measures of network centrality. Some are based on *degree*, which measures the number of direct links between the respective skill and any other skill in the network. Some are based on *betweeness* which measures the number of times the respective skill lies on the shortest path between any two other skills. For the purposes of this paper we focus on the centrality measures *degree* and *weighted degree*, because the focus of this network is on the complementarity, i.e. direct links between any two skills rather than indirect links through paths. The degree of any skill s_i is the number of other different skills it co-occurs with, across all advertisements. Formally, this is the sum of all edges involving s_i:

$$\mathsf{degree}(s_i) = \sum_{j=1}^{J} c(s_i, s_j) \tag{2}$$

The weighted degree is similarly defined, but accounts for the weight of the different edges (i.e., the number of times the same pair of skills co-occur across all advertisements):

$$\mathsf{weighted\ degree}_i = \sum_{j=1}^{J} w_{ij} \times c(s_i, s_j) \tag{3}$$

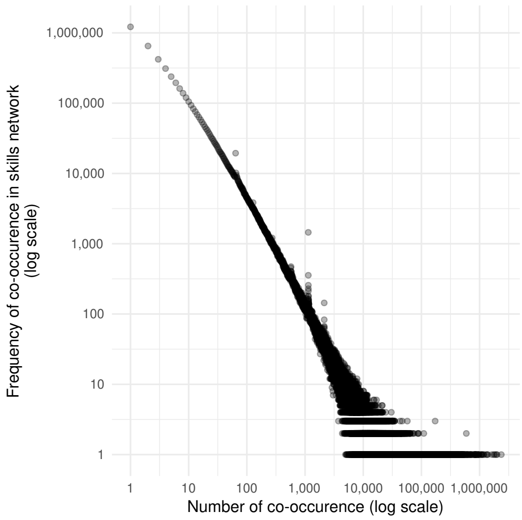
where J represents the total number of all skills in the network that are not s_i. The degree of a skill s_i increases with the total number of other skills s_j with which s_i co-occures at least once. By contrast, the weighted degree also depends on the complementarity of each connected skills pair. That is, the weighted degree of a node indicates how likely this skill can be considered as a complement to other skills. Figure 7(a) shows the distribution of edge weights across the studied skills network. As expected, the majority of skill pairs has a low degree of complementarity. However, there is also a notable share of skill pairs with higher degrees of complementarity (edge weights above 0.5), which makes them likely to end up in the same skill cluster. The full job vacancy dataset yields a network of 13,439 nodes (the number of distinct skills) and 6,021,155 edges (the number of links between pairs of distinct skills). The maximum number of co-occurences for a skill pair is 2,359,101.

To illustrate the scale and properties of the skill network, Figure 7(a), shows that the co-occurences in the network follows a power-law distribution. This means that the most common case is for a pair of skills to be mentioned together only once, something that occurs over 1 million times (top-left corner of panel (a)). Skill pairs being mentioned together two or more times happens increasingly less often, declining at an exponential rate. However, there is a relatively large set of skill pairs that co-occur between 10,000 and over 2 million times each, denoted by the streaks in the bottom-right portion of the panel. Based on this distribution, to allow for more tractable representations and more compact visualisations, we restrict the network for our analysis to a minimum of more than 150 co-occurences. This means that, to be included, a pair of skills should have been mentioned together in at least 150 different advertisements – which results in a network of 7,193 nodes and 619,383 edges.

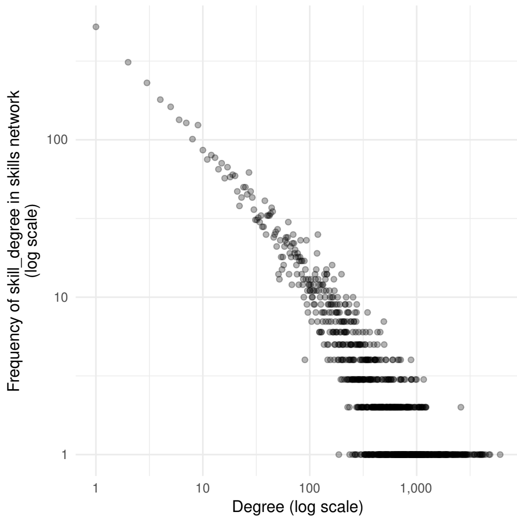
Likewise Figure 7(b) shows that the degree of each skill (i.e., the number of other skills it relates to) also follows a power-law distribution. This indicates that the skills network is indeed structured in communities around important nodes, somewhat like the stylised example in Figure 6 above. The next section detects and classifies the different communities present in the network of skills.

Figure 7: Connections in Skill Network

(a) Frequency of co-occurences n_{ij}



(b) Frequency of degrees
(number of other skills each skill relates to)



4.2 Detecting skill clusters

The network-based representations of skills across job advertisements in the previous section suggests that we may be able to identify distinct clusters of skills based on their patterns of co-occurrence across all job advertisements. The relevant concept in network topology is that of *network communities*.

Intuitively, we want do define communities of nodes in the network (i.e., skill clusters, in our case) which, as a group, are closely linked – namely, many strong pair-wise edges within them, which indicate that those skills are often mentioned together in advertisements – but are also relatively separate from the rest of the network – that is, few or weak edges with other communities in the network, which means that the skills are somewhat distinctive. This property of the network is captured by a metric called *modularity*, which measures the community structure of the network by comparing the weighted sum of edges (see Equation 1) *within* communities relative to the weighted sum of edges *between* communities. The range of modularity is $[-0.5, 1]$, where -0.5 means no community structure, i.e. a network with only random connections and 1 means a strong community structure. Networks with significant community structures typically have a modularity of ≥ 0.3 (Newman and Girvan, 2004).

We identify skill clusters using the Leiden community detection algorithm (Traag et al., 2019), which identifies communities (i.e., skill clusters) by heuristically assigning the nodes (i.e., skills) to different communities based on their close links, in at attempt to maximise modularity. The Leiden algorithm is a successor of the better-known Louvain algorithm (see e.g., Alabdulkareem et al., 2018; Anderson, 2017). Compared to Louvain, Leiden guarantees a better resolution of communities into subcommunities and stronger links within the same community (Traag et al., 2019).

Applying the Leiden community detection algorithm to the skill network shows that it has a significant community structure with a modularity of 0.59, which indicates that there are, in fact, recognisably distinct skill clusters. The Leiden algorithm divides the skills network into 91 distinct skills clusters, where the largest cluster consists of 1929 skills. Among the 91 clusters, 14 have more than 10 distinct skills, while 77 are smaller (or "niche") clusters with 10 or fewer skills (see Table 5 in the Appendix for clusters with more than 10 skills).[5]. Figure 8 shows the nine largest clusters in terms of total number of skills, their top skills (as ranked by their respective degree), and all connections with a minimum weight of 0.05. The size of nodes represents the weighted degree, the width of links represents the weight, and the different skill clusters are coloured.

First, we find that the Leiden algorithm identifies a cluster of transversal skills, such as "communication skills", "planning", "teamwork", "organisational skills" and "Problem solving". This cluster also contains the digital office suite skills "Microsoft Office" among its top skills.[6] The weighted degree of each node in this cluster, as shown by node size, highlights the relevance of these basic digital skills as part of a transversal skills set.

Furthermore, the Leiden algorithm detects three clusters relating to digital skills (among the nine largest clusters detected) with distinct skills: (1) which we describe as *Advanced ICT Skills* with "SQL", "Java" and "Software Development", (2) *ICT Support Skills* with "System administration" and "Troubleshooting", and (3) *AI & Big Data* skills with "big data", "machine learning" and "data science".[7] Since these clusters are identified based on a data-driven approach, without defining these categories in advance, their emergence suggests that each of these clusters could corresponds to job profiles with distinct skill requirements. Moreover, we find that the *Engineering & System Design* cluster, with "Matlab" as one skill and *Marketing & Design* with "Adobe Photoshop" and "Web Site Design" are additional clusters that have some digital skills among their top required skills. Overall, digital skills are a relevant requirement for many different job profiles.

The connections in the figure highlight the results of the Leiden algorithm: the density of links is higher within clusters than between clusters. Nevertheless, there are some notable differences in the connections of the different digital skills clusters compared to *Transversal Office / Administrative* cluster. The *ICT Support* cluster is directly connected with the transversal cluster, while the B*ig Data & AI* and the *Advanced ICT* clusters are only indirectly connected to the transversal skills cluster via the *Advanced ICT* and *ICT Support* clusters.

We measure the relevance of each skill cluster by the average degree of all skills in each cluster (see

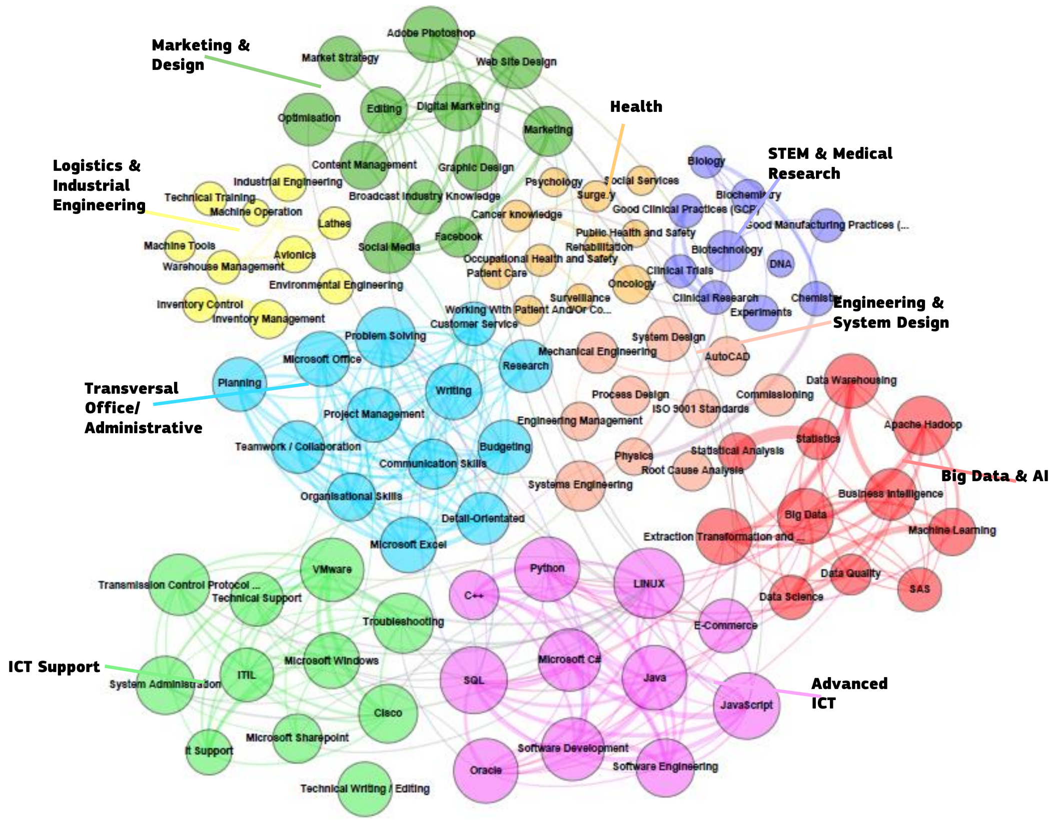
[5]The community structure and number of detected clusters depends on the chosen resolution parameter. For this paper, we tested different resolution parameters ranging from 0.5 to 1.5, where higher resolution yields more, smaller communities. A higher resolution parameter also splits the skills in the cluster of highly transversal skills, such as "Communication skills" and distributes them to several smaller skills with less centrality. The chosen graph has a resolution of 1.0.

[6]Arguably, some skills in this cluster, such as "MS Office" and "Writing" can also be interpreted as basic administrative/office skills which are not necessarily transversal to all types of occupations but only broadly to many "office workplace" type occupations. We assume that these skills occur in the "transversal" skills cluster due to biases of online job advertisements towards job types that are more likely to be advertised online, as discussed in Section 3.2.

[7]The other two relevant skills, "Business intelligence" and "Extraction, transformation and loading" are related to data science and data analysis

Table 5). For clusters with more than 100 skills this value ranges from 63.25 (STEM & Medical Research) to 267.94 (Transversal Office / Administrative). Interestingly, all three core digital skills clusters have an average degree of above 140, which is a comparatively high value. This suggests that digital skills are on average good complements to other skills, which can be non-digital or digital skills. Note that one reason for these high values of complementarity for digital skills could be a potential bias towards the identification of digital skills in the BGT dataset (see Section 3).

Figure 8: Skills Network: Top Skills in Clusters



In summary, the community-detection methodology shows different patterns of digitisation, in terms of whether digital skills are a small part of a broader cluster of mostly non-digital skills (as is the case for office software in the *Transversal Office / Administrative* cluster), a substantial but arguably not defining part of specialised skill clusters (e.g., "Adobe Photoshop" and "Web Site Design" in the *Marketing & Design* cluster), or they define their their own exclusively digital clusters (*Advanced ICT Skills*, *ICT Support Skills*, and *AI & Big Data*). The varying degrees of digitisation across skill clusters reflect the extent to which different digital technologies are mainstreamed: at one end of the spectrum, office software appears to be mainstreamed in the clerical domain, as it is just one skill among many other non-digital ones, and does not signal a high level of specialisation. At the other end, the existence of skills clusters consisting entirely of digital skills could denote some degree of professional specialisation, which could in turn be reflected in wages. The next section addresses the question of whether the skills belonging to the three exclusively digital clusters command a wage premium, compared to other kinds of skills.

4.3 Wage premia of skill clusters

To assess the wage premium for different skill clusters we analyse the variation in the hourly salary proposed in the job vacancies in relation to different skill clusters. For this purpose, we restrict the dataset to job vacancies that actually propose a salary, which reduces the dataset from roughly 61 million observations to 38 million observations (see Table 3 in the Appendix for summary statistics on missing values). To address concerns of selection bias towards those jobs – and their characteristics – that are more likely to propose a salary in their online vacancy, we present in Table 4 summary statistics for both the restricted regression sample and the full sample. We detect no notable difference in the summary statistics presented for both samples, and we see this as evidence that the subset of advertisement that report a salary offer could be representative of all job advertisement in the Nova UK dataset.

We thus estimate the following wage regression:

$$\begin{aligned} log(hourly\_wage_a) =& \beta_0 + \beta_1 \#skills_a + \beta_2 Cluster_a + \beta_3 X_a \\ &+ \mu_t + \mu_c + \mu_{soc4} + \mu_{jt} + \mu_e + \mu_{sic2} + \epsilon \end{aligned} \tag{4}$$

where $log(hourly\_wage_a)$ is the log of the proposed hourly wage in job advertisement a. In advertisements that propose both a minimum and a maximum wage, we set this variable as the minimum hourly wage. $\#skills_a$ represents the total number of skills listed in the advertisement. Given the rather long tail of the distribution of the number of skills per job advertisement (See Figure 1), with a handful of advertisements listing more than 100 skills, we truncate the variable $\#skills_a$ at the 3^{rd} quartile, i.e. at six skills. $Cluster_a$ is a vector of dummy variables that indicate the 14 clusters, which contain more than ten skills, that the Leiden algorithm detected. A variable in $Cluster_a$ is equal to 1 if the job advertisement lists a skill that appears in the corresponding cluster and 0 otherwise. X is a vector of additional control variables that were obtained from the job advertisement.[8] Finally, $\mu_t, \mu_c, \mu_{soc4}, \mu_{jt}, \mu_e$ and μ_{sic2} represent year, city, SOC 4-digit, job titles, employer, and SIC 2-digit fixed effects, respectively.

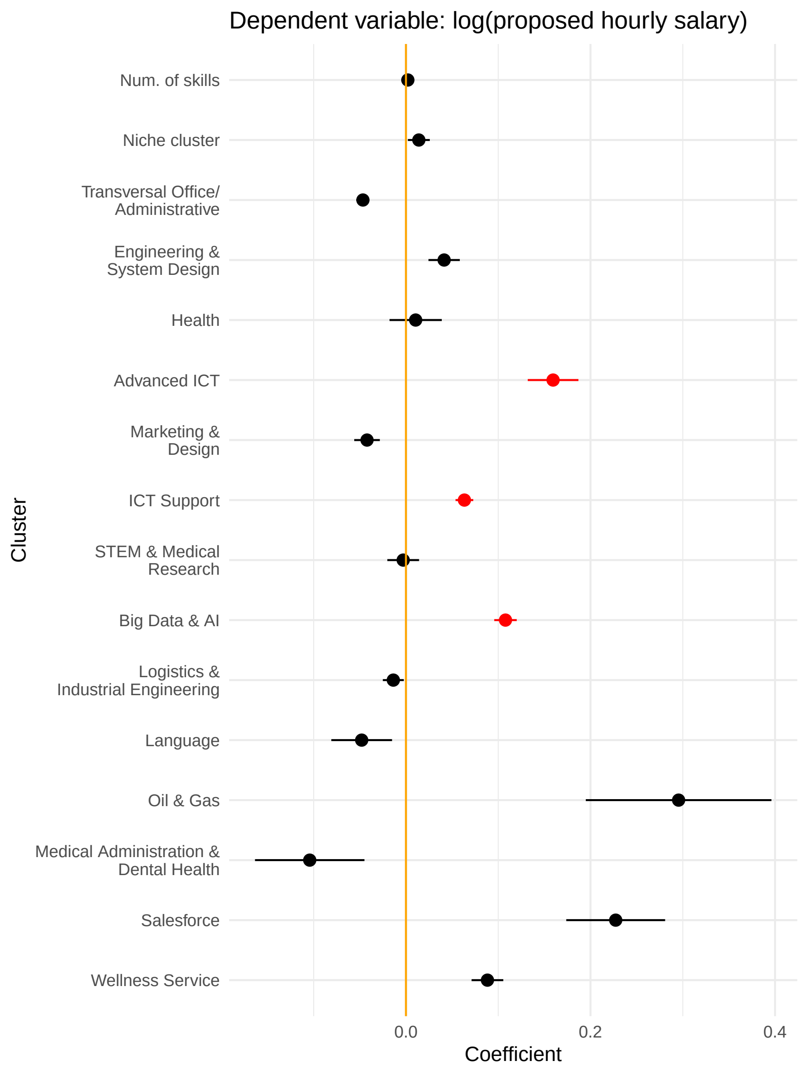
Table 9 presents the results of this wage estimation, where we sequentially add a set of fixed effects from Column (1)–(4) and (6) in order of their appearance in Equation 4. Column (5) is an exception with year, city SOC and employer fixed effects. Note first that in Column (5) the number of observations drops by about 27 million (from 37.82 million to 10.61 million) as we include employer fixed effects (as the employer is mentioned in only a third of advertisements, see Table 3 in appendix), while R^2 deteriorates from 0.78 to 0.59. Similarly, adding industry (SIC2) fixed effects on top (Column (6)) reduces observations by about 19.5 million (from 37.82 to 18.37) without a noteworthy improvement in R^2. By contrast, including occupation (SOC4) (Column (3)) and job title (Column (4)) fixed effects does not notably reduce the number of observations, while it simultaneously improves R^2 substantially, from 0.28 to 0.46 and to 0.78. This is partly a result of the fact that, in practice, over 15 million job titles strings appear only once in the Nova UK dataset, and hence they may uniquely identify as much as a quarter of observations. Even if that were not the case, controlling for job titles arguably leads to model over-specification, but we choose to include it in some specifications as a way to test the stability of our estimated coefficients to extensive controls. Given the low coverage of SIC2 and employer labels and their marginal contribution to explaining the variation in hourly wages, in our analysis we focus more on Columns (1)-(4). Additionally, since job titles may also carry a lot of implicit information about

[8] These variables are: educational degree, years of experience, part-time/full-time position, intern/temporary/permanent position.

skills, we consider Column (3) (controlling for city, year and occupation fixed effects, but not job titles) as the main specification to assess wage premia.

The results of this reference specification are shown in Figure 9, where we also highlight the coefficients of the digital and AI clusters in red.[9]

Figure 9: Wage premia of skills clusters



Notes: Coefficients and 95% confidence intervals of wage regression (Equation 4). Additional control variables not listed include: school degree, years of experience, and indicators for full-time/part-time and intern/permanent/temporary position. Fixed effects: year, city, SOC4, job title

We find that the wage premium for an additional skill listed in a job advertisement is associated with a 0.2% to 2.1% higher hourly salary proposal. However, this effect disappears when we include job title

[9] We show results for the Salesforce cluster for completeness. Salesforce is a software company that has its own programming and development skills specifications in the Nova UK dataset. The Leiden algorithm detects these skills in their own cluster. The skills in this cluster could be considered as advanced digital and AI related skills. This cluster is also associated with a notable wage premium (when not controlling for Salesfore-specific job titles). By nature of Salesforce's job advertisements (with a focus on company-specific skill terms), the Salesforce cluster is not as well connected in the network as the other 13 detected clusters.

fixed effects. That is, on average, any two job advertisements with the same job titles do not propose different hourly salaries when the number of skills listed in the advertisement differs. A reason for this could be that most job titles carry implicit information about the number of skills usually listed in advertisements for the job. Yet, we still find that listing skills from a specific cluster are associated with a wage premium that is not fully explained by job titles alone.

More specifically, we find that job advertisements that list any of the skills in the three digital skills clusters are associated with significant wage premia, even when controlling for job title, employer or industry. Job advertisements that list skills in the *AI & Big Data* cluster are associated with about 10.8% higher salary proposals than job advertisements that do not list these skills, even when controlling for occupation (Table 9, Column (3)). For skills in the *Advanced ICT* cluster the wage premium is about 15.9% and for *ICT Support* the premium is about 6.3%, when controlling for city, year and occupation. The wage premia for the digital skills cluster that we estimate when controlling for job titles (1.1%-4.2%, see Column (4), Table 9) are in line with the 5% and 3% wage premia for AI and Software skills, respectively that Alekseeva et al. (2021) find in BGT US data.[10]

Interestingly, Table 9 shows that the wage premium of the Advanced ICT cluster is in most cases greater than the premium of the Big Data & AI cluster, except when controlling for job titles. This may be because AI job titles are less self-explanatory, which requires to list their skills more explicitly, compared to job titles associated with the Advanced ICT cluster. This, in turn, could be because the Big Data & AI skills cluster has changed relatively more over time compared the Advanced ICT skills cluster (something we explore in the next subsection). If that is the case, the job titles related the Big Data & AI cluster may be associated with a less stable set of skills, compared to job titles related to job titles associated with the Advanced ICT cluster, which are better-standardised. This phenomenon is further explored in Section 5.

4.4 Evolution of skill clusters

To better see whether skill clusters are stable over time, or whether their composition changes, we divide the data for the full time period into two three-year periods, specifically years 2013-2015 and 2017-2019. We exclude the year 2016 from the data to have a clear temporal distance and avoid overlap in temporal skills changes. Note that the network approach for different periods cannot account for intertemporal trends, such that networks, clusters and skills compositions can be very different for different periods. To ensure some stability and avoid high variation due to short-term, year-specific trends, we opted for three-year periods in our analysis.[11]

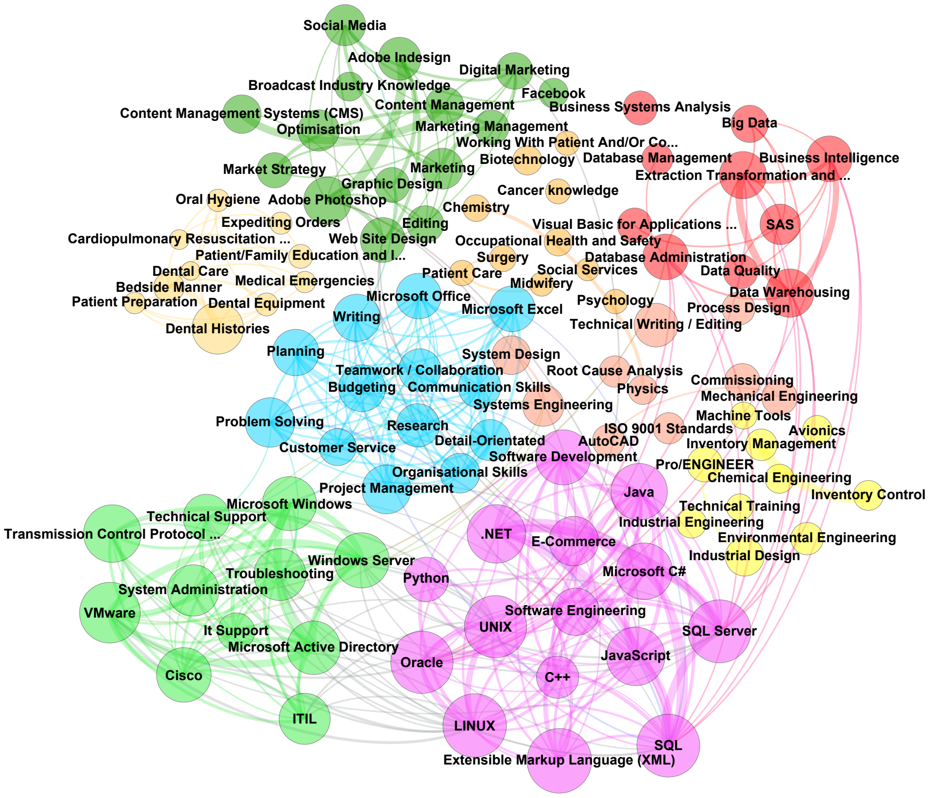
We repeat the Leiden community detection algorithm independently for each time period, which also allows the location of skills to change within their cluster. We present the top skill clusters by cluster size and the top corresponding skills by degree for each period in Figure 10. Despite the lack of inter-temporal dependence in the methodology, we still find that each sub-network breaks down into similar clusters as the full network, which speaks to the gradual, incremental evolution of skill in the workplace. The aggregate flows of skills across clusters between the two time periods is shown in Figure 12 in the Appendix.

There are some differences and similarities to note within the clusters. First, for the ICT Support and the Advanced ICT clusters there are many of the same or similar skills among the top skills between the two time periods. All top skills in the Advanced ICT cluster in period 2017-2019 are also already

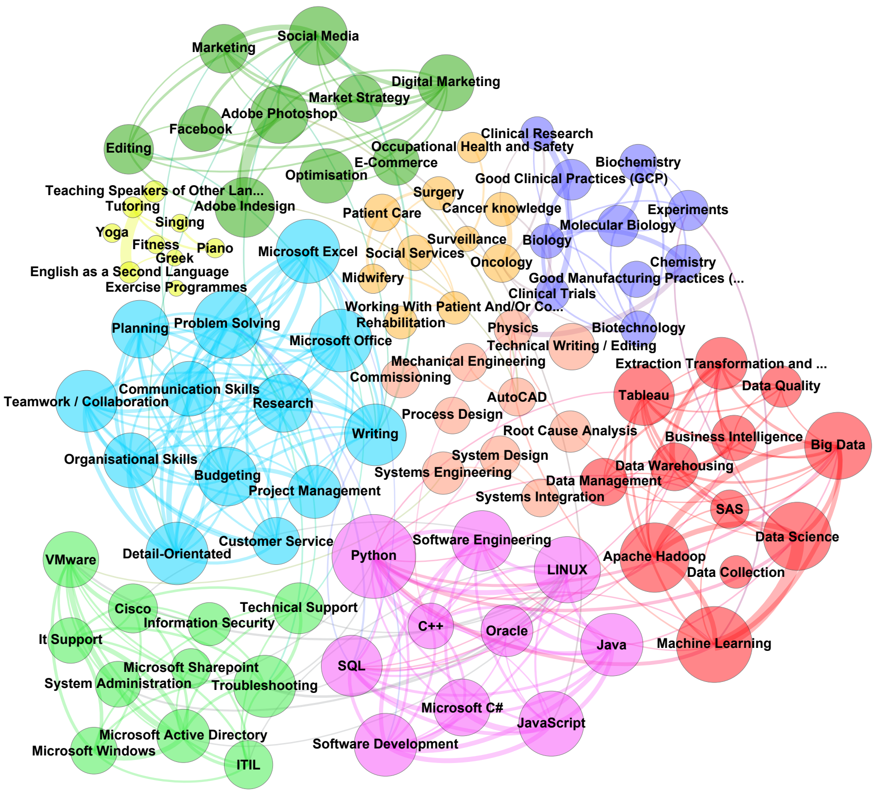
[10] Alekseeva et al. (2021) additionally control for firm fixed effects. Equivalently, when controlling for employer fixed effects, the wage premium for advanced digital skills is 7.6%, for Big Data & AI 3.6% and for ICT support it's 2.3%.

[11] These are the same time intervals as in Samek et al. (2021).

Figure 10: Networks of skills clusters for time periods 2013-2015 and 2017-2019



(a) 2013-2015



(b) 2017-2019

present in the 2013-2015 equivalent cluster. A notable change is the increase of the importance of Python as a complementary advanced digital skill as indicated by an increase in the degree of this node, reflected in its larger size in Figure 12(b). Similarly, the ICT Support cluster has mostly the same top skills in both periods. One noteworthy difference is that *information security* appears as an important skill in the 2017-2019 cluster, which is not present as an important skill in the preceding period. A possible reason for this could be an increased public awareness for cybersecurity stemming from regular cybersecurity attacks of big companies with large personal and sensitive data storage, that have received notable media attention.

Second, there are more changes among the top skills in the AI cluster. the 2013-2015 AI cluster is only assigned as such because it contains AI skills, such as "Artificial Intelligence" itself, "Machine/Deep Learning" or "Natural Language Processing" but the degree of these nodes is, in the 2013-2015 period, not high enough to be listed among the top skills. The top skills that can be assigned to AI in the 2013-2015 AI cluster derive from the list of skills that Burning Glass Technologies labelled "Big Data" ("Big Data", "Oracle") and, in a way, "Data Techniques" ("Database Administration", Data Warehousing, Database Management (see Table 8 in the Appendix). Big Data, Data techniques and SAS remain relevant skills in both periods, but the AI & Big Data cluster in the later period also contains more AI-related skills among its top skills, such as "Machine Learning", "Data Science", and "Apache Hadoop", than any skill in the AI & Big Data cluster in the 2013-2015 period. These findings indicate that there is generally a constantly high demand for different types of digital skills. However, there are differences in the temporal variation of the skills required in different digital domains. The skill clusters of *Advanced ICT* and *ICT Support* are relatively stable over time, meaning that the skills required for these jobs in 2013 are similar or same as in 2019. By contrast, the skills of the *AI & Big Data* cluster seems to be changing and simultaneously gaining importance. This is further indicated by the increase in the average degree and weighted degree of the skills in the AI & Big Data cluster from the early to the late period (see Tables 6 and 7 in the Appendix).

Third, we find that the most prominent skills in the *Transversal Office / Administrative* cluster also do not change over time. Notably in this context, the two basic digital skills, Microsoft Office and Excel, remain relevant skills in the cluster. However, as Figure 12 in the Appendix shows, over time a sizeable minority of skills from this cluster moved to other more specialised clusters. The opposite movement, from other clusters to the Transversal ones, happened to a lesser extent, but there was also a sizeable inflow of skills that were previously unclassified into the transversal cluster.

Finally, Tables 6 and 7 show that the relevance of all digital clusters (as measured by the average degree) increased over time, most notably for the Big Data & AI cluster. In more detail, the average degree for the Advanced ICT cluster increased from 160.60 to 162.51, for the ICT Support cluster from 130.01 to 131.04 and for the Big Data & AI cluster from 96.22 to 113.82.

By showing that meaningful clusters of related skills exist, that they remain relatively constant over time, the community-detection methodology is effective in identifying the degree of specialisation of digital skills, and how these evolve over time. As technology matures, it sometimes expands its reach to become embedded into other domains, and accessible to more occupations. Eventually, it may no longer be thought of as a "technology" at all. To better understand this process of mainstreaming in the case of digital technology, the next section approaches it from the perspective of some of the occupations concerned, to see how their skill requirements changed over time – or did not.

5 Changing skill profiles of selected job titles

This section explores how specific job titles – related to domains with different degrees of digitisation – evolved over time, in terms of the relative composition of the skills demanded of them. By holding job titles constant, we can see a difference in the rate at which the skills of certain jobs can change. From the clustering exercise in the previous section, we identified at least three domains that require distinct sets of exclusively digital skills: *ICT Support*, *Advanced ICT*, and *Big Data and Artificial Intelligence* (which includes Data Science and Machine Learning). The exercise also showed that the cluster of transversal office and administrative skills includes several digital technologies as well, such as office suite software. Finally, the analysis of skill clusters over time suggests that those relating to Big Data and Artificial Intelligence may be changing relatively fast, compared to the rest.

The job-title approach builds on the findings of the skill clusters of the previous section, and complements it in important ways. The network community-detection approach revolved entirely around skills, and showed the structure of the relations between them changed over time. However, the cluster approach is silent on the relative change in relative prominence of skills *within* clusters. It is conceivable that some clusters may remain stable in terms of the type of skills they require, but that the relative frequency of those skills changes over time. Moreover, some of these changes may go hand in hand with the emergence of new job titles, or the changing skill composition of existing job titles. In general, there is a complex dynamic relation between emerging skills, new job titles, and standardised occupational categories, which is beyond the scope of this paper. In this section, we focus only on *existing* job titles, more or less recent and with different degrees of digital intensity, to see how the prevalence of different digital skills for specific job titles changed over time.

We thus look at the evolution over time of the most commonly required skills in job advertisements with selected job titles to understand how they are evolving over time in terms of their digital skill requirement. For the *Advanced ICT* cluster we track the job title "Software Developer" and "Software Engineer". For *IT support* and IT support we as well as "IT Support" "Technical Support" and "Systems admin" respectively. For the *Big Data and Artificial Intelligence* cluster, we consider the job titles of "Data Scientist" or including "Machine Learning". Finally, we consider the job titles "Administrative Support" and "Secretary" to capture jobs in the transversal office and administrative domain.[12]

These collections of job titles have different rates of prevalence, as show in Figure 13 in the Appendix. The "Secretary / Administrative" corresponded to around 4.7% of all UK advertisement in 2019, or over 325,000. The specialised ICT profiles are comparatively less common: "Software Developer / Engineer" amounts to 0.8% of advertisements (or over 56,000 in 2019); "IT/Technical Support" and "Systems admin" to around 0.4% of advertisements (or 27,400). The job title of "Data Scientist", though increasingly prominent in the popular imagination, accounted for 0.1% of all advertisement (a little over 7,000 in 2019).

Figure 11 shows the evolution of the most commonly required skills in job advertisements over the period 2012–2019 for each selected job profile. Figure 11(a) paints a picture of an occupation in flux: over time, some technologies have gained prominence (e.g., *Python*, a programming language), while others have declined (e.g., *R*, another programming language, or Apache *PIG*, a "middleware" data interface). In 2012-2013, many of these (often competing) technologies were roughly equally common, but over time the skills expected of a data scientist have become relatively more polarised. In particular, *Machine Learning*, the foundation of modern Artificial Intelligence, is emerging together with Python, a popular programming language for Machine Learning applications. Nevertheless, the generic

[12] In all cases, we allowed for case-insensitive, partial matching, so that for instance "Systems admin" would match "Systems Administrator for Fintech Startup".

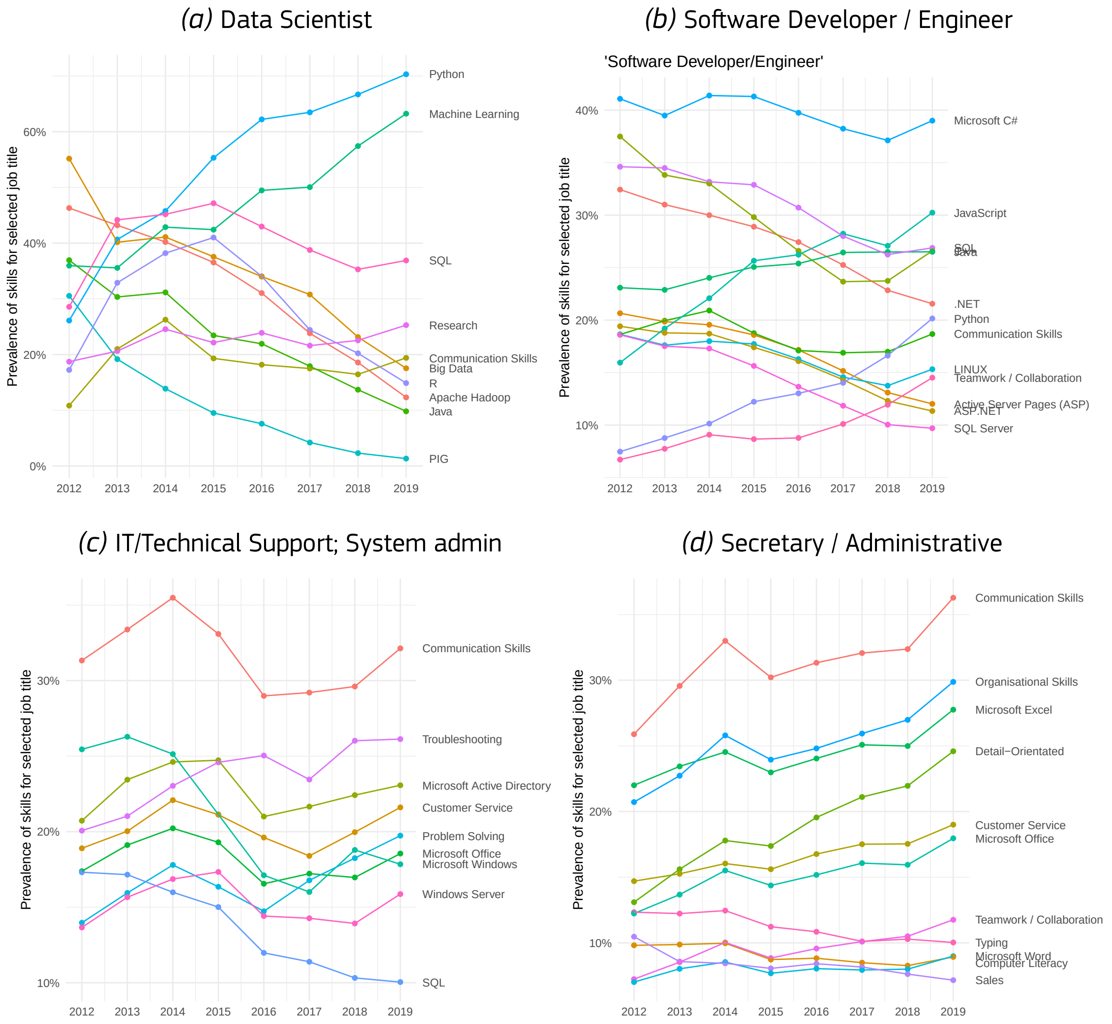
skills *Research* and *Communication Skills* are constantly listed in about 20% of all job advertisements. Some smaller and slower changes can also be found in Figure 11(b) for the software development job profile. We observe a slight decline in *SQL* (a language to manage large relational databases) and *.NET* (a framework mostly used to develop with C#), while the demand for JavaScript (a language that makes webpages interactive) is naturally increasing as websites are increasingly built to personalise user experience and increase user engagement. The skills for the IT support and administrative support job profiles remain comparatively stable over time. Not surprisingly, a constantly high share of job advertisements for the It support job profile demand generic skills such as *Communication Skills* or *Problem Solving*. Figure 11(d) also shows that the importance of the digital office-suit skills *Microsoft Excel* and *Microsoft Word* for administrative support jobs is increasing. Thus, while we observe notable changes in the distribution of skills for the AI job profile, other job profiles exhibit much smaller to no changes. We interpret this as an indication that the occupation of "Data Scientist/Machine Learning Specialist" – and with it, the field of applied Artificial Intelligence – is still undergoing a process of consolidation, which involves standardising tools and skills. Online job advertisements provide us with a rare insight into the process of defining the skills required for emerging occupational profile.

Across these few examples, we notice a tentative pattern: job profiles that have existed for longer tend to mention more often skills with similar names as the job title itself (call them "eponymous skills"), compared to newer and emerging job profiles. We interpret this pattern as evidence of the social establishment of a notion of "competence" associated with more-established profiles. Putting this findings in the context of the analytical framework of Rodrigues et al. (2021) allows to interpret the process of consolidation and professionalisation of skills clusters and job titles. The "skills" found in job advertisements actually correspond to different levels of granularity, which is related to their consolidation over time. New skills related to new technologies or processes will be mentioned more in detail, often explicitly linking to the specific technology or method used. On the contrary, skills that have existed for longer and that are widely used will require less detail, and often they will be implied as part of a more general skill cluster ("project management") or even job title ("Chef", "Driver").

This reflects the underlying processes by which new skills emerge and are bundled into skills clusters, and eventually, job titles. When new technologies or new processes emerge in economic organisations, involving new tasks and requiring new skills, they have to be either bundled into pre-existing skills clusters or job titles, or into new ones, depending on the similarity or complementarity between those new skills and preexisting skill clusters and job titles. Because this is a process that takes place in relatively uncoordinated ways in economic organisations, initially there will be more diversity in the definition and clustering of those new skills, and thus job advertisements will need to be much more specific and detailed when describing them. But as the forces of habit, professionalisation and market exchange tend to standardise and consolidate the new skills into distinct skills clusters and job titles, it will be possible to refer to those skills by just referring to the broader skills clusters in which they are typically embedded – or even more broadly to the associated job title.

With this framework in mind, it is telling to observe how the relatively more established job profiles in our example, such as "Administrative support" or "IT/Technical Support" – which also have comparatively more stable skill profiles – also have relatively more (and more prominent) synonymous and eponymous skills, compared to more recent job profiles like "Data science, Machine Learning". Although these are just a few examples, we are tempted to interpret the presence and prominence of eponymous skills as an indicator of the dynamics of professionalisation and the establishment of an archetype of professional competence. Over time, better-established job titles, with stable skill profiles, tend to crystallise around an idealised type of competence, where the job title itself becomes a by-word for a number of tacit skills.

Figure 11: Changing skillsets in different job profiles



Note: for more meaningful visualisation, the plots exclude the skills that are similar or identical to the job title itself, which tend to be among the most common. Thus panel (a) “Data Scientist” excludes the literal skill *Data Science*; panel (b) “Software Developer / Engineer” excludes *Software Development* and *Software Engineering*; panel (c) excludes *IT Support* and *Technical Support*; panel (d) excludes *Secretarial Skills* and *Administrative Support*

6 Conclusions

This paper studied the demand for digital skills along the spectrum of digital skills, from basic computer literacy to the specialised domain of Artificial Intelligence, using online job advertisement data from the United Kingdom for the period 2012–2020. This dataset provides the longest time series of job advertisements in Europe, which allows to observe some change in the panorama of skills over time. It also offers a particularly rich vocabulary of over 13,000 skills, which are expressed with the terminology of employers, and are therefore well-suited to represent the emergence of new skills, particularly those related to digital technology. The emphasis on technology is partly the result of online job advertisements over-representing white-collar professional occupations relative to the rest, and describing their skill in greater detail, with a particular emphasis on the tools of work.

At first glance, the online job advertisements suggest that only a minority of job advertisements, across all occupations, explicitly mention any kind of digital skill. This contrasts with estimates from representative surveys such as the European Skills and Jobs Survey, which show that a majority of jobs require at least basic digital skills. The demand for basic digital skills is highest among administrative and secretarial occupations, professionals, associate professionals and technical occupations (between 25 % and 32% of all job ads in these occupational categories), where the administrative occupations have a higher demand for office software. The demand for all other specialised software is highest among job ads for professional occupations. In addition, there is also a notable share (10%-14%) of digital skills among the skills required in management and sales occupations job ads. The demand for the most basic "Computer Literacy" type of skills is overall relatively low, probably not because of a factually low demand but because they are often assumed when more advanced skills are required.

The comparative advantage of online job advertisements in describing advanced and specialised digital skills allows to analyse the relation between digital skills and the rest. We represent the skills in terms of a network, based on how often they occur together across job advertisements. We then use the data-driven Leiden community detection algorithm to distinguish clusters of related skills, by focusing exclusively on the relation between skills themselves, without pre-defined categories. This approach allows us to understand which digital technologies constitute distinctive skill clusters, and which ones are embedded in otherwise non-digital domains. In particular, the approach identifies *Advanced ICT*, *ICT Support*, and *Big Data and AI* as an independent clusters with highly complementary skills. We also cluster of *Transversal Office / Administrative* skills, such as "Communication Skills", "Planning" and "Teamwork", that is are well-connected (and hence highly complementary) to most other skill clusters we detected. Among the digital skill clusters, the one related to *Big Data and AI* appears less connected with the generic clerical cluster, which can be taken as evidence that AI is still a relatively esoteric pursuit, even among digitally-intensive occupations.

We also find that digital skills are prominent in skill clusters that primarily relate to "non-digital" domains. For example, skills related to Microsoft Office suite fall in a cluster of *Transversal Office / Administrative* skills, while graphic design and website design are important for a *Marketing & Design* job profile. While it appears that the AI & Big Data cluster is not (yet) as connected as the other two digital skills clusters – namely, it is not a strong complement to complementing other non-AI skills – we find that these connections are increasing over time. Moreover, we find that skills in any of the digital skills clusters yield considerable wage premia: job advertisements mentioning a skill in the Advanced ICT cluster propose on average 15.9% higher minimum salaries than other job advertisements in the same occupational group Similarly, the wage premium is 10.8% for skills in the AI & Big Data cluster in the same occupational group and 6.3 % in the ICT Support cluster for the same occupational group. Even when controlling for job titles, we still find notable positive wage premia for all three digital clusters (1.1% to 4.2%), suggesting that digital skills even make a difference in terms of wage proposals for jobs with the same title. However, the analysis of skill clusters and job profiles over time suggests that the cluster of skills in demand for the AI & Big Data job profile is highly in flux, meaning that its composition changed substantially between 2012 and 2019. By contrast, for most skills in the other two digital skills clusters remains relatively stable.

The detection of distinct (digital) skill clusters suggests that they may corresponds to separate job profiles with different skill requirements. Indeed, the findings above are also confirmed by an analysis of specific job titles, and how their skill composition has evolved over time. In particular, the increasing relevance of particular programming languages or methods for Data Science as such shows that this new occupation is less mature than others.

The combination of high wage premia for digital occupation higher-than-average variability in their skills may also be evidence that higher wages are necessary to attract employees in fast-changing domains, and hence that the very mutability of the skills in question commands a premium.

These results have to be interpreted in light of some limitations. The first is the representativeness of the data on online job advertisements, which necessarily presents the perspective of employers, and comes from the United Kingdom, which means it may not be representative of the European labour market. Although the dataset has the longest time series of its kind in Europe, reaching back to 2012, that remains a relatively short period of time to observe systemic occupational and skill transitions.

Overall, our work has implication for the European Skills Agenda and its Digital Education Action Plan by providing evidence on the scale and scope of the digital transition in the world of work. The first is that the more basic levels of digital skills, including computer literacy and office software, have become so pervasive and necessary that they are often no longer explicitly required in advertisements. However, basic digital skills clearly remain necessary, and are not yet universal, hence the importance of ensuring basic proficiency to foster labour market inclusion.

Second, advanced ICT occupations, including those related to Artificial Intelligence, although limited in terms of employment volume, are distinctive in scope, seem to evolve relatively rapidly, and pay notable wage premia. This may inform attempts to set targets on the number of qualified ICT professionals to be trained.

Third, the pace of digital transition of individual occupations appears to be mostly gradual, with little evidence of widespread radical disruptions in the sets of skills that define most domains. Although specialised skill profiles and emerging job titles seem to show substantial ebb and flow in their (digital) skills, better-established occupations do not show evidence of being radically transformed in short periods of time. Only in specialised fields like Artificial Intelligence, which have limited employment scale, does the set of skills required change rapidly, meaning that practitioners need to learn to frequently learn new skills, which indicates that the technology is still undergoing a process of consolidation. This state of flux may complicate any attempt to codify and set standards on “Artificial Intelligence” skills, such as developing educational curricula or professional qualifications.

Future research will explore the complex relation between skill clusters, job titles, and occupational categories to better understand how new job profiles emerge over time, and how existing ones are changing.

References

Acemoglu, D., Autor, D., Hazell, J. and Restrepo, P., 'Ai and jobs: Evidence from online vacancies', Working Paper 28257, National Bureau of Economic Research, December 2020. . URL http://www.nber.org/papers/w28257.

Acemoglu, D. and Restrepo, P., '8. artificial intelligence, automation, and work', University of Chicago Press, 2019.

Alabdulkareem, A., Frank, M. R., Sun, L., AlShebli, B., Hidalgo, C. and Rahwan, I., 'Unpacking the polarization of workplace skills', *Science advances*, Vol. 4, No 7, 2018, p. eaao6030.

Alekseeva, L., Azar, J., Gine, M., Samila, S. and Taska, B., 'The demand for ai skills in the labor market', *Labour Economics*, 2021, p. 102002.

Anderson, K. A., 'Skill networks and measures of complex human capital', *Proceedings of the National Academy of Sciences*, Vol. 114, No 48, 2017, pp. 12720–12724.

Attewell, P., 'What is skill?', *Work and occupations*, Vol. 17, No 4, 1990, pp. 422–448.

Babina, T., Fedyk, A., He, A. X. and Hodson, J., 'Artificial intelligence, firm growth, and industry concentration', *Firm Growth, and Industry Concentration (November 22, 2020)*, 2020.

Brynjolfsson, E., Mitchell, T. and Rock, D., 'What Can Machines Learn, and What Does It Mean for Occupations and the Economy?', In 'AEA Papers and Proceedings', Vol. 108. pp. 43–47.

Brynjolfsson, E., Rock, D. and Syverson, C., 'The Productivity J-Curve: How Intangibles Complement General Purpose Technologies', *American Economic Journal: Macroeconomics*, Vol. 13, No 1, 2021, pp. 333–72. . URL https://www.aeaweb.org/articles?id=10.1257/mac.20180386.

Cedefop, 'Insights into skill shortages and skill mismatch: Learning from cedefop's european skills and jobs survey', Publications Office, 2018. . URL https://data.europa.eu/doi/10.2801/897740.

Dawson, N., Williams, M.-A. and Rizoiu, M.-A., 'Skill-driven recommendations for job transition pathways', *Plos one*, Vol. 16, No 8, 2021, p. e0254722.

Dillender, M. and Forsythe, E., 'Computerization of white collar jobs', Working Paper 29866, National Bureau of Economic Research, March 2022. . URL http://www.nber.org/papers/w29866.

European Commission, 'Digital economy and society index (desi) 2020', Tech. rep., 2020. URL https://op.europa.eu/s/pdcy.

Fagan, S. and Gençay, R., 'An introduction to textual econometrics', *Handbook of empirical economics and finance*, 2011, pp. 133–154.

Felten, E. W., Raj, M. and Seamans, R., 'A Method to Link Advances in Artificial Intelligence to Occupational Abilities', In 'AEA Papers and Proceedings', Vol. 108. pp. 54–57.

Fernández-Macías, E., 'Automation, digitalisation and platforms: Implications for work and employment', 2018.

Gofman, M. and Jin, Z., 'Artificial intelligence, human capital, and innovation', *Human Capital, and Innovation (January 7, 2020)*, 2020.

Goos, M., Arntz, M., Zierahn, U., Gregory, T., Gomez, S. C., Vázquez, I. G. and Jonkers, K., 'The impact of technological innovation on the future of work', Tech. rep., JRC Working Papers Series on Labour, Education and Technology, 2019.

Montobbio, F., Staccioli, J., Virgillito, M. E. and Vivarelli, M., 'Labour-saving automation and occupational exposure: a text-similarity measure', Tech. rep., Global Labor Organization Discussion Paper, 2021.

Newman, M. E. and Girvan, M., 'Finding and evaluating community structure in networks', *Physical review E*, Vol. 69, No 2, 2004, p. 026113.

Righi, R., Samoili, S., Cobo, M. L., Baillet, M. V.-P., Cardona, M. and De Prato, G., 'The ai techno-economic complex system: Worldwide landscape, thematic subdomains and technological collaborations', *Telecommunications Policy*, Vol. 44, No 6, 2020, p. 101943.

Rodrigues, M., Fernández-Macías, E. and Sostero, M., 'A unified conceptual framework of tasks, skills and competences', Tech. rep., JRC Working Papers Series on Labour, Education and Technology, 2021.

Samek, L., Squicciarini, M. and Cammeraat, E., 'The human capital behind ai', , No 120, 2021. . URL `https://www.oecd-ilibrary.org/content/paper/2e278150-en`.

Samoili, S., Lopez Cobo, M., Gomez Gutierrez, E., De Prato, G., F., M.-P. and Delipetrev, B., 'AI watch: defining Artificial Intelligence : towards an operational definition and taxonomy of artificial intelligence.', Publications Office, LU, 2020. URL `https://data.europa.eu/doi/10.2760/382730`.

Shoham, Y., Perrault, R., Brynjolfsson, E., Clark, J., Manyika, J., Niebles, J. C., Lyons, T., Etchemendy, J. and Bauer, Z., 'The AI Index 2018 Annual Report', *AI Index Steering Committee, Human-Centered AI Initiative, Stanford University.*, Vol. 202018, 2018.

Sostero, M. and Fernández-Macías, E., 'The professional lens: What online job advertisements can say about occupational task profiles', Tech. rep., JRC Working Papers Series on Labour, Education and Technology, 2021.

Squicciarini, M. and Nachtigall, H., 'Demand for ai skills in jobs: Evidence from online job postings', 2021.

Tolan, S., Pesole, A., Martínez-Plumed, F., Fernández-Macías, E., Hernández-Orallo, J. and Gómez, E., 'Measuring the occupational impact of ai: tasks, cognitive abilities and ai benchmarks', *Journal of Artificial Intelligence Research*, Vol. 71, 2021, pp. 191–236.

Traag, V. A., Waltman, L. and Van Eck, N. J., 'From louvain to leiden: guaranteeing well-connected communities', *Scientific reports*, Vol. 9, No 1, 2019, pp. 1–12.

Vuorikari, R., Kluzer, S. and Punie, Y., 'Digcomp 2.2: The digital competence framework for citizens - with new examples of knowledge, skills and attitudes', Scientific analysis or review KJ-NA-31006-EN-N (online),KJ-NA-31006-EN-C (print), Luxembourg (Luxembourg), 2022. .

Webb, M., 'The Impact of Artificial Intelligence on the Labor Market', *Department of Economics, Stanford University, Stanford, California*, 2020.

Appendix

Table 3: Overview of Nova UK dataset of online job advertisements.

Variables in data (*topics*)	**Coverage** (% of ads)	**Quality** assessment
Job title	100	Good (cleaned)
(*Occupation*): UK SOC (4-digit) Unit/Minor/Sub-Major/Major; BGT code: group/ career area	98.7	Fair
Employer identity	33	Good (cleaned)
(*Location*): City, Language, Country, Nation, Region, Travel-to-work Area, Coordinates, Local authority	78	Very good
(*Sector/Industry*): UK SIC Code	51	Fair: 1–4 digits
Skills	≈4/ad	
– BGT skill classification (cluster/family);	70 inst.	Fair
– flags: foreign language, baseline, specialized, software	100 inst.	Fair
(*Education*): Min/Max/Preferred ed. level, major/license	17	Good
(*Certifications or licences*): UK NQF levels, years of experience required (min/max)	16	Good
(*Contract/conditions*): permanent/temporary contract, working hours, internship, work from home	75-100	Good
(*Salary*): (min-max range salary range; hourly salary)	62	
(*Misc.*): Company stock ticker, local enterprise partnership	Very low	Good

Source: own elaboration from Burning Glass Nova UK dataset, January 2012 to January 2020.

Table 4: Summary Statistics

	Regression Sample		Full Sample	
	Mean	SD	Mean	SD
Number of skills	5.095	4.842	5.142	5.118
Number of skill clusters	1.525	0.942	1.513	0.977
Hourly salary	15.618	9.551	15.618	9.551
Required years of experience	0.394	1.377	0.398	1.411
Skill clusters				
Niche cluster	0.010	0.098	0.011	0.105
Transversal Office / Administrative	0.783	0.412	0.777	0.416
Engineering & System Design	0.178	0.383	0.175	0.380
Health	0.147	0.354	0.135	0.342
Advanced ICT	0.137	0.344	0.135	0.342
Marketing & Design	0.104	0.305	0.107	0.309
ICT Support	0.089	0.284	0.090	0.287
STEM & Medical Research	0.020	0.140	0.023	0.150
Big Data & AI	0.035	0.184	0.038	0.190
Logistics & Industrial Engineering	0.017	0.129	0.017	0.128
Language	0.002	0.049	0.003	0.052
Oil & Gas	0.000	0.013	0.000	0.022
Medical Administration & Dental Health	0.000	0.019	0.000	0.021

Continued on next page

Table 4 – continued from previous page

	Regression Sample N	Regression Sample %	Full Sample N	Full Sample %
Salesforce	0.001	0.033	0.001	0.033
Wellness Service	0.001	0.024	0.001	0.033
	N	%	N	%
Job hours				
N/A	7,968,744	21.07	15,872,301	26.14
full time	28,495,544	75.35	42,565,886	70.10
part time	1,351,497	3.57	2,287,332	3.77
Contract type				
N/A	7,532,484	19.92	15,094,203	24.86
intern	885	0.00	7,263	0.01
permanent	25,781,226	68.18	38,602,909	63.57
temporary	4,501,190	11.90	7,021,144	11.56
Educational degree				
N/A	31,460,986	83.20	50,376,505	82.96
0 to 11 years	1,695,518	4.48	2,657,816	4.38
12 to 15 years	1,411,156	3.73	2,177,104	3.59
16 to 17 years	2,970,089	7.85	5,033,014	8.29
More than 17 years	278,036	0.74	481,080	0.79
Occupational group				
Managers, Directors & Senior Officials	4,106,974	10.86	6,560,813	10.80
Professional occ.	12,600,947	33.32	20,282,505	33.40
Associate Professional & Technical occ.	6,600,155	17.45	10,459,543	17.22
Administrative & Secretarial occ.	3,327,564	8.80	5,101,237	8.40
Skilled Trades occ.	2,333,487	6.17	3,710,592	6.11
Caring, Leisure & Other Service occ.	2,023,690	5.35	3,356,406	5.53
Sales & Customer Service occ.	3,579,915	9.47	5,769,033	9.50
Process, Plant & Machine Operatives	1,269,690	3.36	1,943,671	3.20
Elementary occ.	1,528,961	4.04	2,727,433	4.49
Sector group				
Agriculture, Forestry & Fishing	22,806	0.06	51,295	0.08
Mining & Quarrying	44,486	0.12	116,338	0.19
Manufacturing	2,370,125	6.27	3,647,871	6.01
Electricity, Gas, Steam & Air Conditioning Supply	54,564	0.14	89,806	0.15
Water Supply; Sewerage, Waste Mgt.	118,864	0.31	173,050	0.28
Construction	382,843	1.01	754,484	1.24
Wholesale & Retail Trade; Vehicle Repair	1,344,553	3.56	2,662,785	4.38
Transportation & Storage	661,971	1.75	1,014,856	1.67
Accommodation & Food Service Activities	1,220,468	3.23	2,350,706	3.87
Information & Communication	475,722	1.26	1,139,551	1.88
Financial & Insurance Activities	710,686	1.88	1,457,036	2.40
Real Estate Activities	386,768	1.02	640,151	1.05
Professional, Scientific & Technical Activities	1,741,570	4.61	3,468,809	5.71
Administrative & Support Service Activities	952,219	2.52	1,575,111	2.59

Continued on next page

Table 4 – continued from previous page

	Regression Sample		Full Sample	
	N	%	N	%
Public Administration, Defence, Social Security	716,418	1.89	952,778	1.57
Education	2,136,862	5.65	3,738,942	6.16
Human Health & Social Work Activities	4,437,322	11.73	6,075,761	10.01
Arts, Entertainment & Recreation	209,215	0.55	394,641	0.65
Other Service Activities	350,665	0.93	626,560	1.03
Households	20,179	0.05	29,969	0.05
Extraterritorial Organisations & Bodies	1,763	0.00	3,408	0.01
N	37,815,785		60,725,519	

Table 5: Main skill clusters (with ≥10 skills each)

	Cluster label	Top 10 skill	N Skills	Avg. degree	Avg. w. degree
1	Transversal Office / Administrative	Communication Skills, Teamwork / Collaboration, Planning, Research, Problem Solving, Organisational Skills, Writing, Project Management, Detail-Orientated, Budgeting	1929	267.94	0.92
2	Engineering & System Design	Commissioning, Systems Engineering, Mechanical Engineering, Physics, System Design, Process Design, AutoCAD, Root Cause Analysis, Engineering Management, ISO 9001 Standards	1181	126.15	0.71
3	Health	Working With Patient And/Or Condition: Mental Health, Patient Care, Social Services, Surgery, Occupational Health and Safety, Cancer knowledge, Surveillance, Oncology, Public Health and Safety, Psychology	949	71.30	0.58
4	Advanced ICT	SQL, Software Development, Oracle, Java, LINUX, JavaScript, E-Commerce, Software Engineering, Microsoft C#, Python	727	262.60	1.91
5	Marketing & Design	Optimisation, Social Media, Marketing, Editing, Facebook, Adobe Photoshop, Web Site Design, Broadcast Industry Knowledge, Market Strategy, Content Management	617	169.08	1.01
6	ICT Support	Technical Support, Troubleshooting, Technical Writing / Editing, Microsoft Windows, ITIL, Microsoft Sharepoint, System Administration, It Support, Cisco, VMware	548	187.95	1.44
7	STEM & Medical Research	Chemistry, Experiments, Biotechnology, Biology, Clinical Research, Good Manufacturing Practices (GMP), Clinical Trials, Good Clinical Practices (GCP), DNA, Biochemistry	427	63.25	0.61
8	Big Data & AI	Business Intelligence, Big Data, SAS, Data Warehousing, Data Quality, Extraction Transformation and Loading (ETL), Statistics, Apache Hadoop, Machine Learning, Statistical Analysis	367	148.33	1.29
9	Logistics & Industrial Engineering	Technical Training, Inventory Management, Inventory Control, Avionics, Industrial Engineering, Lathes, Machine Tools, Machine Operation, Warehouse Management, Environmental Engineering	132	131.25	3.89
10	Language	Japanese, Hindi, Sign Language, Teaching Speakers of Other Languages, Turkish, Urdu, English as a Second Language, Greek, Hebrew, Environmental Studies	31	50.52	0.62
11	Oil & Gas	Drilling Operations, Well Control, Petrel, Offshore Drilling, Wireline, Geophysics, Petrophysics, Coiled Tubing, Workover, Well Testing	21	19.95	0.35
12	Medical Administration & Dental Health	Issuing Receipts, Dental Equipment, Bedside Manner, Equipment Inventory, Payment Receiving, Treatment Preparation, Dental Histories, Dental Industry Knowledge, Dental Technology, Plaque Control	20	27.90	2.31
13	Salesforce	Salesforce.Com Development, VisualForce, Salesforce.Com Administration, Apex Code, Salesforce Platform Skills, Salesforce Sales Cloud, Salesforce.Com Visualforce, Salesforce Object Query Language (SOQL), Salesforce.Com Sales Cloud, Salesforce.Com Service Cloud	17	77.18	0.99
14	Wellness Service	Hair Styling, Massage Therapy, Aromatherapy, Sports Massage, Injury Prevention, Massage, Reflexology, Strength and Conditioning, Microdermabrasion, Hot Stone Massage	16	23.75	0.47

Table 6: Main skill clusters in 2013–2015 (with ≥10 skills each)

	Cluster label	Top 10 skill	N Skills	Avg. degree	Avg. w. degree
1	Transversal Office / Administrative	Communication Skills, Planning, Teamwork / Collaboration, Project Management, Research, Problem Solving, Organisational Skills, Budgeting, Writing, Customer Service	1620	167.20	0.30
2	Engineering & System Design	Commissioning, Technical Writing / Editing, Systems Engineering, Mechanical Engineering, System Design, AutoCAD, Process Design, Physics, ISO 9001 Standards, Root Cause Analysis	798	80.62	0.27
3	Health	Chemistry, Patient Care, Working With Patient And/Or Condition: Mental Health, Surgery, Biotechnology, Occupational Health and Safety, Social Services, Cancer knowledge, Psychology, Midwifery	787	43.86	0.22
4	Advanced ICT	SQL, Software Development, Oracle, Java, LINUX, E-Commerce, Microsoft C#, JavaScript, Software Engineering, SQL Server	618	160.60	0.65
5	Marketing & Design	Social Media, Optimisation, Marketing, Editing, Adobe Photoshop, Web Site Design, Facebook, Market Strategy, Content Management, Broadcast Industry Knowledge	461	103.87	0.34
6	ICT Support	Technical Support, Troubleshooting, ITIL, Microsoft Windows, System Administration, VMware, Cisco, Transmission Control Protocol / Internet Protocol (TCP / IP), Microsoft Active Directory, Windows Server	389	130.01	0.59
7	Big Data & AI	Business Intelligence, Data Warehousing, SAS, Database Administration, Big Data, Business Systems Analysis, Extraction Transformation and Loading (ETL), Database Management, Data Quality, Visual Basic for Applications (VBA)	316	96.22	0.45
8	Logistics & Industrial Engineering	Technical Training, Inventory Management, Inventory Control, Chemical Engineering, Industrial Design, Avionics, Industrial Engineering, Machine Tools, Pro/ENGINEER, Environmental Engineering	108	104.69	1.79
9	Dental health	Expediting Orders, Cardiopulmonary Resuscitation (CPR), Patient/Family Education and Instruction, Patient Preparation, Dental Care, Medical Emergencies, Oral Hygiene, Dental Equipment, Bedside Manner, Dental Histories	29	33.21	1.48
10	Language	Japanese, Polish, Hindi, Teaching Speakers of Other Languages, Turkish, Dyslexia Diagnosis / Treatment, Sign Language, Latin, Urdu, Hebrew	28	25.71	0.16
11	ICT System administration	Mainframe, COBOL, Adobe ColdFusion, Customer Information Control System (CICS), Guidewire, IBM iSeries, Adabas, Role Playing Games(RPG), Job Control Language (JCL), AS/400	26	35.27	0.22
12	Oil exploration	Geographic Information System (GIS), ArcGIS, OpenLayers, Petrel, Esri Software, Mapping, Geophysics, Petrophysics, Remote Sensing, Oracle Spatial	17	18.88	0.12
13	Salesforce	Salesforce.Com Development, VisualForce, Apex Code, Salesforce.Com Administration, Salesforce Sales Cloud, Salesforce.Com Visualforce, Salesforce Platform Skills, Salesforce Service Cloud, Salesforce.Com Service Cloud, Salesforce Object Query Language (SOQL)	14	31.07	0.29

Table 7: Main skill clusters in 2017–2019 (with ≥10 skills each)

	Cluster label	Top 10 skill	N Skills	Avg. degree	Avg. w. degree
1	Transversal Office / Administrative	Communication Skills, Teamwork / Collaboration, Planning, Problem Solving, Research, Organisational Skills, Detail-Orientated, Writing, Budgeting, Microsoft Excel	1745	172.25	0.35
2	Engineering & System Design	Commissioning, Technical Writing / Editing, Systems Engineering, Mechanical Engineering, Physics, System Design, Root Cause Analysis, Process Design, Systems Integration, AutoCAD	943	76.55	0.26
3	Health	Patient Care, Working With Patient And/Or Condition: Mental Health, Social Services, Surgery, Cancer knowledge, Occupational Health and Safety, Oncology, Rehabilitation, Midwifery, Surveillance	707	48.13	0.26
4	Advanced ICT	SQL, Software Development, Java, Python, Oracle, LINUX, JavaScript, Software Engineering, Microsoft C#, C++	628	162.51	0.72
5	Marketing & Design	Optimisation, Social Media, E-Commerce, Marketing, Editing, Facebook, Adobe Photoshop, Market Strategy, Digital Marketing, Adobe Indesign	482	118.58	0.46
6	ICT Support	Troubleshooting, Technical Support, ITIL, Microsoft Windows, Microsoft Sharepoint, It Support, Information Security, VMware, Cisco, Microsoft Active Directory	345	131.04	0.49
7	STEM & Medical Research	Chemistry, Experiments, Biotechnology, Biology, Good Clinical Practices (GCP), Good Manufacturing Practices (GMP), Clinical Research, Clinical Trials, Biochemistry, Molecular Biology	319	45.15	0.29
8	Big Data & AI	Data Management, Big Data, Business Intelligence, Machine Learning, Data Science, Data Quality, Data Warehousing, Data Collection, Tableau, Apache Hadoop	266	113.82	0.64
9	Language	Fitness, Yoga, Tutoring, Singing, Teaching Speakers of Other Languages, English as a Second Language, Greek, Piano, Exercise Programmes, Guitar	44	20.61	0.21
10	Oracle	Business Process Modelling, Business Process Modelling Notation (BPMN), Oracle E-Business Suite Financials, Oracle Fusion, Process Modelling, Oracle Cloud, Business Process Monitor (BPM), Oracle E-Business Suite, Oracle Financials, Ebusiness	28	52.25	0.19
11	ICT System administration	Mainframe, COBOL, Guidewire, Customer Information Control System (CICS), IBM iSeries, Job Control Language (JCL), Adabas, Role Playing Games(RPG), AS/400, Mainframe Systems	16	26.56	0.10
12	Salesforce	Salesforce.Com Development, Salesforce.Com Administration, VisualForce, Salesforce Platform Skills, Apex Code, Salesforce Sales Cloud, Salesforce.Com Sales Cloud, Salesforce.Com Service Cloud, Salesforce Object Query Language (SOQL), Salesforce Service Cloud	13	48.62	0.50
13	Restaurant and retail	Safety Training, Asset Protection, Report Maintenance, Point of Sale System, Identifying and Evaluating Defects, Allergies, Procuring Goods and Services, Employee Coaching, Life-Safety Systems, Fine Motor Skills	12	63.00	0.22

Table 8: AI Skills, from different sources, by number of mention across all job advertisements

Skill name	List origin	N Ads
Artificial Intelligence	BGT: Artificial Intelligence	48323
Embedded Software	BGT: Automation Engineering	112315
Automation Systems	BGT: Automation Engineering	23120
Rockwell Automation	BGT: Automation Engineering	7828
ABB	BGT: Automation Engineering	4873
Machine Vision	BGT: Automation Engineering	3889
Motion Control Systems	BGT: Automation Engineering	3042
Wonderware InTouch	BGT: Automation Engineering	2414
Arduino	BGT: Automation Engineering	2341
DeviceNet	BGT: Automation Engineering	2328
Machine Control Systems	BGT: Automation Engineering	2172
Human Machine Interface (HMI) Control Systems	BGT: Automation Engineering	2143
RSView	BGT: Automation Engineering	1608
ControlNet	BGT: Automation Engineering	1027
Big Data	BGT: Big Data	202009
Apache Hadoop	BGT: Big Data	106277
Splunk	BGT: Big Data	29029
Apache Hive	BGT: Big Data	28634
Cassandra	BGT: Big Data	17641
PIG	BGT: Big Data	12882
Apache Spark	BGT: Big Data	9054
Sqoop	BGT: Big Data	3987
Oracle Big Data	BGT: Big Data	2298
Greenplum	BGT: Big Data	2254
AWS Elastic MapReduce (EMR)	BGT: Big Data	2177
Apache Accumulo	BGT: Big Data	989
Hives	BGT: Big Data	449
Talend Big Data	BGT: Big Data	383
IBM InfoSphere BigInsights	BGT: Big Data	345
Datameer	BGT: Big Data	248
Platfora Big Data Discovery Platform	BGT: Big Data	232
Statistical Analysis	BGT: Data Analysis	80052
Big Data Analytics	BGT: Data Analysis	41311
Data Reports	BGT: Data Analysis	16394
Data Trending	BGT: Data Analysis	8274
Alteryx	BGT: Data Analysis	7508
Quantitative Data Analysis	BGT: Data Analysis	6868
Verint Systems	BGT: Data Analysis	5015
CANalyzer	BGT: Data Analysis	3032
Exploratory Analysis	BGT: Data Analysis	1445
Data Quality Assessment	BGT: Data Analysis	1231
Qualitative Data Analysis	BGT: Data Analysis	1055
HP Vertica	BGT: Data Analysis	230
Atlas.ti	BGT: Data Analysis	74
HP Sprinter	BGT: Data Analysis	72

Continued on next page

Table 8 – continued from previous page

Skill name	List origin	N Ads
Expectation-Maximisation (EM) Algorithm	BGT: Data Analysis	10
Data Mining	BGT: Data Mining	59395
Data Capture	BGT: Data Mining	53931
Text Mining	BGT: Data Mining	3805
Analysis Computing Tools	BGT: Data Mining	1927
WEKA	BGT: Data Mining	859
Knowledge Discovery	BGT: Data Mining	619
Conceptual Data Models	BGT: Data Mining	593
CRISP-DM	BGT: Data Mining	187
Web Mining	BGT: Data Mining	149
Unstructured Information Management Architecture	BGT: Data Mining	79
SEMMA	BGT: Data Mining	39
Data Science	BGT: Data Science	105178
Predictive Models	BGT: Data Science	23213
Predictive Analytics	BGT: Data Science	14599
Cluster Analysis	BGT: Data Science	11055
Pandas	BGT: Data Science	7767
Chi Square Automatic Interaction Detection (CHAID)	BGT: Data Science	6035
Pattern Recognition	BGT: Data Science	3135
Time Series Models	BGT: Data Science	3023
Ontologies	BGT: Data Science	2379
Factor Analysis	BGT: Data Science	2069
Monte Carlo Simulation	BGT: Data Science	1965
Social Network Analysis	BGT: Data Science	1810
Bayesian Methods	BGT: Data Science	1428
Time Series Forecasting	BGT: Data Science	1175
Information Extraction	BGT: Data Science	1003
K-Means	BGT: Data Science	905
Bayesian Networks	BGT: Data Science	665
Markov Chains	BGT: Data Science	656
Principal Component Analysis (PCA)	BGT: Data Science	622
Naive Bayes	BGT: Data Science	541
Bayesian Modelling	BGT: Data Science	399
Kernel Methods	BGT: Data Science	167
Graph-Based Algorithms	BGT: Data Science	151
Stochastic Optimisation	BGT: Data Science	54
Fuzzy Clustering	BGT: Data Science	1
Data Services Industry Knowledge	BGT: Data Services Industry Knowledge	619
Data Protection Industry Knowledge	BGT: Data Services Industry Knowledge	68
Database Development Industry Knowledge	BGT: Data Services Industry Knowledge	3
Data Mining Industry Knowledge	BGT: Data Services Industry Knowledge	1
Data Collection	BGT: Data Techniques	121764
Data Manipulation	BGT: Data Techniques	72128

Continued on next page

Table 8 – continued from previous page

Skill name	List origin	N Ads
Data Governance	BGT: Data Techniques	44605
Pipeline (Computing)	BGT: Data Techniques	11165
Data Conversion	BGT: Data Techniques	8313
Data Documentation	BGT: Data Techniques	5309
Data Cleaning	BGT: Data Techniques	5063
Data Encryption	BGT: Data Techniques	3144
Data Evaluation	BGT: Data Techniques	1854
Semi-Structured Data	BGT: Data Techniques	517
Machine Learning	BGT: Machine Learning	111656
Computer Vision	BGT: Machine Learning	14669
Deep Learning	BGT: Machine Learning	13653
Neural Networks	BGT: Machine Learning	8932
Decision Trees	BGT: Machine Learning	7015
OpenCV	BGT: Machine Learning	2642
Recommender Systems	BGT: Machine Learning	2331
Mahout	BGT: Machine Learning	2327
Random Forests	BGT: Machine Learning	2211
Support Vector Machines (SVM)	BGT: Machine Learning	2093
Object Recognition	BGT: Machine Learning	532
Object Tracking	BGT: Machine Learning	329
Natural Language Processing	BGT: Natural Language Processing	22719
Speech Recognition	BGT: Natural Language Processing	2949
Sentiment Analysis / Opinion Mining	BGT: Natural Language Processing	2081
Computational Linguistics	BGT: Natural Language Processing	1616
Tokenisation	BGT: Natural Language Processing	750
Machine Translation (MT)	BGT: Natural Language Processing	584
Natural Language Toolkit (NLTK)	BGT: Natural Language Processing	493
Automatic Speech Recognition (ASR)	BGT: Natural Language Processing	492
Text to Speech (TTS)	BGT: Natural Language Processing	482
OpenNLP	BGT: Natural Language Processing	139
Nearest Neighbour Algorithm	BGT: Natural Language Processing	51
Latent Dirichlet Allocation	BGT: Natural Language Processing	44
Sentiment Classification	BGT: Natural Language Processing	30
Lexical Semantics	BGT: Natural Language Processing	21
Lexical Acquisition	BGT: Natural Language Processing	14
Latent Semantic Analysis	BGT: Natural Language Processing	8
Lexalytics	BGT: Natural Language Processing	3
Information Retrieval	JRC: AI WATCH	5213
Regression Algorithms	JRC: AI WATCH	4651
Classification Algorithms	JRC: AI WATCH	3399
Supervised Learning (Machine Learning)	JRC: AI WATCH	3194
Unsupervised Learning	JRC: AI WATCH	2796
Autonomous Systems	JRC: AI WATCH	2756
Caffe Deep Learning Framework	JRC: AI WATCH	1641
Bayesian Inference	JRC: AI WATCH	1597
Convolutional Neural Network (CNN)	JRC: AI WATCH	1165

Continued on next page

Table 8 – continued from previous page

Skill name	List origin	N Ads
Recurrent Neural Network (RNN)	JRC: AI WATCH	698
Genetic Algorithms	JRC: AI WATCH	588
Boosting (Machine Learning)	JRC: AI WATCH	580
Clustering Algorithms	JRC: AI WATCH	490
Network Intelligence	JRC: AI WATCH	353
Cascading Big Data Applications	JRC: AI WATCH	177
Semi-Supervised Learning	JRC: AI WATCH	147
Classification Software	JRC: AI WATCH	51
Unmanned Vehicle Systems	JRC: AI WATCH	36
Maximum Entropy Classifier	JRC: AI WATCH	19
DBSCAN (Density-Based Spatial Clustering of Applications with Noise)	JRC: AI WATCH	17
Stochastic Gradient Descent (SGD)	JRC: AI WATCH	16
Optical Character Recognition (OCRS)	JRC: AI WATCH	7
TensorFlow	JRC: desk research	7572
Scikit-learn	JRC: desk research	3686
Theano	JRC: desk research	1422
Torch	JRC: desk research	590
AdaBoost algorithm	JRC: desk research	8

Table 9: Wage premia of skill clusters

Dependent variable:	log(proposed hourly salary)					
	(1)	(2)	(3)	(4)	(5)	(6)
# skills (truncated at 6)	0.021***	0.017***	0.002+	0.000	0.004**	-0.001+
	(0.003)	(0.002)	(0.001)	(0.000)	(0.001)	(0.000)
Big Data & AI	0.209***	0.178***	0.108***	0.042***	0.036***	0.029***
	(0.005)	(0.009)	(0.006)	(0.004)	(0.003)	(0.004)
Advanced ICT	0.293***	0.269***	0.159***	0.032***	0.076***	0.032***
	(0.025)	(0.019)	(0.014)	(0.003)	(0.003)	(0.003)
ICT Support	0.078***	0.072***	0.063***	0.011**	0.023***	0.011*
	(0.008)	(0.007)	(0.005)	(0.002)	(0.003)	(0.005)
Salesforce	0.373***	0.298***	0.227***	0.011	0.077**	0.011
	(0.026)	(0.030)	(0.027)	(0.008)	(0.016)	(0.015)
Niche cluster	0.011	0.008	0.014+	0.010*	0.008*	0.008*
	(0.006)	(0.006)	(0.006)	(0.003)	(0.003)	(0.003)
Transversal Office/	-0.084***	-0.091***	-0.047***	-0.015***	-0.025***	-0.013***
Administrative	(0.006)	(0.003)	(0.003)	(0.001)	(0.001)	(0.001)
Engineering &	0.087**	0.106***	0.041**	0.017**	0.034***	0.022***
System Design	(0.023)	(0.012)	(0.009)	(0.004)	(0.003)	(0.003)
Health	0.019	0.037	0.011	-0.004	0.027***	0.003
	(0.029)	(0.025)	(0.014)	(0.004)	(0.003)	(0.003)
Marketing & Design	-0.018**	-0.051**	-0.042***	-0.012**	-0.015**	-0.002
	(0.004)	(0.011)	(0.007)	(0.003)	(0.003)	(0.003)
STEM & Medical	0.013	0.006	-0.003	0.011*	-0.012*	0.015**
Research	(0.009)	(0.007)	(0.009)	(0.004)	(0.004)	(0.003)
Logistics &	-0.088***	-0.056***	-0.014*	0.010*	-0.017**	0.011**
Industrial Engineering	(0.016)	(0.008)	(0.006)	(0.003)	(0.005)	(0.003)
Language	-0.071*	-0.095*	-0.048*	-0.017***	-0.007	-0.018***
	(0.026)	(0.034)	(0.017)	(0.003)	(0.006)	(0.003)
Oil & Gas	0.309***	0.295***	0.295***	0.102**	0.103+	0.078*
	(0.056)	(0.055)	(0.051)	(0.029)	(0.050)	(0.025)
Medical Administration	-0.271***	-0.262***	-0.104**	-0.025+	-0.076**	-0.038**
& Dental Health	(0.037)	(0.034)	(0.030)	(0.012)	(0.015)	(0.011)
Wellness Service	-0.158***	-0.161***	0.088***	0.009	0.035**	0.017
	(0.028)	(0.028)	(0.009)	(0.009)	(0.008)	(0.010)
schooling $\leq$ 11	-0.486***	-0.442***	-0.300***	-0.039***	-0.221***	-0.032***
	(0.068)	(0.059)	(0.048)	(0.006)	(0.026)	(0.006)
11 < schooling $\leq$ 15	-0.081*	-0.066*	-0.064***	-0.014**	-0.048***	-0.010**
	(0.027)	(0.021)	(0.011)	(0.003)	(0.007)	(0.002)
15 < schooling $\leq$ 17	0.089***	0.072**	0.004	-0.004	0.027**	0.008+
	(0.018)	(0.021)	(0.011)	(0.004)	(0.008)	(0.004)
schooling > 17	0.204***	0.180***	0.060***	0.016**	0.057***	0.014*
	(0.023)	(0.027)	(0.008)	(0.004)	(0.005)	(0.005)
experience	0.050***	0.045***	0.028***	0.008***	0.022***	0.012***
	(0.003)	(0.004)	(0.003)	(0.001)	(0.001)	(0.001)
experience^2	-0.002***	-0.002**	-0.001*	0.000+	-0.001***	-0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
N	37,815,785	37,815,785	37,815,785	37,815,042	10,613,234	18,365,601
R2 Adj.	0.173	0.24	0.446	0.78	0.594	0.779
FE: Year	X	X	X	X	X	X
FE: City		X	X	X	X	X
FE: SOC (4 digit)			X	X	X	X
FE: Job title				X		X
FE: SIC (2 digit)						X
FE: Employer					X	

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001.
Not listed control variables: job type (intern,temporary,permanent), job hours (part-time, full-time)

Figure 12: Changing compositions of skill clusters

Flows of skills between the skill clusters of 2013−2105 and those of 2017−2019

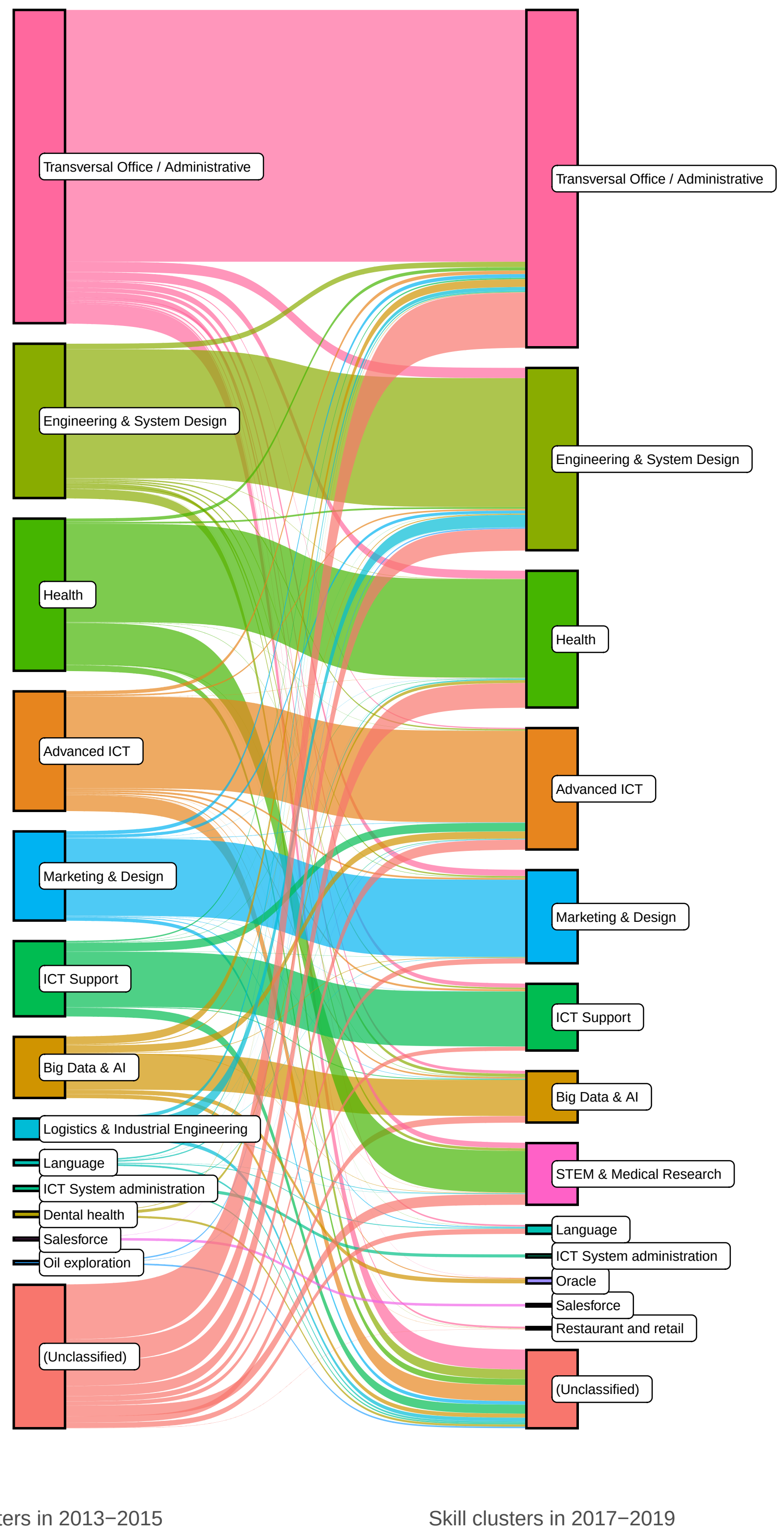
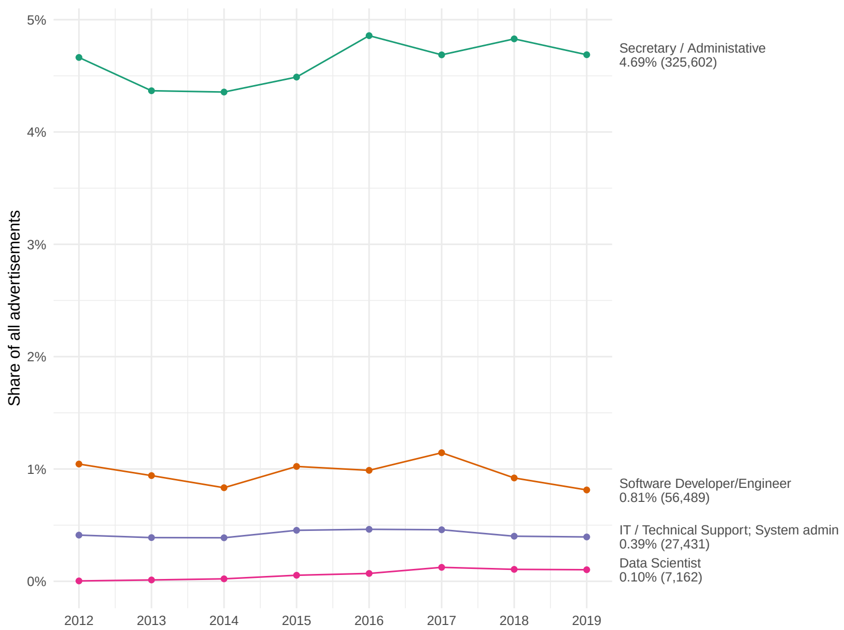


Figure 13: Prevalence of selected job titles among UK online job advertisements



GETTING IN TOUCH WITH THE EU

In person

All over the European Union there are hundreds of Europe Direct centres. You can find the address of the centre nearest you online (european-union.europa.eu/contact-eu/meet-us_en).

On the phone or in writing

Europe Direct is a service that answers your questions about the European Union. You can contact this service:

— by freephone: 00 800 6 7 8 9 10 11 (certain operators may charge for these calls),

— at the following standard number: +32 22999696,

— via the following form: european-union.europa.eu/contact-eu/write-us_en.

FINDING INFORMATION ABOUT THE EU

Online

Information about the European Union in all the official languages of the EU is available on the Europa website (european-union.europa.eu).

EU publications

You can view or order EU publications at op.europa.eu/en/publications. Multiple copies of free publications can be obtained by contacting Europe Direct or your local documentation centre (european-union.europa.eu/contact-eu/meet-us_en).

EU law and related documents

For access to legal information from the EU, including all EU law since 1951 in all the official language versions, go to EUR-Lex (eur-lex.europa.eu).

Open data from the EU

The portal data.europa.eu provides access to open datasets from the EU institutions, bodies and agencies. These can be downloaded and reused for free, for both commercial and non-commercial purposes. The portal also provides access to a wealth of datasets from European countries.

The European Commission's science and knowledge service

Joint Research Centre

JRC Mission

As the science and knowledge service of the European Commission, the Joint Research Centre's mission is to support EU policies with independent evidence throughout the whole policy cycle.

EU Science Hub
joint-research-centre.ec.europa.eu

@EU_ScienceHub

EU Science Hub - Joint Research Centre

EU Science, Research and Innovation

EU Science Hub

EU Science