IS THE WORKFORCE READY FOR THE JOBS OF THE FUTURE? DATA-INFORMED SKILLS AND TRAINING FORESIGHT

FABIAN STEPHANY AND ROSEMARY LUCKIN

For many newly emerging jobs, labour-market mismatches prevail as workers and firms are unable to apply precise occupation taxonomies and training lags behind workforce needs. We report on how data can enable useful foresight about skill requirements and training needs, even when that data has not been collected for this express purpose. First, we show how online generated freelance data can help monitor labour-market developments in the short run. Second, in the long run, we illustrate how data can shed light on development of workplace-ready aptitudes among students, even when these are not the direct focus of instruction. This combination of data-intensive activities can inform the immediate and long-term needs for education and training in order to help individuals develop the ability to learn, train and retrain as often and as much as needed.

Fabian Stephany (fabian.stephany@oii.ox.ac.uk) is a Departmental Research Lecturer at Oxford Internet Institute, University of Oxford
Rosemary Luckin (r.luckin@ucl.ac.uk) is a professor at London Knowledge Lab, Institute of Education, London

bruegel

1 Introduction

Technological and social change alter the required skill composition of the workforce. For many newly emerging jobs, precise skill requirements are evolving and unclear, resulting in a mismatch between occupational taxonomies and these novel roles in fields including data management, digital design, and autonomous systems. We explore and discuss two implications that arise from this mismatch between workforce talent and job vacancies. First, the mismatch has harmful consequences for workers and firms, which are unable to attain their optimal productivity. Second, the mismatch has significant and complex implications for educators and trainers, whose focus is on the young people who represent the near and mid-term future workforce. These implications present something of a 'wicked' problem, with many moving interdependencies. Apparent solutions are likely to uncover yet more challenges. In this work, we try to address both the immediate need for systematic oversight and the long-run necessity for sustainable solutions in reskilling.

In the short run, the first problematic implication of the talent mismatch might be addressed through the use of online-generated data that could help provide a farsighted perspective on labour-market developments. By monitoring the skill formation on online labour platforms, we can see the emergence of new occupations and identify the skills required for this new type of work. The accurate identification of future labour-market needs can enable economies to provide immediate 'emergency' upskilling to the impacted sections of their workforce. However, this is only a short-term fix.

Within a longer time horizon, a sustainable solution will require radically different action in education and training reform. Such reform must ensure that the workforce is 'future-proofed' in two ways. First, training and education for each specific predicted workforce expertise must be timely. It must be provided *in advance* of the workforce-need becoming critical. To ensure that specific training stays ahead of workforce needs, a better dynamic connection must be made between emerging and evolving occupational analysis and the frameworks that underpin what and how young people are taught and trained. Second, education and training must develop within people the motivation and ability to be extremely good at learning, re-learning, training and retraining, as often and as much as is needed.

2 The EU skill gap and strategies to close it

Technological change drives skill mismatches

Technology is changing fundamentally the way we work. Technological and social transformation change the skill composition of work (Acemoglu and Autor, 2011). The work that is thereby eliminated

has different skill requirements than the newly created jobs, which leads to the paradox of simultaneous unemployment and labour shortages (Autor, 2015). Often, precise skill requirements for mastering emerging technologies remain opaque (De Mauro *et al*, 2018) and – despite their growing demands – new occupations, in fields including data management, digital design and autonomous systems, are not acknowledged by official employment taxonomies. This is bad news for both firms and workers, as they need to 'speak' the same language when hiring great talent or finding a new job.

History suggests that this skills gap, even more so than the elimination of jobs *per se*, causes heightened economic inequality (Card and DiNardo, 2002) and retards firm growth (Krueger and Kumar, 2004) during times of technological and social transformation. In the situation of rapidly changing skill requirements and nameless new occupations, systematic oversight is essential. However, individuals often lack foresight about which occupations are about to emerge or demand for which skills is rising and falling. They might get locked into path dependencies that may result in dead ends that prevent them from reskilling into new areas (Escobari *et al*, 2019).

Skill mismatches can impact earnings negatively, as individuals might accept a less-desirable job because of greater competition, as the example of high-skilled knowledge workers, who juggle multiple, short-term and often precarious online jobs shows. This temporary impact might also lead to permanent effects in the form of human capital depreciation. Skill mismatches and skill shortages distort the optimal allocation of resources and thereby reduce average productivity. In terms of GDP loss, Mavromaras *et al* (2007) proxied the individual productivity loss with the estimated wage penalty associated with over-skilling, and multiplied this penalty by the number of overskilled workers by educational attainment level, concluding that the costs of over-skilling amount to about 2.6 percent of GDP.

The Beveridge curve that describes the relationship between unemployment and job vacancy rates allows us to study the impact of business cycles on skill mismatches and their development over time. In times of economic contractions, vacancy rates decrease and unemployment rises. In times of skill mismatches or increased search intensity, shifts in the curve can occur when unemployment rises given a specific level of vacancies. Figure 1 shows that in the European Union, the Beveridge curve shifted outwards during the 2008-2021 period. After 2013, job vacancy rates increased and unemployment rates declined. At time of writing, vacancy rates are higher than 10 years ago, with a slightly higher level of unemployment, indicating that despite economic recovery, many firms are not able to fill their vacancies.

Figure 1: In the EU, the Beveridge curve shifted outwards during the 2008-2021 period; job vacancy rates increased and unemployment rates declined
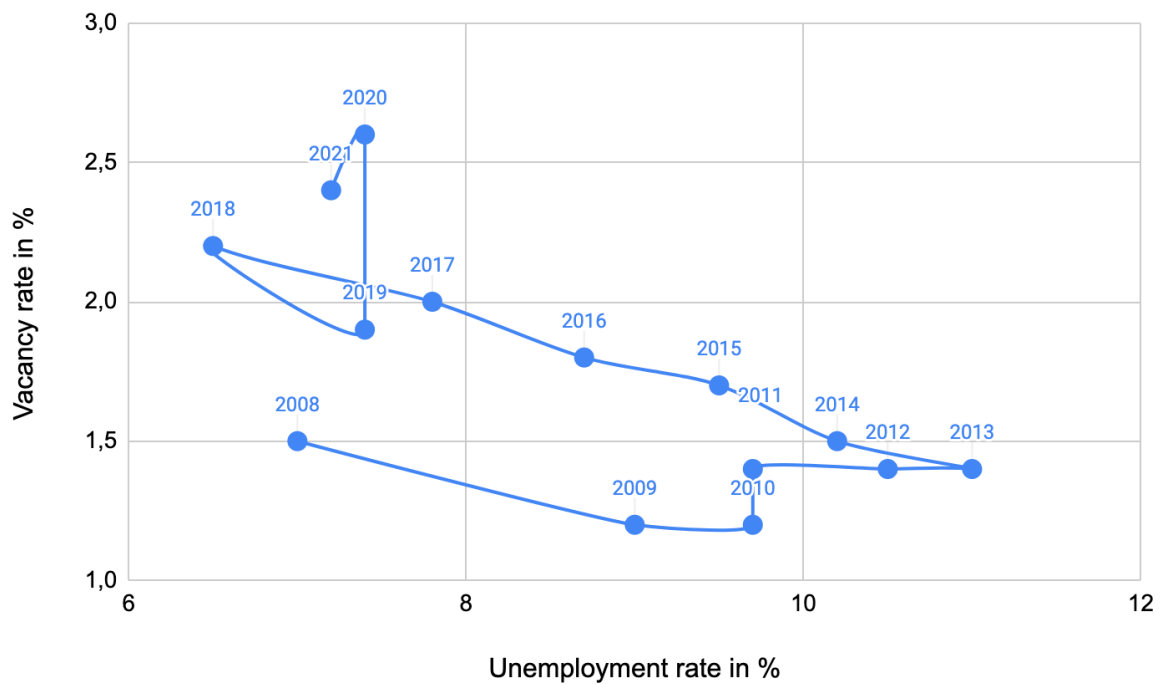
Source: Own calculations based on Eurostat.

In economic downturns, matching efficiency, defined as the quality of the matching process involving unemployed workers and unfilled vacancies, declines. Active labour market policies (ALMP) can help improve the matching efficiency by re- and up-skilling measures, which improve the matching prospects for the long-term unemployed. Re- and up-skilling policies will be increasingly relevant as job polarisation due to automation and digitisation increases, and longer unemployment together with rapidly changing skill requirement due to technological change worsens re-employment prospects.

For the EU, skills and job forecasts by CEDEFOP (Pouliakas, 2021) indicate that further job market polarisation is on the rise with rising employment shares for professionals, managers and technicians, and a decline in labour-market demand for clerks, craft workers and plant and machine operators. Typically, jobs at low risk of automation require professional training or tertiary education, but workers, and similarly employers, in domains that are most exposed to high automation risk are also less likely to invest in training (see Nedelkoska and Quintini, 2018) and often have limited access to it. A central question for ALMP in targeting a reduction of the skill mismatch is therefore what type of training can effectively allow people to upgrade skills and move to jobs that are less automatable (Tamm, 2018; Schmidpeter and Winter-Ebmer, 2018).

With increasing international labour-market competition and rapid technological change, skill requirements will continue to shift. Research shows that skills complementary to new technology requirements facilitate the adaptation to changing job requirements and carry additional value for employees. For firms that want to adopt new technologies and the practices that come with them, employees' skills are becoming increasingly important (EIB, 2018).

Existing policy solutions try to close the skills gap

Most policies targeting the skills mismatch try to enhance the responsiveness of the education and training sector to the newly emerging demand from labour markets. This includes, for example, an enhancement of youth employability by reforming VET (vocational education and training) education, or better forecasts of future skills needed to meet labour-market demands. Likewise, policy can target skill mismatches by reducing the information asymmetry between jobseekers, workers and firms, because they might be using different taxonomies to describe skill requirements and existing capabilities (Colahan *et al*, 2017).

The responsibility for the development of 'future-proofed' skills among schools and employers is unclear. For schools it is often difficult to foresee the labour-market demands of the next few years (Cappelli, 2014), while constraints on what should be taught can lead to inflexibility. In the absence of skill-demand forecasts, pupils might invest in more academic skills as they want to cushion the potential costs of skill obsolescence in the medium to long run (Brunello and Rocco, 2017). Skill mismatches that are not solved by market mechanisms can be targeted by policies that adjust the under-provision of education or training. In Europe, levy-grant schemes, tax deductions and co-financing programmes targeted at individuals are examples of public policies encouraging adult training that include co-financing programmes targeted at firms (see Brunello and Wruuck, 2019).

Lastly, on the European level, skill policies can and must respond proactively to skill mismatches caused by structural trends such as technological change, ie digitalisation and automation, which lead to labour-market polarisation or inequality. It will be increasingly important to ensure the smooth transition between jobs exposed to structural change, such as automation pressure, and the development of new and sustainable qualifications. This is a challenge both for firms requiring talent skilled with new capabilities, and for job seekers, who enter the European labour market or try to find new jobs via adult re-skilling. The monitoring of occupation taxonomies and skill requirements via online labour market (OLM) data, as presented here, is one idea in a stream of research that leverages novel

data sources and methods of computational social science in order to recommend future-proofed skills against the backdrop of ongoing technological change.

The European Commission's Pact for Skills[1], launched in 2020, recognises both the need for a data-driven and targeted approach to reskilling, and the inclusion of public-private partnerships in the process. The overall goal of the Pact is to maximise the impact and effectiveness of skills investment. A focal point of the Pact is the aspect of upskilling and reskilling in the vocational training sector. For a successful implementation, two aspects are important. On one hand, industry needs for specific skills have to be made explicit; on the other hand, the unique training history of workers needs to be acknowledged. The approach to monitor occupation taxonomies and skill requirements via online labour market data, as presented here, is a proposal for a targeted and near-real time reskilling advice regarding both industry needs and worker skills.

3 Improving skill-matching efficiency with online labour-market data

In the absence of institutional support, independent professionals today develop new skills incrementally, adding closely related skills to their existing portfolios (Lehdonvirta *et al*, 2019). Lehdonvirta *et al* (2019) examined the skill development of freelancers using online labour platforms, which are global marketplaces that match millions of buyers and sellers of digitally delivered work in various occupational domains (Horton, 2010). These platforms are websites that mediate between buyers and sellers of remotely deliverable cognitive work (Horton, 2010). The sellers of labour on OLMs are either people in regular employment earning additional income via the Internet as freelancers, or they are self-employed independent contractors. The buyers of labour range from individuals and early-stage startups to Fortune 500 companies (Corporaal and Lehdonvirta, 2017). OLMs can be further subdivided into micro-task platforms, eg Amazon Mechanical Turk, where payment is on a piece-rate basis, or freelancing platforms, such as UpWork, where payment is on an hourly or milestone basis (Lehdonvirta, 2018). Between 2017 and 2020, the global market for online labour grew approximately 50 percent (Kässi and Lehdonvirta, 2018). In the wake of the COVID-19 pandemic and it's significant economic repercussions across industries (Stephany *et al*, 2020a), OLMs continue to increase in popularity due to a general trend of increasing remote working (Stephany *et al*. 2020b).

While the idea of studying OLM data for skill monitoring is new, a few social data science scholars have explored alternative approaches using online generated data for investigating skill formation. De Mauro *et al* (2018), for example, showcased the skill complexity of the new profession of data science with

---

[1] See https://ec.europa.eu/social/main.jsp?catId=1517&langId=en.

data retrieved from various job boards. Similarly, Calanca *et al* (2019) demonstrated the increasing relevance of soft-skills in a large body of online job vacancies. Bastian *et al* (2014), on the other hand, made use of data from LinkedIn, the world's most popular professional online social network, to compare the relevance of certain 'hard skills' across industry domains. Alternatively to OLM data, information from online job vacancy portals, such as Indeed or Glassdoor, or professional social network sites, like LinkedIn, could be used to provide information about skill developments and the emergence of novel occupational domains. However, all three data avenues have unique advantages and shortcomings, as summarised in Table 1.

Table 1: Compared to data from online job vacancy sites and career portals, online labour market data allows for the study of both the demand and supply side of work, including relevant information on prices

Data source	Demand side	Supply side	Price information	Industry coverage
Online job vacancies	Yes	No	Seldom	Large
Networking sites	No	Yes	No	Medium
Online labour markets	Yes	Yes	Yes	Small

While online job vacancies cover a large segment of the labour market, including many industry sectors and also potentially non-digital and manual work, they seldom include information on price levels and give no indication of the possible supply in the targeted population. Data from professional social media sites, like LinkedIn, on the other hand, enables an in-depth analysis of skill compositions in the population; however, no price or income information is revealed and matching efficiencies can't be evaluated in the absence of demand-side data. OLM data, lastly, only covers a small segment of the labour market, ie digitalised tasks from jobs in the professional services sector. However, this data does contain information on both the demand and supply sides of skills. In addition, it is possible to observe the matching process and price, eg hourly rate, for each job with a certain skill bundle attached to it. These properties make OLMs an interesting data source for studying skill formation, skill matching and the evaluation of individual skills or skill bundles (Stephany, 2021).

With regard to skill formation, Stephany (2021) showed that online labour markets provide relevant data for active labour-market policies in the domain of reskilling, as they are characterised by a high level of skill elasticity. This means that, in contrast to employees, for whom market competition and market premia are, to a large extent, mitigated by the firm, online freelancers have several strong and immediate incentives to acquire a new capability once it becomes marketable. Online freelancers who quickly acquire a newly demanded skill can cash in the global market premium attached to this new capacity. For employees, this positive incentive to develop a new skill might be weaker, as it is mediated by a firm that relies on a small set of locally defined customers only. However, online workers have a strong incentive to constantly re-skill, as they are exposed to global competition. Here, again, employees are shielded by their company, which mitigates market competition on the employee level.

In light of this strong skill elasticity of online workers, the work by Stephany (2021) showed how online labour-market data enables us to monitor skill rebundling in a global workforce in near real-time with up-to-date skill bundles on a granular level. The study used the rich toolbox of network analysis for the characterisation of skill relationships and argued that online labour platforms have become early 'laboratories' of the rebundling of skill sets. OLM data allows us to monitor skill rebundling in a global workforce by near real-time reporting location, wages asked, previous income, gender attributes (forenames), and up-to-date skill bundles on a granular level. This process involved various protection and encryption methods to respect data protection law. For example, when dealing with individual workers, personally identifiable information, such as locations or IP addresses were anonymised. The information on gender or income is only processed on an aggregated basis.

Here, we showcase the potential of online labour market data with the example of three novel domains: General Data Protection Regulation (GDPR), robotics, and virtual reality. The data for this work stems from the Online Labour Index database (Stephany *et al*, 2021). Many of the job profiles that emerge on online labour platforms through the workers' versatile combinations of different established and new skills give birth to novel occupational domains long before they are accredited by official employment taxonomies. We propose the exploration of OLM data to improve skill matching efficiencies with regard to two aspects:

1. **Demand**: The data from OLM projects provides not only information on the extent of labour demand in a specific domain but also on the evaluation in terms of hourly wages.

2. **Context**: Reducing the asymmetric information between jobseekers, workers and firms, as OLMs provide a detailed description of occupational domains or adjacent capacities that new domains require.

Figures 2a-c show three exemplary job profiles from a popular global remote work platform: Autonomous Robotics Engineer, GDPR Expert, Virtual Reality (VR) Producer. While labour demand for all of these three domains is certainly growing, official occupational taxonomies, such as the Standard Occupational Classification System (SOC), do not list them yet[2].

Figure 2a-c: The example profiles of three online freelancers show how 'conventional' and novel skills, in the domains of robotics, compliance, and design are re-combined in new skill domains

Mohamed H.
Autonomous Robotics, ROS1/2, AWS, Robotics-IoT, Web, Embedded Engineer
Egypt

$60.00 /hr   100% Job Success   TOP RATED

Autonomous Robotics Software R&D Engineer | Founder @ DevServers We have special offers for long-term collaboration! In the past 2 years, I managed to help many clients with their Robotics projects over the world (Canada, U…

Computer Vision   Machine Learning   Robotics   Robot Operating System (ROS)   Reinforcement Learning   Deep Learn >

Associated with
DevServers

Maria A.
GDPR Compliance & Data Privacy consulting
Germany

$180.00 /hr $2k+ earned

GDPR consultant and trainer. Privacy researcher • Certified Information Privacy Manager (CIPM by IAPP) • TÜV (Germany) Certified DPO • Trained in Strategic Privacy by Design, • Both in-house and external DPO experience • 5…

GDPR   Data Privacy   Policies & Procedures   Employee Training   Regulatory Compliance   Gap Analysis   Contract >

Dmitri L.
VR Producer, UX/UI Designer, 360 Video VFX Artist
Israel

$75.00 /hr $80k+ earned   100% Job Success   TOP RATED

I produce VR apps: presentations, training, tours, experiences. 5+ years in the industry. If you need a cost-efficient VR solution - I will build it, from scratch to the final product. I have 4 types of VR glasses to test my apps. ✅ Key skills: UI/UX…

AR/VR Consultation   VR   UI/UX Prototyping   Project Management   Product Management   3D Animation   CGI   >

The rich dataset of online freelance profiles enables us to identify newly emerging skill domains and to observe the combination of skills they require in near real-time. The skill-based occupation model (Stephany, 2021), enables us to assess which skills are required for these new types of jobs. The forward-looking perspective of online labour-market data and the granularity of skill portfolios allow us

---

[2] https://www.bls.gov/oes/current/oes_stru.htm.

to categorise new skill domains within existing taxonomies of skills and occupations, across time and space.

Figure 3 a-c illustrates the skill composition of the above-mentioned three jobs in 2021 and relates the required skills to established taxonomies, eg the exemplary job of a VR producer contains 50 percent development and IT skills, 37.5 percent skills in creative and design work, and 12.5 percent skills from admin customer support. This taxonomy breakdown allows us not only to categorise novel jobs in an existing framework, which facilitates the matching of labour supply and demand, but also helps workers to navigate when reskilling towards a novel skill domain.

Figure 3a-c: Via the granular depiction of skill compositions, we can categorise new and relevant jobs into existing skill and occupation taxonomies
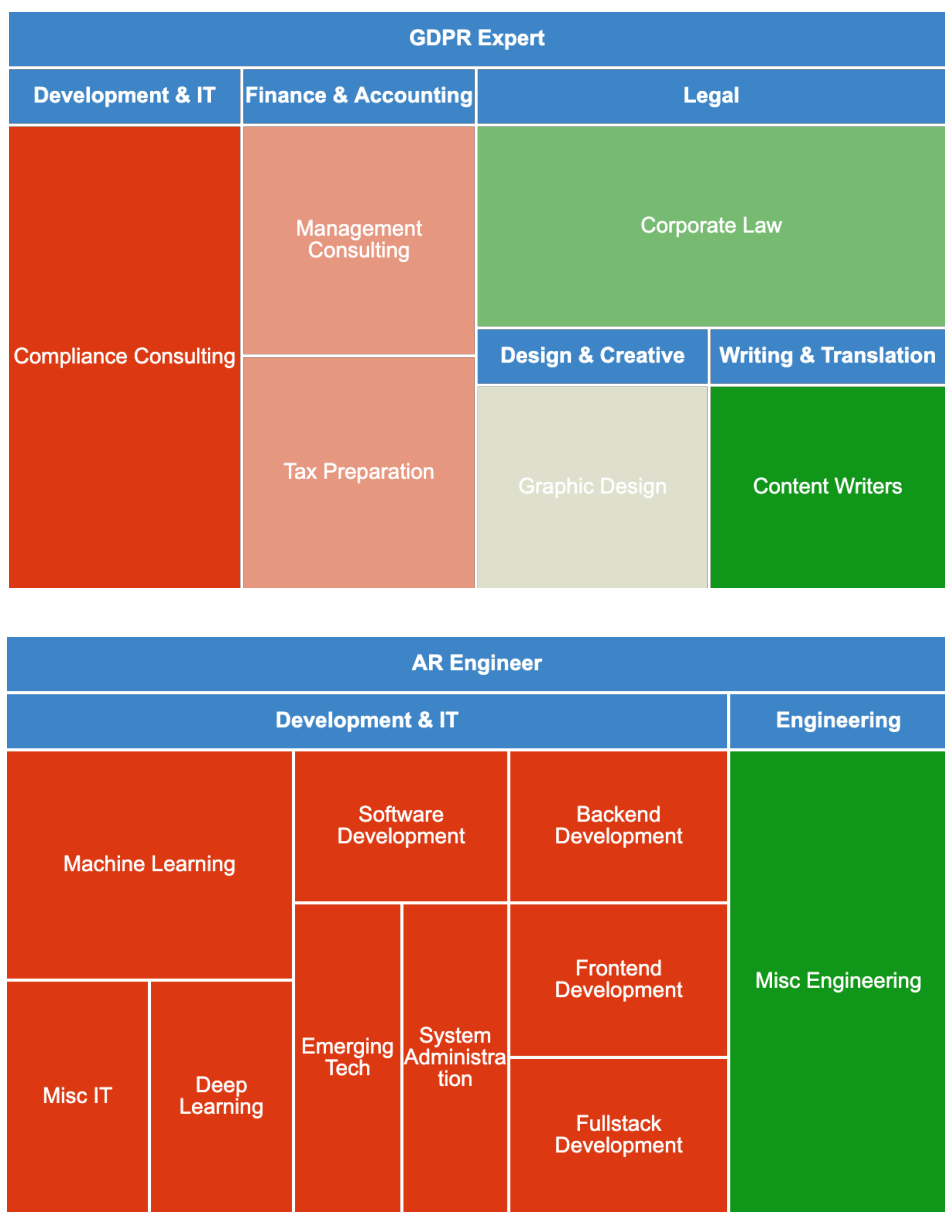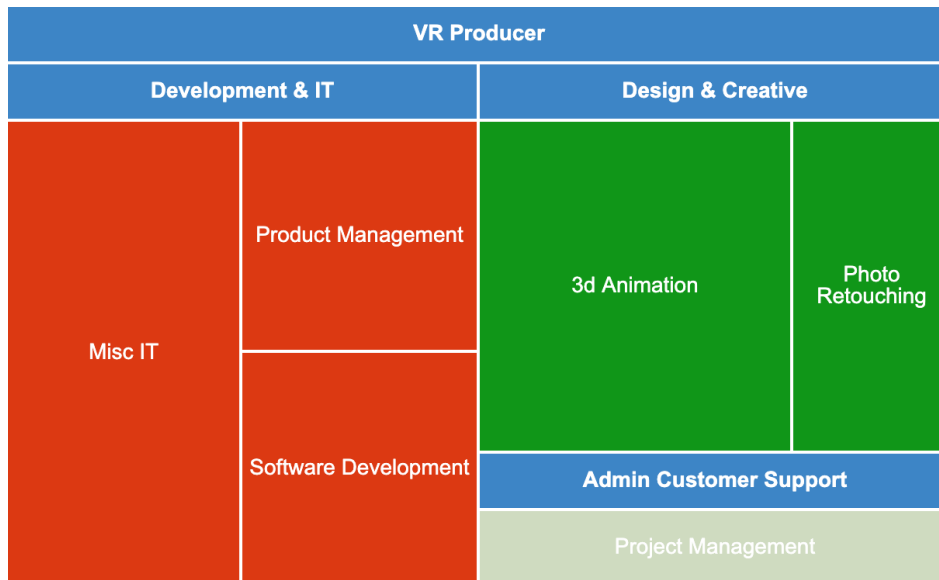
The granularity of online labour-market data also allows us to dissect the trends and skill compositions of newly emerging domains across time. Figure 4 showcases the demand for online freelance projects in the domain of GDPR (EU General Data Protection Regulation) work between 2017 and 2019. Here, we see that with the increasing legal relevance of GDPR configurations, online labour markets responded with an increase in demand, rapidly rising in early 2018 and peaking in May 2018 (with 22 times as many projects, as in the first month of the time series), the month when the EU made GDPR application mandatory in all member states. After this, the demand for GDPR work remained on a lower, yet, stable level.

Figure 4: The demand for online freelance services in the field of GDPR has grown with the entry into force of the regulation in all EU countries on 25 May 2018
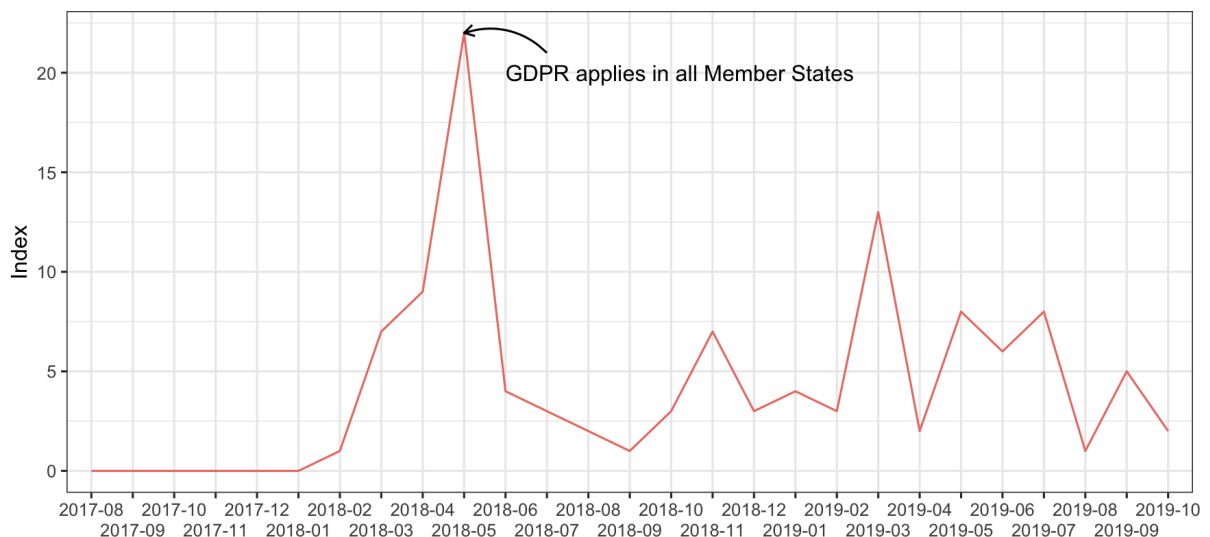
Besides tracking the quantity of labour demand in newly emerging domains, OLM data allows us to shed light on the hourly remuneration for this work. Figure 5 depicts the monthly project wages for work around GDPR, robotics, and virtual reality, as dollars per hour. These wages represent market evaluations of the work, as they are the final project remunerations agreed upon by project buyers and workers. We see that work around GDPR has a significantly higher level of wages than projects around VR and robotics. In addition, the most recent part of the time series indicates an upwards trend for project wages in the field of GDPR.

Figure 5: We observe significantly higher project wage levels for work around GDPR than in the new domains of robotics and VR; likewise, wages for GDPR work show an upwards trend
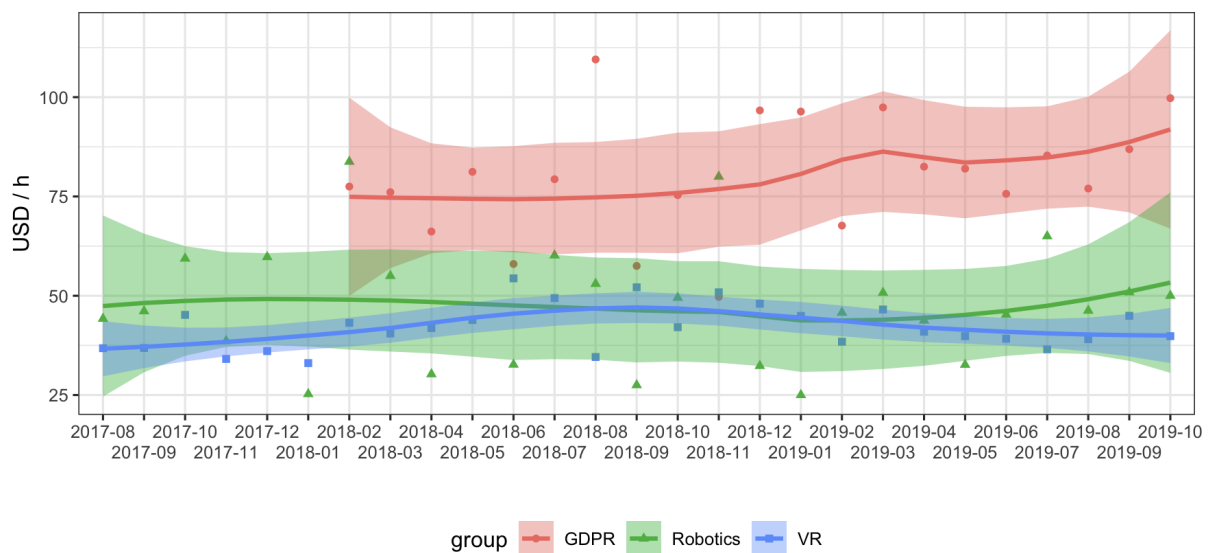

In particular for novel skill domains, skill mismatches are related to information asymmetries between jobseekers and firms on the specific content of work (de Mauro, 2018). Here, OLM data can likewise help to reduce asymmetries, as it allows a granular depiction of the skill domains related to a new field. For the case of robotics work, Figure 6 shows how skill domains are distributed. While the majority of skills, not surprisingly, stem from the field of engineering or web/mobile/software development, the second group appears to become more relevant in the last time period. This might indicate a shift in the skill content or online mediated work around robotics.

Figure 6: For the field of robotics, we are able to decompose online work by occupational categories, which reveals that the share of projects in web, mobile and software development is increasing.
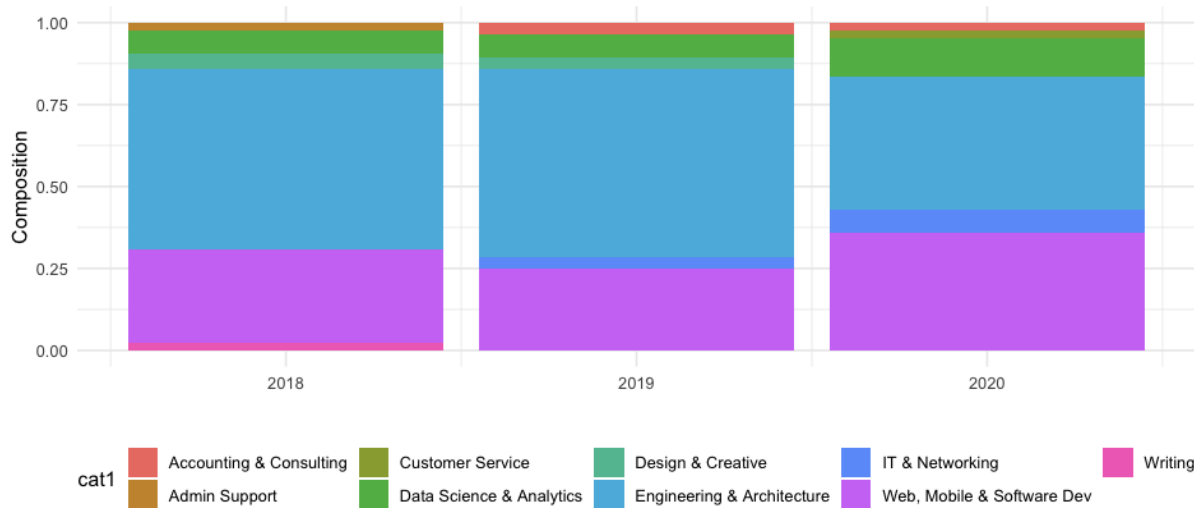
As we would like to further unravel the dynamics behind the rise of the web/mobile/software development skill domain illustrated in Figure 6, OLM allows us to dissect project profiles more granularly, on the level of individual skills. Figure 7 illustrates how the demand for individual skills and competencies around robotics changed from 2019 to 2020 within the domain of web/mobile/software development. In accordance with the general decline of engineering as a skill domain, we also see that skills from this domain are less likely to be requested in 2020 projects than a year previously. On the other hand, skills around programming and handling embedded systems (*pcd-design* or *embedded-systems*) or capacities for automation (*arduino, machine-learning, python*) have become more popular.

Figure 7: Within the domain of web, mobile, and software development, projects related to robotics are increasingly requesting skills for handling and programming embedded systems, while mechanical design and engineering was less demanded in 2020 than one year previously
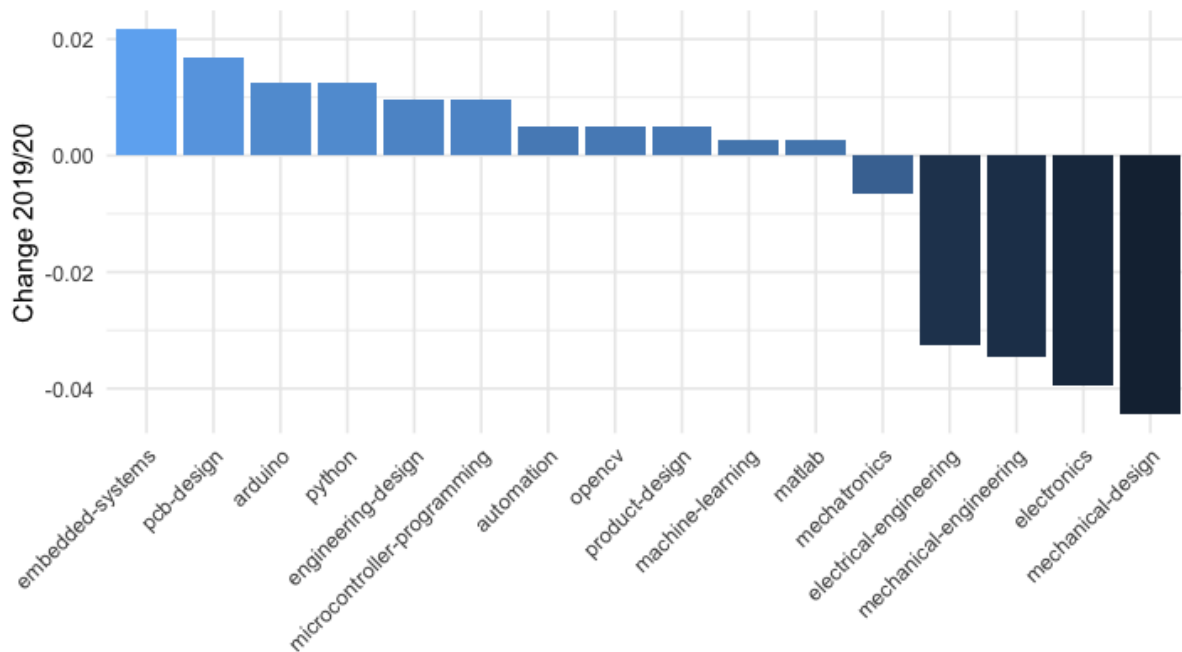
In perspective, these findings show that OLM data allows us to monitor both the demand for and the occupational or skill context of newly emerging phenomena, such as GDPR, robotics or VR.

To address the immediate need for systematic oversight of recent skill requirements of new professions, this granular and market data-driven perspective can be a first step to help designing farsighted and proactive retraining policies for in-demand skills. However, this approach can only deliver the data basis to building a truly 'future-proofed' workforce in the long run.

4 Building a 'future-proofed' workforce

Addressing near-term specific skill requirements

Future workforce development is not an easily-solved challenge, nor does it readily yield to a solution that will eradicate the problem in the long term. The harsh reality is that it is impossible to predict the future with any certainty; indeed we are psychologically averse to accepting the possibility of significant change (Quoidbach, 2013). With respect to identifying the *precise* nature of future workforce requirements, the best we can hope to achieve is to predict *specific* future skills requirements *in the very near term,* for which data can be collected, collated and analysed. As already identified in this paper,

the skills and expertise that combine to create emerging job profiles and skills in online labour platforms can be identified and tracked. There is therefore the potential to develop systems to ensure that training in specific skills stays ahead of immediate needs. However, further action is required to build more dynamic connections between emerging and evolving occupational analysis and the frameworks that underpin how people are educated, trained and re-trained.

Addressing the longer-term workforce requirements

A greater challenge than identifying specific skills as they emerge in the marketplace is preparing people for the requirements of the workforce in the longer term. There is a substantial and urgent need to devise a way to enable education and training systems to integrate and begin the teaching of the capabilities, aptitudes and associated skills that will better prepare people to thrive in a future workforce for which the precise skill requirements *cannot be accurately predicted, identified or specified.*

There is little clear agreement about exactly *how* to address the long-term workforce education and training challenge (see for example, Autor, 2019; Susskind and Susskind, 2015; Pew, 2021). However, there is a broad consensus about *what* the skills and abilities are that employers require of the future workforce. There have been numerous studies, a considerable amount of thinking and countless reports and policy briefs (see for example, WEF, 2020; McKinsey, 2021; and OECD, 2020) resulting in a broad consensus that employers want people with 'no-regrets' higher level skills and capabilities (PWC, 2020), including *critical thinking, creativity, communication, collaboration, problem solving, self-efficacy* and the *ability to learn* (see for example KPMG, 2018). In addition, employers want a workforce of lifelong learners who can easily pivot into new roles as and when required, quickly assimilating their new roles' specific skills.

However, currently, these 'no-regrets' higher level capabilities are not effectively provided or evidenced by any education sector in a manner that meets employers' needs. There is therefore clearly a substantial need to develop policies that will enable the integration of these future workforce skills and abilities into the curricula and pedagogies adopted by educational and training institutions. The size of this task should not be underestimated and is beyond the scope of this paper. However, there are some smaller steps that are tractable and can yield some useful prescience, understanding and progress. Just as the data collected by freelance recruitment platforms enables monitoring of both the demand for and the context of newly emerging occupations and their associated specific skills, so too data, already being collected by some educational organisations, can be used to monitor, understand and better support their development.

How data can help provide evidence of existing 'no-regrets' skill development

A fundamental aspect of any new approach to long-term workforce development through education and training system design must acknowledge the essential need to build within people the motivation and the ability to be extremely good at learning, re-learning, training and retraining, as often and as much as needed (Luckin, 2018). Learning to Learn (LTL), or Learning how to Learn, is just one of the 'no-regrets' higher level capabilities for the future workforce that employers require. For example, LTL was mentioned as one of the eight key competencies recommended for lifelong learning adopted by the Education Council and the European Parliament in December 2006 and revised in 2018 (Education Council, 2006 and 2018). There are many attempts to define LTL. For example, in one of the largest research projects on LTL, University of Helsinki Researchers developed a framework for assessing LTL, and defined learning to learn as:

 "the ability and willingness to adapt to novel tasks, activating one's commitment to thinking and the perspective of hope by means of maintaining one's cognitive and affective self-regulation in and of learning action" (Hautamäki *et al*, 2002, p. 39).

To explain the way in which a future workforce ability, such as LTL, can be identified within existing educational practices and therefore leveraged to better effect for the future, we provide a case study from research undertaken with a large US higher education institute (HEI) during 2021. Further information about this work can be found in Kent *et al* (2022). The research was precipitated by a request from the University's Provost, who wished to know if there was any evidence that students were developing the ability to learn as well as developing an understanding of the subject discipline of their degree.

Definition

The first step in seeking evidence of an ability, such as LTL, was to conduct a rapid evidence review and engage members of university staff in conversations to identify the sub-constructs of LTL of particular relevance to their context. This process resulted in agreement that LTL was conceptualised as a process that starts before the students arrive at the HEI, develops during their time studying and continues after graduation. Further refinement resulted in agreement between researchers and University staff about an initial definition: *Learning to Learn is a process of improvement in self-regulated learning (SRL). Self-regulated learners are "meta-cognitively, motivationally, and behaviourally active participants in their own learning process"* (Zimmerman,1989, p. 4).

15

Mapping LTL to evidence

The agreement about a definition was a key milestone, allowing attention to move to the ways in which evidence about students' LTL, according to this agreed definition, can be identified within the data that is available. Whilst a range of approaches has been adopted previously to study the proxies for LTL and its sub-constructs within data, the data available for the particular population being studied may not be amenable to the use of such proxies. For example, heart-rate variability (Spann *et al*, 2017), question-asking and engagement in online discussions (Pardo *et al*, 2016), the association of learning strategies with teachers' feedback (Matcha *et al*, 2019), and clickstream data from learning management systems (Cicchinelli *et al*, 2018; Motz *et al*, 2019), have all been used previously, but not all these data features may be available. In the case of the HEI being reported here, we were restricted to existing data sources for a cohort of students who were either studying a basic biology course (2,056 students) or a basic psychology course (1,895 students) during 2019 and 2020 (six terms in total). The data sources available were: academic performance, enrolment data and socio-demographic data (a record per student, course, year and term); assignment datasets from the University's learning management system (LMS) on a daily basis, and weekly reports of clickstream data, again from the LMS.

A further challenge to the limitations of the existing data is the lack of agreed semantics about the behavioural activity and complexities associated with LTL concepts. The key to addressing this challenge and connecting available data sources to the types of evidence required for evidencing LTL was the creation of an ontology that mapped the theoretical aspects of LTL onto the university's existing data sources to provide a framework within which an analytical methodology could be constructed. Specifically, the ontology focussed on the LTL sub-concept of self-regulated Learning (SRL) and specified the associations between theoretical constructs, such as reflection, and their possible operationalisations, such as in an email or work done on a group assignment. The ontology also made explicit that a theoretical construct might be operationalised in more than one way and that an observable behaviour might contribute to the operationalisation of more than one theoretical construct.

### Coping with changes in students' LTL over time

In addition to the basic mapping of theoretical concepts to data sources, the ontology can be used to capture the dynamic chronology of LTL as students develop through their studies. This dynamic chronology was represented by the concepts of a *snapshot: a time-stamped event capturing data about a student's behaviour*, and *a sequence of snapshots: the order and frequency of individual snapshots*. Furthermore, a coding system was devised to classify snapshots that provided some evidence of LTL,

and a rule base was created to identify significant correlations with well-established metrics, such as a student's Grade Point Average. The rule base enabled two engineered features: an LTL classification scheme with the types: Analysis, Assessment, Engagement, Feedback, Motivation, Practice, Reflection and Review; and a measure of the strength of the LTL evidence using a numerical scale from -3 to 3; negative values indicated evidence associated negatively with LTL. A positive value indicated evidence positively associated with LTL. This work enabled us to quantify how the LTL evidence strength increased or decreased across time. Process mining (Van Der Aalst, 2012), was used to model each student's LTL journey, operationalised by the student's sequences of snapshots. This was complemented by dimension reduction and clustering along with factor analysis to identify whether LTL features formed a unified dimension that was distinct from other, well-researched dimensions, such as academic performance and engagement. The dimensions identified were clustered (Trivedi *et al*, 2015) to identify a set of learner profiles (del Valle and Duffy, 2007).

***Learning from the data about students' development of LTL***

Principal components analysis (PCA) revealed three components consistent with the theoretical assumptions about LTL with strong loadings of *engagement* items on component 1, *grades* on component 2, and *LTL evidence* items on component 3.

Figure 8: Percentage of cases in each cluster; eight cases were not clustered due to missing values

Cluster 1
6,3%

Cluster 3
26,7%
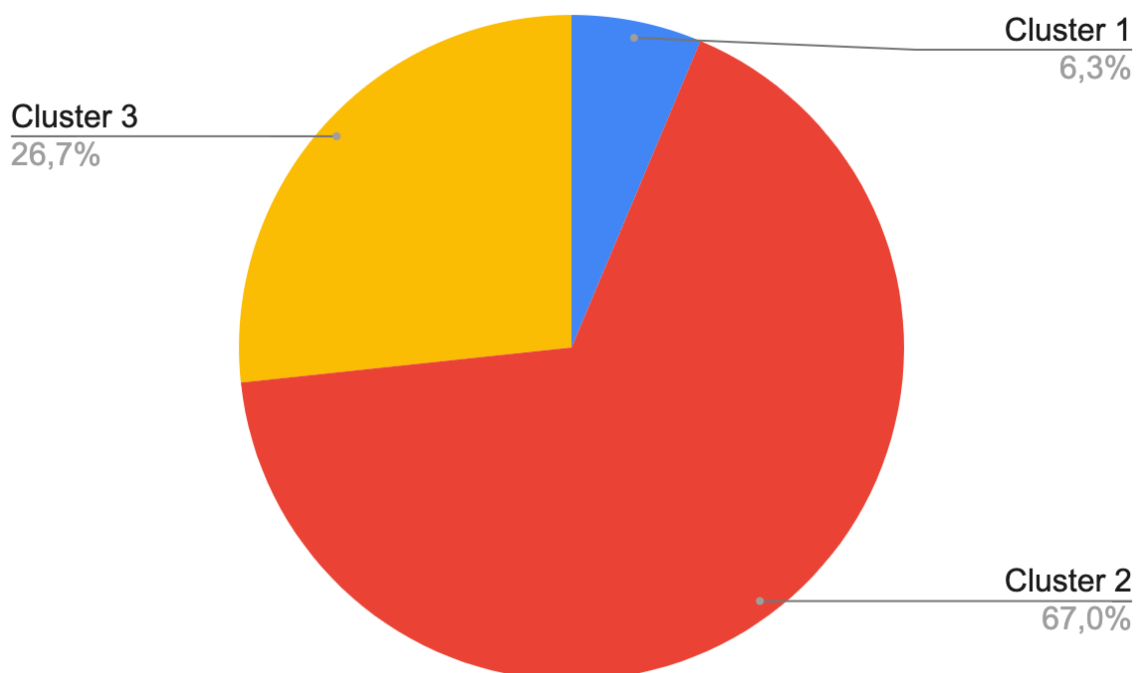
Cluster 2
67,0%

We then added the five LTL evidence variables: reflection, motivation, review, unclassified and analysis to the three components from the PCA. K-means cluster analysis (Lloyd, 1982) was used to identify three clusters of students: cluster 1 with 267 cases, cluster 2 with 2,818 cases, and cluster 3 with 1123 cases (Figure 8).

Figure 9: The three clusters using a final centres profiling: students in both clusters 1 & 2 had a higher level for LTL evidence than students in cluster 3
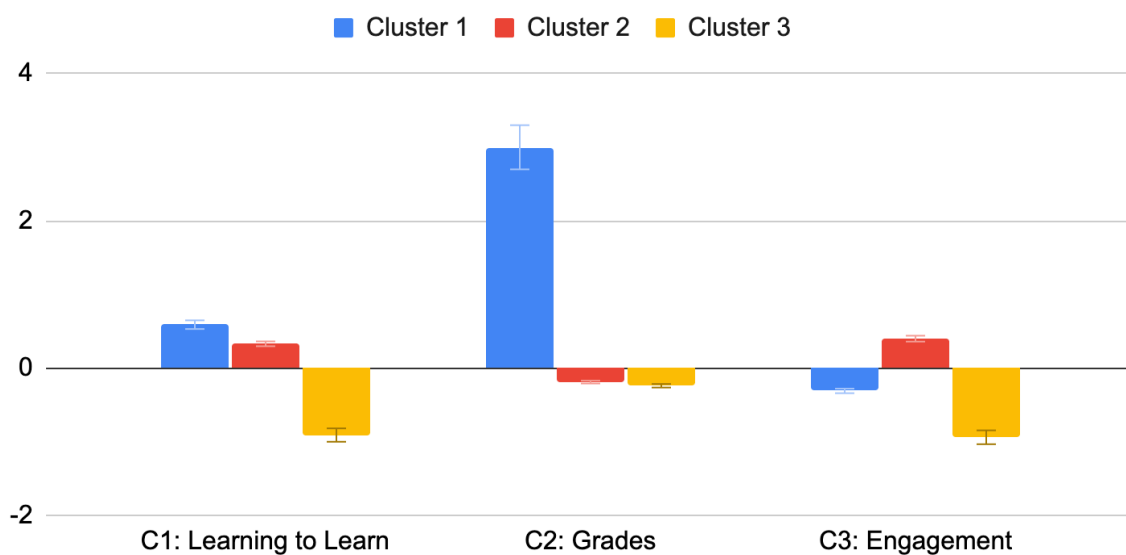
This analysis indicated that students in both clusters 1 and 2 had a higher level of LTL evidence than students in cluster 3. This difference was statistically significant. In addition, students in cluster 2 had a higher level of engagement than students in cluster 1 (this too, was statistically significant), despite achieving lower grades than students in cluster 1. In short, cluster 2 students had higher levels of engagement, and were more likely to transform their positive engagement levels into positive practice. However, this cluster of students was not able to transform this engagement into higher academic performance. Interestingly, the LTL evidence for students in cluster 2 was lower than the cluster 1 students, which might explain their difficulty in translating engagement into performance. In general, LTL levels were in a decline throughout the term, which could suggest the need for a half-term intervention to refresh and boost students' reflective capabilities. Process mining revealed that students in cluster 3 had the highest number of instances of negative engagement and subsequent negative assessment levels. It is interesting to note that this cluster also performed poorly academically and had low engagement levels.

This short case study illustrates that existing data can be used to explore the development of 'no-regrets' higher level capabilities and workplace skills, such as LTL, even when the data was not collected with the express purpose of tracking LTL. Such analysis and modelling could enable future training to be adjusted quickly to better meet the needs of students with respect to developing these longer term over-arching capabilities of value to employers. The approach described here could also be used to develop better data sets and more accurate analytical tools to help educators better address the needs of employers in advance of significant curriculum reform.

Validating the skills the workforce needs

A related challenge that impacts on both near-term skill development and longer-term workforce skill development is assessing both the precise skills that we are able to specify and the higher level capabilities that will be important for members of the future workforce. A move away from the current over-reliance on exams and tests that still dominates in many regions will be essential (Harris and Jones, 2021; Staton, 2021; Walker, 2021). Many current qualification mechanisms still take too long to create and validate. Consequently, even for the skills and labour requirements that can be predicted with some level of certainty as they emerge, there are still difficulties in validating an individuals' success or failure in reaching the required proficiency.

In addition, training and validating the knowledge and skills of teachers and trainers is a major challenge. This is not least because it requires a change in mindset that results in educators and trainers accepting an inevitable lifelong commitment to continually updating their knowledge and skills, in order to address the needs of the people who must be prepared for the roles and occupations that emerge and evolve. Educator training and professional development is rarely dynamic, as will now be required, and a change in culture and expectation in relation to training teachers and trainers will be required (Dover *et al*, 2019; Leask and Younie, 2013)

Future workforce models must therefore, out of necessity, be essentially fluid and evolving. This makes them rather like fog, hard to capture, incapable of full definition and amenable to the type of structured regimes common today within many education and training systems. The 'no-regrets' higher level capabilities are slightly less ephemeral and can, in principle, be captured and evidenced. However, even the most basic datasets required to harness the fledgling evidence about the development of these capabilities, so important to the workplace, are often unavailable, inconsistent and of poor quality. An essential first step is for all education and training leaders to recognise that they must pay greater attention to the consistent collection of data about learners as they progress through their studies,

beyond simple performance metrics. Those leaders and their teams must then engage in learning about what they can gain from such data and how they can use this to leverage valuable information to inform the way that they meet employers' needs as quickly as possible, rather than waiting for substantial education and training reform that may take some time. Perhaps most importantly, policymakers must whet the appetite of industry and educational leaders to work together to use the data available to them to change the way they prepare people for the workplace now, rather than waiting for clearer agreement on how the accepted all-important 'no-regrets' capabilities should be integrated into rubrics, curricular and pedagogy.

The willingness of policymakers, politicians, industry bosses and educational leaders to crush the existing inertia in attitudes to changing global circumstances and the need for transformational reform is probably the toughest challenge to address. A change of mindset is essential to move away from systems that have apparently served their purpose well for over a century, but which are no longer fit for purpose. This is also likely to lead to changes in levels and loci of control if economic and social prosperity is to be preserved (Brighouse, 2020).

5 Discussion and recommendations

The global workforce is under constant pressure to reskill, as technological change accelerates and new technologies reshuffle skill requirements. Labour-market mismatches must be avoided but traditional reskilling, via national education policies, is too slow for the fast pace of technological change while precise skill requirements for emerging new technologies are too fuzzy. In addition, economic tightening due to situations like the COVID-19 lockdown further accelerates digitalisation trends while forcing workers to reskill from home.

The skill-based occupation modelling of online labour-market data could help workers and firms to recognise newly emerging occupations in time and leverage their already existing capacities in reskilling towards these new types of jobs. Establishing online labour-market data as an additional resource for occupation foresight could help in matching adequate talent with job vacancies and in developing far-sighted reskilling suggestions in a future of technological disruption. The willingness of organisations to use existing educational and training data to learn where and how they are already developing the capabilities and skills the workforce will require in the longer term is essential. As is their willingness to make small, but significant changes to their practice and organisation in accordance with what they learn from this data, to further enhance the development of these capabilities. The collection

of further data that is more closely aligned with these agreed workplace capabilities would of course enrich what can be learnt and boost how existing courses, practices and organisation can be improved.

Accordingly, the European Commission's 2022 Data Act proposal[3] identified the importance of and difficulty in accessing business (and platform) data for public sector bodies, while acknowledging the protection of businesses' interests. In detail, Article 15 of the proposed Data Act provides the legal basis for making private sector data available on the basis of a *"public emergency"*. However, establishing a case of 'public emergency' is a high hurdle for potential data recipients. The drastic repercussions of climate change and the outbreak of a pandemic might qualify as a public emergency in the spirit of Article 15, but do rising economic inequalities or growing skill mismatches? 'Public interest', rather than 'public emergency' could be the focal point of Article 15. Public bodies acting in the interest of the public should qualify as data recipients with the mandate to reach out to private data holders. To further ease data access in the interest of the public, future amendments to the Data Act should include data retrieval via web-scraping – an automated form of data extraction from websites. Web-scraping is a valuable option if data is publicly available and platform providers are too slow, unwilling or technically unable to share data with public sector recipients in the interest of the Data Act. At the same time, data recipients are required to assure that neither the business interest of the platform nor privacy of the respective platform users, according to GDPR, are at risk.

In the long run, however, training and education for each specific predicted workforce expertise must be provided before the need within the workforce becomes critical. To ensure that specific training stays ahead of need, a better dynamic connection must be made between emerging and evolving occupational analysis and the frameworks that underpin what and how young people are taught and trained.

---

[3] Available at https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52022PC0068.

References

Acemoglu, D. and D. Autor (2011) 'Skills, tasks and technologies: Implications for employment and earnings', in *Handbook of Labor Economics* (Vol. 4, pp. 1043-1171), Elsevier

Anderson, K.A. (2017) 'Skill networks and measures of complex human capital', *Proceedings of the National Academy of Sciences*, *114*(48): 12720-12724

Autor, D. (2015) 'Why are there still so many jobs? The history and future of workplace automation', *Journal of Economic Perspectives* 29, no. 3: 3-30

Autor, D. (2019) 'Work of the past, work of the future', *Working Paper* No. 25588, National Bureau of Economic Research

Bastian, M., M. Hayes, W. Vaughan, S. Shah, P. Skomoroch, H. Kim and C. Lloyd (2014) 'Linkedin skills: large-scale topic extraction and inference', in *Proceedings of the 8th ACM Conference on Recommender Systems* (pp. 1-8)

Brunello, G. and L. Rocco (2017) 'The effects of vocational education on adult skills, employment and wages: What can we learn from PIAAC?', *SERIEs* 8(4): 315-343

Brunello, G. and P. Wruuck (2019) 'Skill shortages and skill mismatch in Europe: A review of the literature', *EIB Working Papers* 2019/05, European Investment Bank

Calanca, F., L. Sayfullina, L. Minkus, C. Wagner and E. Malmi (2019) 'Responsible team players wanted: an analysis of soft skill requirements in job advertisements', *EPJ Data Science*, 8(1): 1-20

Cappelli, P. (2014) 'Skill gaps, skill shortages and skill mismatches: Evidence for the US', *Working Paper* No. 20382, National Bureau of Economic Research.

Card, D. and J.E. DiNardo (2002) 'Skill-biased technological change and rising wage inequality: Some problems and puzzles', *Journal of Labor Economics*, 20(4): 733-783

Cicchinelli, A., E. Veas, A. Pardo, V. Pammer-Schindler, A. Fessl, C. Barreiros and S. Lindstädt (2018) 'Finding traces of self-regulated learning in activity streams', in *Proceedings of the 8th International conference on learning analytics and knowledge* (LAK) (pp. 191-200)

Coolahan, J., S. Drudy, P. Hogan and S. McGuiness (2017) *Towards a better future: A review of the Irish school system*, Irish Primary Principals Network and the National Association of Principals

De Mauro, A., M. Greco, M. Grimaldi and P. Ritala (2018) 'Human resources for Big Data professions: A systematic classification of job roles and required skill sets', *Information Processing & Management* 54(5): 807-817

del Valle, R. and T.M. Duffy (2007) 'Online learning: Learner characteristics and their approaches to managing learning', *Instructional Science* 37(2): 129–149

Dover A., N. Henning, R. Agarwal-Rangnath and E. Dotson (2018) 'It's Heart Work: Critical Case Studies, Critical Professional Development, and Fostering Hope Among Social Justice–Oriented Teacher Educators', *Multicultural Perspectives* 20(4): 229-239, DOI: 10.1080/15210960.2018.1527154

Education Council (2006) 'Recommendation of the European Parliament and the Council of 18 December 2006 on key competencies for lifelong learning', Brussels: Official Journal of the European Union

Education Council (2018) 'Council Recommendation on key competences for lifelong learning', ST/9009/2018/INIT, 1–13

Escobari, M., I. Seyal and M. Meaney (2019) *Realism about Reskilling: Upgrading the Career Prospects of America's Low-Wage Workers*, Workforce of the Future Initiative, Center for Universal Education at The Brookings Institution

European Investment Bank (2018) *Retooling Europe's Economy*, Luxembourg

Harris, A. and M. Jones (2021) *Leading in disruptive times: a spotlight on assessment*, School Leadership & Management, DOI: 10.1080/13632434.2021.1887643

Hautamäki, J., P. Arinen, S. Eronen, A. Hautamäki, S. Kupiainen, B. Lindblom, M. Niemivirta, L. Pakaslahti, P. Rantanen and P. Scheinin (2002) *Assessing learning-to-learn: a framework*, Helsinki, National Board of Education

Horton, J.J. (2010) 'Online Labor Markets', in A. Saberi (ed) *Internet and Network Economics* (pp. 515–522), Springer: Berlin Heidelberg

Huber, S.G. and C. Helm (2020) 'COVID-19 and Schooling: Evaluation, Assessment and Accountability in Times of Crises—Reacting Quickly to Explore Key Issues for Policy, Practice and Research with the School Barometer', *Education Assessment Evaluation and Accountability* 32: 237–270, https://doi.org/10.1007/s11092-020-09322-y

Jain, A. and R. Sushant (2020) 'Implications of emerging technologies on the future of work', *IIMB Management Review*, Volume 32, Issue 4: 448-454, https://doi.org/10.1016/j.iimb.2020.11.004

Illanes, P., S. Lund, M. Mourshed, S. Rutherford and M. Tyreman (2018) *Retraining and reskilling workers in the age of automation*, McKinsey Global Institute

Kent, C., A. Akanji, B. du Boulay, I. Bashir, T. Fikes, S.A. Rodríguez De Jesús ... R. Luckin (2022) 'Mind the Gap: From Typical LMS Traces to Learning to Learn Journeys', in G. Trajkovski, M. Demeter and H. Hayes (eds) *Applying Data Science and Learning Analytics Throughout a Learner's Lifespan*, IGI Global

KPMG (2018) *Learning to Learn: The Future-Proof Skill*, available at https://assets.kpmg/content/dam/kpmg/uk/pdf/2018/10/learning-to-learn_report.pdf

Krueger, D. and K.B. Kumar (2004) 'Skill-specific rather than general education: A reason for US–Europe growth differences?' *Journal of Economic Growth* 9(2): 167-207

Leask, M. and S. Younie (2013) 'National models for continuing professional development: the challenges of twenty-first-century knowledge management', *Professional Development in Education* 39(2): 273-287, DOI: 10.1080/19415257.2012.749801

Lehdonvirta, V., A. Margaryan and H.U.W. Davies (2019) *Skills formation and skills matching in online platform work: policies and practices for promoting crowdworkers' continuous learning*, CrowdLearn

Luckin, R. (2018) *Machine Learning and Human Intelligence The Future of Education for the 21st Century*, University College London Institute of Education Press

Matcha, W., D. Gasevic, N.A.A. Uzir, J. Jovanovic and A. Pardo (2019) 'Analytics of Learning Strategies: Associations with Academic Performance and Feedback', *LAK19: Proceedings of the 9th International Conference on Learning Analytics & Knowledge*, https://doi.org/10.1145/3303772.3303787

Motz, B., J. Quick, N. Schroeder, J. Zook and M. Gunkel (2019) 'The validity and utility of activity logs as a measure of student engagement', *LAK19: Proceedings of the 9th International Conference on Learning Analytics & Knowledge*, https://doi.org/10.1145/3303772.3303789

McKinsey (2021) 'Defining the skills citizens will need in the future world of work', available at https://www.mckinsey.com/industries/public-and-social-sector/our-insights/defining-the-skills-citizens-will-need-in-the-future-world-of-work

Nedelkoska, L. and G. Quintini (2018) 'Automation, skills use and training', *OECD Social, Employment and Migration Working Papers* No. 202, OECD Publishing, Paris, https://doi.org/10.1787/2e2f4eea-en

OECD (2020) *The future of work*, available at https://www.oecd.org/future-of-work/

Pardo, A., F. Han and R.A. Ellis (2016) 'Exploring the relation between self-regulation, online activities, and academic performance: A case study', in *Proceedings of the 6th International Conference on Learning Analytics & Knowledge*

Pew Research Center (2021) 'Experts Say the 'New Normal' in 2025 Will Be Far More Tech-Driven, Presenting More Big Challenges', available at https://www.pewresearch.org/internet/2021/02/18/experts-say-the-new-normal-in-2025-will-be-far-more-tech-driven-presenting-more-big-challenges/

Pouliakas, K. (2021) 'Understanding Technological Change and Skill Needs: Technology and Skills Foresight', *Cedefop Practical Guide 3*, Cedefop-European Centre for the Development of Vocational Training

PWC (2020) *Upskilling: Building confidence in an uncertain world*, available at https://www.pwc.com/gx/en/ceo-survey/2020/trends/pwc-talent-trends-2020.pdf

Quoidbach, J., D. Gilbert and T. Wildon (2013) 'The End of History Illusion', *Science*, 4 January: 96-98

Saton, B. (2021) 'Educators around world seek to take an axe to exam-based learning', *Financial Times*, available at https://www.ft.com/content/9d64e479-182c-4dbd-96fe-0c26272a5875

Schmidpeter, B. and R. Winter-Ebmer (2018) 'How do automation and offshorability influence unemployment duration and subsequent job quality?' *IZA Discussion Paper* 11736, IZA Institute of Labor Economics

Spann, C.A., J. Schaeffer and G. Siemens (2017) 'Expanding the scope of learning analytics data: Preliminary findings on attention and self-regulation using wearable technology', in *Proceedings of the 7th international learning analytics & knowledge conference* (LAK) (pp. 203-207)

Stephany, F. (2021) 'One size does not fit all: Constructing complementary digital reskilling strategies using online labour market data', *Big Data & Society*, 8(1), https://doi.org/10.1177/20539517211003120

Stephany, F., M. Dunn, S. Sawyer and V. Lehdonvirta (2020) 'Distancing Bonus Or Downscaling Loss? The Changing Livelihood of Us Online Workers in Times of COVID-19', *Tijdschrift voor economische en sociale geografie* 111(3): 561-573

Stephany, F., O. Kässi, U. Rani and V. Lehdonvirta (2021) 'Online Labour Index 2020: New ways to measure the world's remote freelancing market', *Big Data & Society*, https://doi.org/10.1177/20539517211043240

Stephany, F., N. Stoehr, P. Darius, L. Neuhäuser, O. Teutloff and F. Braesemann (2020) 'The CoRisk-Index: A data-mining approach to identify industry-specific risk assessments related to COVID-19 in real-time', *arXiv* preprint arXiv:2003.12432

Susskind, R.E. and D. Susskind (2015) *The future of the professions: How technology will transform the work of human experts*, Oxford University Press, New York

Tamm, M. (2018) 'Training and changes in job tasks', *Economics of Education Review* 67: 137-147

Trivedi, S., Z.A. Pards and N.T. Heffernan (2015) 'The utility of clustering in prediction tasks', *arXiv* (September): 1–11

Walker, M. (2021) 'The future of qualifications and assessment', *CIEA Blog*, available at https://www.herts.ac.uk/ciea/articles/the-future-of-qualifications-and-assessment

WEF (2020) 'These are the top 10 job skills of tomorrow — and how long it takes to learn them', available at https://www.weforum.org/agenda/2020/10/top-10-work-skills-of-tomorrow-how-long-it-takes-to-learn-them

Zimmerman, B.J. (1989) 'A social cognitive view of self-regulated academic learning', *Journal of Educational Psychology* 81: 329-339