What is the Price of a Skill?
Revealing the Complementary Value of Skills

Version September 2022

Fabian Stephany✦ ✤ * 📧, Ole Teutloff♥*, Vili Lehdonvirta✦

✦ Oxford Internet Institute, University of Oxford, Oxford OX1 3JS

✤ Humboldt Institute for Internet and Society, Berlin

♥ University of Copenhagen · Centre for Social Data Science

** These authors contributed equally to the work.*

📧 *fabian.stephany@oii.ox.ac.uk*

Abstract

The global workforce is urged to constantly reskill, as technological change favours particular new skills while making others redundant. But which skills are most marketable and have a sustainable demand? We propose a model for skill evaluation that attaches a premium to a skill based on near real-time online labour market data. The model allows us to isolate the economic return of an individual skill measured as a premium on hourly wages. We demonstrate that the value of a specific skill is strongly determined by complementarity - that is with how many other high-value skills a competency can be combined. We introduce the idea of "hub skills" to the field of human capital formation, i.e., high-return skills that can be recombined with many valuable complements. Specifically, we show that the value of a skill is relative, as it depends on the capacities it is combined with. For most skills, their value is highest when used in combination with skills of the same type. In addition, we find that supply and demand and the membership in specific skill communities, such as finance and legal or software and development, determine the value of a skill. We illustrate that AI skills are hub skills, as they can be combined with other high-value skills to generate beneficial complementarities. The value of some of these in-demand skills has increased significantly over the last years. Furthermore, we contrast our skill premia with automation probabilities and find that some skills are very susceptible to automation despite their high economic value. The model and metrics of our work can inform digital re-skilling to reduce labour market mismatches. In cooperation with online platforms and education providers, researchers and policy makers should consider using this blueprint to provide learners with personalised skill recommendations that complement their existing capacities and fit their occupational background.

Highlights:

- We expand literature on human capital formation by quantifying the economic value of skills.
- We demonstrate that the value of a skill depends on complementarity; by the number and value of skills it can be combined with.
- We reveal that the value of a skill is relative; it depends on the type of skills it is combined with.
- We showcase that AI skills are most valuable as they have high levels of skill complementarity.

Introduction

Technological change is not "skill-neutral". Innovations create new jobs but these jobs are characterised by new tasks that require new sets of skills. As a result, the landscape of skills and occupations changes. Despite recurring fears of mass unemployment, current literature suggests that firms are using technologies to automate specific tasks rather than entire occupations (Autor, 2015). Some occupations will disappear but the remaining ones will change, and entirely new jobs will emerge (Acemoglu & Autor, 2011; Brynjolfsson & Mitchell, 2017; Frey & Osborne, 2017). The work that is eliminated has different skill requirements than the newly created jobs, resulting in the paradoxical situation of simultaneous unemployment and labour shortage (Autor, 2015). In other words, workers risk being pushed out of employment while companies struggle to find suitable employees to pursue innovative activities. To stay in employment, workers need to learn new skills and combine existing skills in new ways. To stay competitive, employers need to invest in reskilling and talent acquisition. However, for many of the newly emerging jobs, precise skill requirements are unclear and constantly evolving. How can policy makers, businesses, and individuals decide which skills to invest in?

In this paper, we introduce a method to measure the economic value of skills in human capital formation. We propose that the value of a skill is, mostly, determined by its complementarity, that is, by how easily it can be combined with other skills of high value. The value of one and the same skill depends on the type of its complements. We show that skills are most valuable when applied in combination with other similar skills. We test our assumption with the use of rich and near-real-time online labour market data. We do this against the background of widespread uncertainty among policy-makers, employers, and workers of how to address the growing skill mismatch and resulting labour market inefficiencies. The conventional policy response – aligning training programmes with changing labour market demand – is becoming increasingly ineffectual as technological and social transformation outpaces national training systems (Collins & Halverson, 2018). Likewise, large employers struggle to keep the skills of their workforce up to date (Illanes et al., 2018). Employers, workers and education providers seem uncertain about which new, often digital, skill is the first step towards a successful re-skilling trajectory. To provide reliable information on which skills are most marketable and have a sustainable demand, we propose an evaluation of skills based on complementarity as core features that determine a skill's economic value.

To this end, we use data from one of the most popular online freelancing platforms. The online labour market (short OLM) data on platform transactions ("projects" or "gigs") contains skill requirements and price information. Using these information and methods from network science we construct a network of co-occurring skills. This skill network provides us with an endogenous categorisation of skills and illustrates the context dependency of human capital. The monetary value of freelance projects (in USD per hour) allows us to statistically assess the economic value of individual skills. We use various analyses to derive the monetary value of the 962 most popular skills. Moreover, we build a regression model to test our assumptions about what drives the variance in skill premia. To illustrate our approach we visualise and explain the value of skills that are frequently used in the domain of AI work. Using this example, we show how the value of a specific skill depends on supply and demand, skill community, and - most importantly - complementarity. We show why "AI skills" are particularly valuable, as they are frequently combined with other high-value skills. We contrast our skill premia with automation probabilities and show that some skills are very susceptible to automation despite their high economic value.

Our paper adds to four related streams of literature. Firstly, we contribute to an established debate on how to measure human capital. We attach an interpretable economic price to individual skills and we highlight the relevance of complementarity for a skill's value. Secondly, our work relates to scholarly work on how technological change and automation change the demand for skills. Here, our work suggests a pragmatic approach to near real-time monitoring the demand and value of individual skills on fast-changing labour markets. Thirdly, we contribute to the growing body of research that uses digital trace data to study labour market developments. In contrast to recently suggested data sources that examine either the market's supply or demand side exclusively, our approach takes both demand and supply into account and, crucially, allows us to attach price tags to individual skills. Lastly, we build on existing attempts to model the complex relationship between skills using network science. We show that a skill's network centrality captures its complementarity which is strongly predictive of its economic value. To guide our empirical analysis, we synthesise these scholarly debates into three hypotheses about what determines the value of a skill.

Our findings are applicable in several ways. They allow us to identify what skills and what combination of skills are in demand and successful (as hourly wage premium). Conceptualising the relationship between skills as a network, enables us to identify the most valuable complement to an existing skill bundle. Thereby, we could support workers in building individual, data-driven reskilling trajectories. We illustrate our method to skill evaluation with online freelancing data. However, the method can be extended to other data sources, such as online job ads or profiles from career building webpages, that cover larger parts of the traditional labour market.

Background

Human Capital Formation

The conceptualisation of human capital is one of the most prominently discussed topics in economic sociology and labour economics for the last decades, as novel concepts and metrics have emerged constantly (Angrist et al., 2019). Conventional measures of human capital often rely on the count of years of experience, training, or education or divide workers into categories, e.g., of labourers and management (Willis, 1986). However, a growing body of literature suggests that years of training and broad worker categories fail to address the importance of skill specialisation, diversity, and recombination in knowledge generation (Aggarwal & Woolley, 2013; Hong & Page, 2004; Lazear, 2004; Ren & Argote, 2011; Woolley et al., 2010). In addition, the rise of the knowledge economy (Powell & Snellman, 2004) has sparked new interest in a more nuanced measure of skill composition. In this context, several papers have taken skill diversity and individual cognitive abilities into account to estimate their effect on wages (Altonji, 2010; Autor & Handel, 2013; Borghans et al., 2008; Bowles et al., 2001; Heckman et al., 2006; Ingram & Neumann, 2006).

A central conclusion of past contributions on skill diversity is that the relationship between wages and skills does not only depend on a worker's individual skills but also, on how they are combined. The question of skill complementarities arises (Allinson & Hayes, 1996). For some skills (e.g., programming in JavaScript and visualisation techniques) it can be argued that skill complementarities emerge. The bundle of skills is more valuable than the sum of its parts. Oftentimes, skill complementarities are likely to be limited to one occupational domain, e.g., programming in python and translating Russian should have little complementarities. Certainly, the value of additional skills depends on the skill portfolio that the worker already possesses (Altonji, 2010). Although the existing literature recognizes the importance of skill complementarities, there have been little attempts to model it formally at the level of skills. One notable exception is Neffke (2019) who investigates the value of complementary co-workers. Modelling skill complementarities between workers shows that the value of what a person knows depends on the skillsets of whom they work with (Neffke, 2019).

Instead, our work focuses on the complementarity between individual skills. We model this complex relationship using network science. We observe that skills of similar economic value cluster together. Our results underline the importance of complementarity. Skills that are frequently combined with a specific set of other skills tend to be of higher economic value. Moreover, skills that match the overall skill background of workers are of higher value to them then skills that are distant to their skill domain. This illustrates that the economic value of a skill is relative as it depends on a worker's existing skill bundle.

Automation and Technological Change

Modelling human capital at the levels of skills represents a promising avenue to analyse the impact of technological change on labour markets. The periodic warning that automation and new technologies are going to replace large numbers of jobs is a recurring theme in economic literature (Acemoglu & Autor, 2011; Brynjolfsson & McAfee, 2014; Frey & Osborne, 2017). However, measuring the effects of new technologies on the future of work is not trivial. Looking at education and wages at the aggregate level has proven to be insufficient (Acemoglu & Restrepo, 2018; Beaudry et al., 2016; Brynjolfsson & McAfee, 2014). Treating occupations as homogenous entities with a certain automation probability risks being overly simplistic and could even lead to false conclusions (Frank et al., 2019). Skill requirements of occupations are dynamic because technological innovations change the demand for specific skills and thereby the skill composition of occupations. Aggregating skills into occupations therefore obfuscates the differential impact of new technologies (Frank et al., 2019). Acemoglu & Autor (2011) propose a task-based model in which occupations are classified into routine or nonroutine and physical or cognitive based on their task content. This approach proved to be powerful in explaining the hollowing out of middle-class jobs.

Yet, even this level of aggregation might still be too coarse. Disaggregating to the level of specific skills might be necessary to understand technology's complex impact on labour markets. If we conceptualise occupations as dynamic containers of skills and machines as replacing individual elements within these containers, then we need to move the analysis to the level of skills to fully grasp how technology is redefining the demand for human skills. However, reliable data on workplace skills is not widely available. In fact, Frank et al. (2019) identify sparse skills data as the number one barrier to forecasting the future of work. Recent studies (e.g. Acemoglu & Autor, 2011; Alabdulkareem et al., 2018; Frey & Osborne, 2017) rely on survey-based data such as O*NET for the US (or PIAAC for the OECD). While valuable in many ways, O*NET data is relatively static and therefore not ideal to study a fast changing economy (Frank et al., 2019). The skills included in O*NET and the corresponding skill taxonomy tend to be outpaced by the dynamic frontier of technological innovation time and again.

We propose a data source and methodology to help overcome the roadblock of sparse skills data. Using project-level skill information from online labour platforms, we study skills at their most granular level and at almost real-time. The actuality of our data allows us to monitor ongoing changes in the demand for and the value of skills. Together with our historical time series data, we are able to determine whether the value of a skill is on the rise or in decline. Moreover, applying tools from network science, we build a dynamic, bottom-up and data generated classification of skills based on how they are used. Thereby, we can observe how the relationship between skills changes as they are being recombined in new ways. Lastly, we break down the occupation-level automation probability by Frey & Osborne (2017) to the level of specific skills and illustrate how automation risk relates to the economic value of skills. Taken together, these elements could help policy makers, education providers and workers identify skills that are worth investing in during times of fast-changing labour market.

Understanding Labour Markets with Digital Trace Data

With the rise of digital platforms as economic intermediaries more and more data about labour market processes has become available online. As automated retrieval of large amounts of this economic transaction data has become more feasible in recent years, the study of labour markets with digital trace data is gaining momentum. Besides relying on traditional survey data, researchers have started to investigate skill developments with the new digital data sources.

While the proposed idea of using online labour market data for skill monitoring is novel, several scholars have explored alternative sources of online generated data to investigate skill formation. De Mauro et al. (2018), for example, examined the skill complexity of the new profession of data science with data retrieved from various job boards. Similarly, both Börner et al. (2018) and Calanca et al. (2019) demonstrate the increasing relevance of soft skills based on online job vacancy data. Bastian et al. (2014), on the other hand, make 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. The methodology that we propose in this paper is applicable beyond OLM data. Our method could be used to analyse data from online job vacancy portals, such as Indeed or Glassdoor, or data from professional social network sites, like LinkedIn, in order to study skill development and detect the emergence of novel occupational domains. However, the different data sources have particular advantages and shortcomings, as summarised in Table 1.

Data Source	Demand Side	Supply Side	Price Information	Broad Coverage
Online Job Vacancies	✓	✗	?	✓
Networking Sites	✗	✓	✗	?
Online Labour Platforms	✓	✓	✓	✗

Table 1 *Compared to data from online job vacancy sites and career portals, online labour market data allow for the study of both the demand and supply side of work, including relevant information on prices. Source: Stephany & Luckin (2022).*

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 on the possible supply in the targeted population. Data from professional social media sites, like LinkedIn, on the other hand, allow for an in-depth analysis of skill compositions in the population. However, no price or income information is revealed, and matching efficiencies cannot be evaluated in the absence of demand side data. OLM data, e.g. from platforms like UpWork or Fiverr, only cover a small segment of the labour market, namely digitised tasks from jobs in the professional service sector. However, OLM data have the major advantage of containing information on both the demand and supply side of skills. In addition, it is possible to observe the matching process and price, e.g., hourly rate, for each job with a particular skill bundle attached to it.

The Network Value of Skills

The availability of novel data sources representing manifold aspects of interconnectedness have enabled complexity scientists to capture some of the network-related value of skills. Waters & Shutters (2022) examine an U.S. skill network, focusing on the relationship between network centrality and economic performance. They find that occupations with higher skill centrality are associated with greater annual salaries, and metropolitan areas with higher skill centrality have higher productivity rates. Overall, these results suggest that the application of traditional network metrics to this view of cities as complex systems can offer new insights into the dynamics of regional economies. When it comes to the complementary value of a skill, Dave et al. (2018) show that the relative positioning of a skill in a skill network contains information on how likely it is for a capacity to be acquired. In their simulation of skill acquisition they show that it is most likely for learners to acquire a skill that is close to the bundle of their already existing capacities.

Similarly, del Rio-Chanona et al. (2021) use network metrics to track worker transitions between occupations due to automation shocks. They show that transitions between occupations are more likely if skill similarities are high. This leads to the assumption that certain "bridging" skills are particularly valuable, as they enable for a transition from one (less profitable) occupation to the next. This complementary aspect of skills is again outlined by Alabdulkareem et al. (2018) who argue that much of the polarisation across industries and occupations can be explained by skill polarisation, e.g., groups of workers that are "stuck" in a low value segment of the skill network with little proximity to central skills that would allow them to shift into more profitable occupations.

While all of the above mentioned novel contributions highlight the importance of understanding the complexity of skills via network methods, they do not explicitly seek to measure and explain the economic value of a skill. Our contribution is to return to the assumption that a skill's network centrality matters, as it captures its complementarity, which should be reflected by its economic value. Moreover, modelling the relationship between skills as a network allows us to map skills by their similarity (with skills being close-by in the network being more similar). Via the network structure, we can categorise skills and workers (based on their skill profile) into communities (also referred to as domains). This can be used to approximate how close or distant a given skill is to the skill bunde of a particular worker.

What Determines the Value of a Skill?

As a theoretical contribution, our work aims to explain the economic value of a skill. Based on the theoretical considerations outlined in the previous sections, we argue that three aspects mainly influence the economic value of a skill:

1) **Supply and demand**. Our assumption rests on an intuitive understanding of market dynamics. We perceive the value of a skill as a price, which is conventionally determined by forces of supply and demand. We expect a skill that is often requested and seldom offered by workers to be of high value.
2) **Community**. Past literature has pointed towards a premium on belonging to a specific community of skills (Weeden, 2002). The community membership is an approximation for signalling various aspects of human capital, e.g., software or legal skills are prestigious and difficult to learn. This should lead to systematic level differences in skill values across communities of skills.

3) **Complementarity**. Lastly, reviewing the network science literature on skills, we assume that a skill's economic value is determined by its complementarity. In other words, how likely it is for a certain skill to be combined with other (high value) competencies. Using a network-based approach (Anderson, 2017; Dave et al., 2018; Waters & Shutters, 2022), we measure a skill's complementarity by measuring its (weighted) centrality in the network of skills.

However, the literature also shows that the value of skills (e.g. Anderson, 2017 or Neffke, 2019) is relative as it depends on how they are being combined. So, does a skill not have the same value for all workers? We hypothesise that the value of each skill indeed depends on its complements, as proposed by Stephany (2021): One and the skill can have a different economic value depending on the skills it is combined with. To test this assumption, we compare the economic value of individual skills in combination with different skill bundles.

Method

Online Labour Platform Data

The data for this analysis stems from one of the most popular online freelancing platforms, also referred to as online labour markets - short OLM - as introduced by Horton (2010). These platforms are websites that mediate between buyers and sellers of remotely deliverable cognitive work (Horton, 2010). The sellers of work on OLMs are either people in regular employment earning additional income by moonlighting via the Internet as freelancers or they are self-employed independent contractors (Stephany et al., 2020). The buyers of work range from individuals and early-stage startups to Fortune 500 companies (Lehdonvirta & Corporaal, 2017). OLMs can be subdivided into microtask platforms, e.g, Amazon Mechanical Turk, where payment is on a piece rate basis and freelancing platforms, such as Fiverr, where payment is on an hourly or milestone basis. Between 2017 and 2020, the global market for online labour has grown approximately 50% (Stephany et al., 2021). OLM data allow us to monitor skills in a global workforce on a granular level and in near real-time. The data includes worker and employer location, project wages, previous income, and project-level skill requirements.

Building a Network of Skills

This work conceptualises the relationship between skills using a network approach. The data consists of a sample of 49,884 freelance projects posted between 2014 and 2021 with multidimensional skill requirements completed by U.S. based workers. Using the skill information, we construct a network in which 4,583 skills are represented as nodes that are connected by a link if a worker applies both in a freelance project. The links are weighted according to how often two skills co-occur, as illustrated in Figure 1.

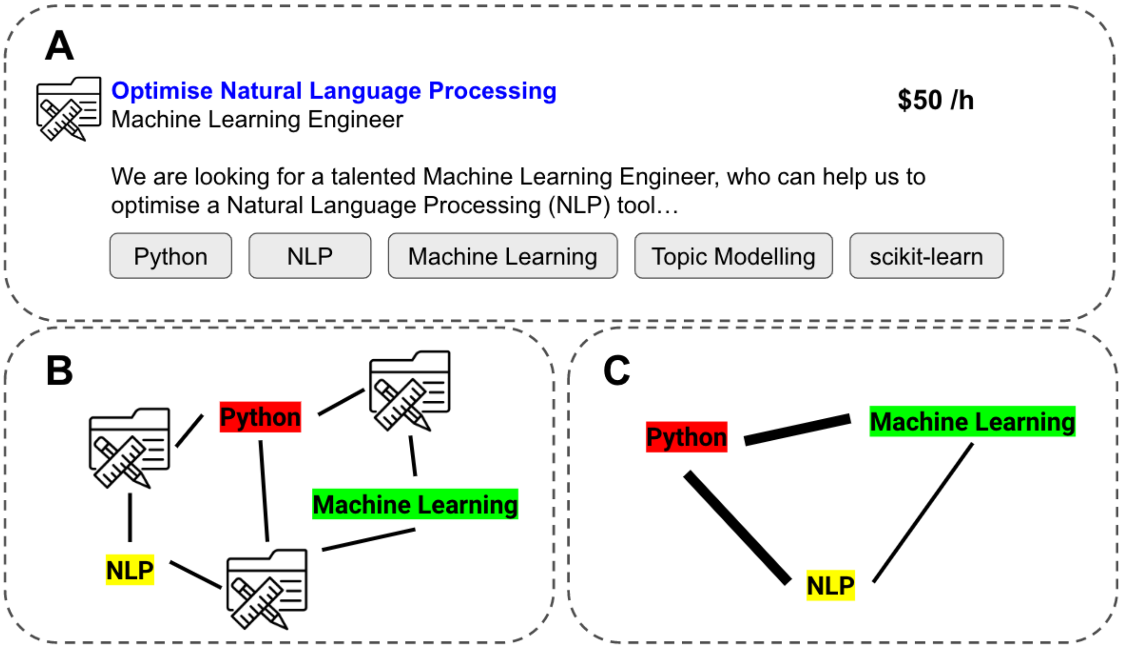

Figure 1: *From the bipartite network connecting freelancers and skills (B), we derive a unipartite network of skills (C) where two skills are connected if a worker applies both in a freelance project.*

Firstly, this skill network provides us with an endogenous categorisation of skills and illustrates the context dependency of human capital. Secondly, the value of each project, measured in USD per hour, allows us to statistically assess the value of individual skills.

Measuring the Value of Skills

The "premium" of each skill is calculated by comparing the mean rate of projects that require the respective skill with the mean of those projects that do not require it:

$$premium = \frac{\sum_{i=1}^{n} wage/n}{\sum_{j=1}^{m} wage/m} - 1$$

By calculating this premium, we derive the economic value of the 962 most popular skills, that is skills that occur in at least 20 projects[1]. Alternatively, we could derive the value of a skill via a regression model where the "price" of a skill is represented by the coefficient of a respective skill dummy. Both metrics of price and premium are highly correlated with a pearson correlation coefficient of 0.68. While the price metric allows us to control for differences in worker experiences and time effects, it is less intuitive to interpret compared to the skill premium. We use the price metric in our regression model for additional robustness checks.

[1] This popularity threshold has been determined empirically, as illustrated in Figure A1 in the Appendix.

Lastly, with our metric of skill value at hand, we investigate how the variance in values can be explained by the features proposed in our hypotheses. We build a regression model that allows us to test our assumptions about what drives skill prices. We include the following features:

1. Supply (Number of workers commanding respective skill)
2. Demand (Number of projects requesting respective skill)
3. Community (Dummy for the seven skill communities)
4. Complementarity (Weighted pagerank centrality)

$$premium_j = \beta_0 + \beta_1 * supply_j + \beta_2 * demand_j + \beta_3 * community_j + \beta_4 * complementarity_j + e_j,$$
$$j \in 1, \ldots, 962$$

We use a weighted pagerank centrality measure as our proxy for a skill's complementarity. We borrow this concept from various network science applications. For the example of aviation markets (Chung et al., 2020) or citation networks (Ding, 2011) it has been shown that the complementarity of nodes, such as relevant airport hubs or popular scholars, can be described by a weighted form of eigenvector (often pagerank) centrality. In our scenario of estimating the complementarity of skills, we assume that both the eigenvector centrality and the respective economic value of adjacent skills that can be reached from a specific skill node matter for our complementarity. We modify the pagerank centrality formula:

$$PR_i = \frac{1-d}{n} + d \sum_{j=1}^{n} \frac{PR_j}{c_j}$$

by adding a normalised weight v (the economic value of each skill):

$$PR_i = \frac{1-d}{n} + d \sum_{j=1}^{n} \frac{PR_j * v_j}{c_j}$$

Measuring the Complementary Value of Skills

For the representation of skill complementarities, we employ a network structure of how skills are combined with each other. This structure also tells us something about the relationship between skills and workers. Skills that are very distant to each other in the network are not frequently applied by the same worker. This representation should give us some indication about the costs and benefits of complementarity. Accordingly, we assume that it is most beneficial and least costly, in terms of economic value, to combine skills that are close-by in the skill space. To test this hypothesis, we represent skills and workers in a simplified fashion. Just as we attribute skills to specific communities, we use the same structure to group workers by their most frequent type of skill. We then calculate the value of each skill for the respective group of workers. This allows us to compare the value of one and the same skill in context with different other skills, e.g., the value of Python for workers (and skill bundles) from the domain of Design with the value of Python for workers from the domain of Admin.

Results

Creating a Network of Skills

We begin modelling the complex relationship of skills by creating a network, as described in the Method section. Figure 2A illustrates the resulting network of skills ("skill space"). We use the Louvain method for community detection (Blondel et al., 2008). Skills cluster into seven communities according to their field of application. We label the clusters based on the type of most prevalent skills in each of the communities: "Finance & Legal, "Software & Tech", "Marketing", "Design", "Audio & Video", "Writing", and "Admin". In addition to the community labels, Figure 2A also highlights three of the most prevalent skills for each of the communities. The most prevalent skills highlighted here reveal that the network is based on skill similarity. For example, it is worth noticing that the skill Creative Writing is adjacent to the Design community while it has a very high distance from the writing skill French Translation. Another example of this type is the Audio & Video skill Motion Graphics, which is much closer to the Design skill Drafting then to the same cluster skill Music Composition.

To a certain extent these skill communities overlap with the occupation taxonomy provided by the freelance platform (see Figure A2), i.e., the majority of skills in the Design community also fall into the occupation category Design & Creative. However, other communities, such as Admin distribute more broadly across occupations of Writing, Admin & Support, and Legal. This is in line with findings by Anderson (2017) and matches the categorisation of online labour market skills by Kässi & Lehdonvirta (2018).

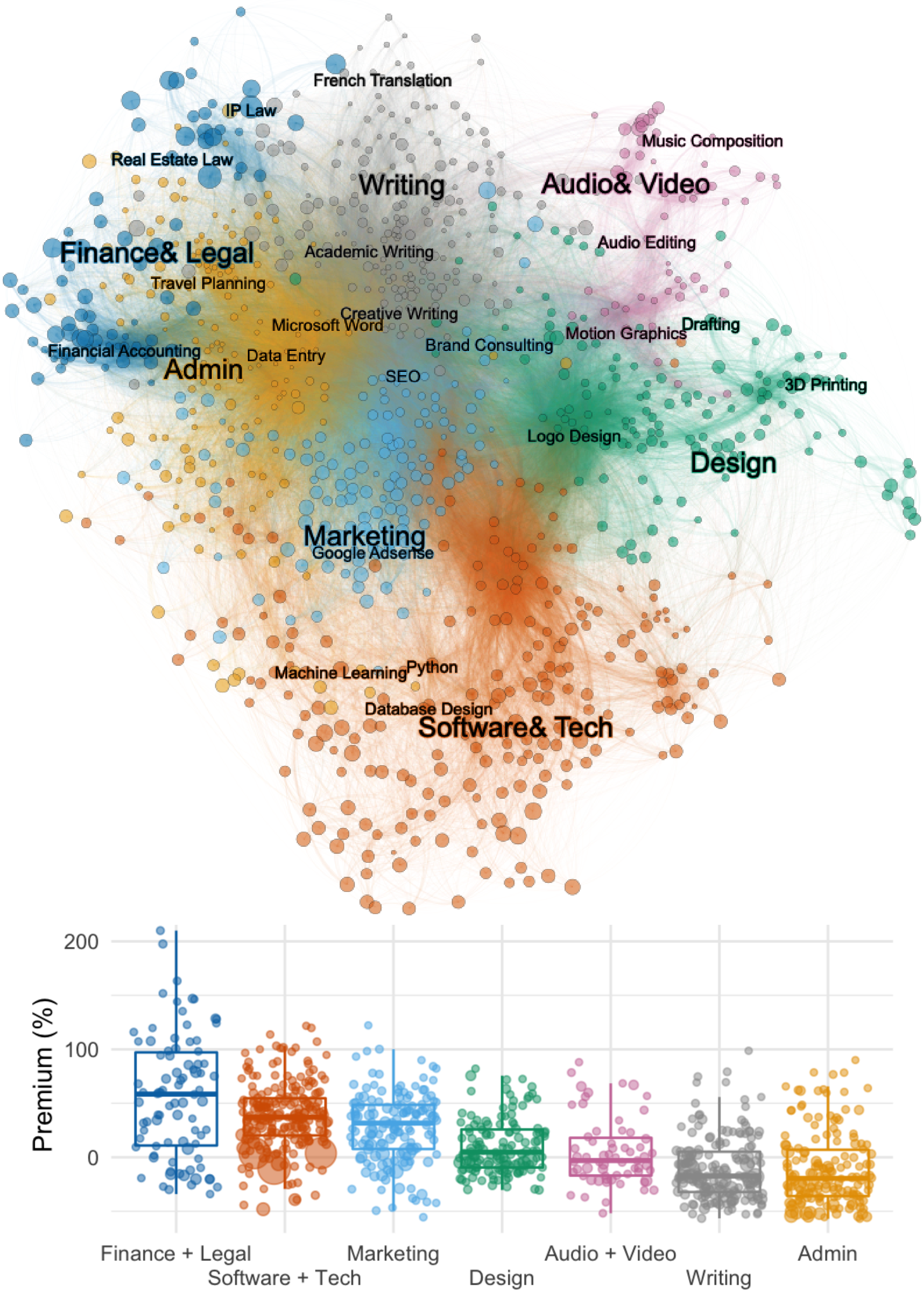

Figure 2 (A) *Skills cluster into seven groups according to their field of application. We label the clusters based on the type of most prevalent skills in each of the communities: "Finance & Legal, "Software & Tech", "Marketing", "Design", "Audio & Video", "Writing", and "Admin".* ***(B)*** *Valuable skills - the node size represents the premium of each skill - are not distributed at random across the skill space.*

In contrast to the official occupation taxonomy, our clustering allows us to fully group all skills into communities based on their actual application. In the next stage of the analysis this helps us to evaluate the contribution of community membership to a skill's economic value.

Explaining the Value of a Skill

In a second step, we estimate the individual value of the 962 most popular skills as outlined in the Method Section. We apply both the price and premium evaluation of each of the 962 skills (Table A1 shows the top 20 and bottom 20 skills by premium). Skill price and premium are strongly correlated across application clusters (with a pearson correlation of 0.69), indicating that they both consistently measure a similar concept. However, in contrast to the skill price, the premium does not take into account other factors that could explain a project wage, such as worker experience or time components. Nevertheless, the strong correlation between the two metrics encourages us to continue working with the price premium as our preferred measure, as it is easy to interpret.

The premia attached to each skill in Table A1 are derived from actual market demand. While the exact monetary values are specific to one online labour platform in the U.S. they allow us to make quantitative comparisons between skills. We also observe that price and premium are strongly correlated across skill domains. We see skills from the domain of Admin work with relatively low skill premia and prices, while skills from Software & Tech. or Finance & Legal show higher values. It is worth noticing that the correlation between the two metrics is lowest for skills from Finance & Legal work. Our assumption here is that other factors than the presence of a specific skill determine project wages, such as the experience of the worker.

Figure 1A reveals that valuable skills - the node size represents the premium of each skill - are not distributed at random across the skill space. Instead, they seem to have distinct characteristics and a distinct positioning in the skill space. Valuable skills are also not equally distributed across skill communities, as Figure 1B shows. Skills in the domain of Finance & Legal have, on average, a significantly higher premium that skills in Marketing, which have in return a higher premium than skills in Admin. This finding points towards a confirmation of our second hypothesis.

To further test our hypotheses, we run multiple regression models explaining the skill premium (and price), as described in the Method Section. In a stepwise fashion, our models include supply (number of projects), demand (number of workers), skill communities, and centrality measures (degree and weighted pagerank centrality) of the respective skill. The results of the regression analysis are shown in Table 3.

			Premium			Price
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Market						
Supply	-8.54***	-86.78***	-50.98***	-56.36***	-29.93***	-0.16***
Number of worker (log)	(1.04)	(8.39)	(7.46)	(7.69)	(5.43)	(0.04)
Demand		75.22***	42.14***	44.61***	21.87***	0.12***
Number of projects (log)		(8.01)	(7.10)	(7.14)	(5.17)	(0.04)
Community						
Reference: "Software & Tech"						
Audio & Video			-30.90***	-28.45***	15.22***	0.10***
			(4.70)	(4.77)	(3.74)	(0.03)
Design			-26.22***	-25.23***	7.13**	0.03
			(3.63)	(3.64)	(2.85)	(0.02)
Marketing			-8.59**	-11.74***	14.57***	0.06***
			(3.51)	(3.69)	(2.65)	(0.02)
Writing			-40.90***	-40.87***	15.84***	0.09***
			(3.47)	(3.46)	(3.15)	(0.02)
Finance & Legal			16.54***	16.94***	24.37***	0.03
			(4.18)	(4.17)	(3.03)	(0.02)
Admin			-45.09***	-46.27***	4.80	0.05***
			(3.47)	(3.49)	(3.02)	(0.02)
Complementarity						
Degree Centrality (log)				6.54***		
				(2.40)		
Weighted Pagerank (Premium)					1.92***	
					(0.06)	
Weighted Pagerank (Price)						2.52***
						(0.10)
Constant	53.14***	59.14***	70.35***	54.48***	20.05***	-1.45***
	(4.73)	(4.57)	(4.31)	(7.25)	(3.55)	(0.11)
Observations	962	962	962	962	962	962
Adjusted R^2	0.06	0.14	0.37	0.37	0.67	0.45
Note:			**p<0.1, **p<0.05, ***p<0.01*			

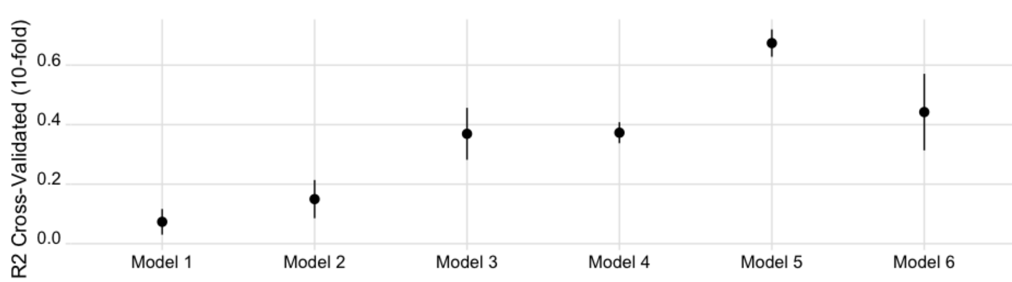

Table 2 (A) *The value of a skill, measured by skill premia in models 1-5 is determined by supply, demand, community, and complementarity.* ***(B)*** *The out of sample R-Squared of each model increases consistently when adding new controls.*

Model 1 and 2 show the effect of supply and demand on the value of skills. Here, we see, in alignment with our hypothesis that a skill with lower supply and higher demand have higher values on average. In model 3 we add dummies for the seven skill communities, while excluding the domain "Software & Tech." as a reference group. We see that the value of a skill clearly depends on the community it belongs to, as only skills from "Finance & Legal" work have a higher premium than our reference group while skills from all other domains are worth less. Skills from "Admin" work have the lowest values on average. In model 4 and 5 we add various network metrics. First, in model 4 we add the degree centrality (logged) of the skill to the model, which reveals that more centrally connected skills have higher premia on average. It is worth noticing that this effect is not driven by the overall occurrence of a skill, which is controlled for by introducing the number of projects. Finally, in model 5, we add the weighted pagerank centrality of the skill. This significantly improves the model performance. The complementarity of a skill plays a crucial role for its value, as proposed in our hypothesis of the complementarity. Model 6 repeats this full model setting for skill prices (and complementarity based on price information). It shows that our assumptions hold regardless of the skill evaluation metric. The lower panel of Table 2 shows that the out of sample R-Squared of each model increases consistently when adding new controls.

Interestingly, when introducing the pagerank centrality the coefficients for the skill communities change noticeably. Now skills from all domains, but for Admin, appear to have higher values than skills from Software & Tech. work. We think that this finding is driven by the fact that much of the value of skills in Software & Tech. is determined by their complementarity. This assumption is confirmed when turning to Figure 3, which plots the 962 skills across their complementarity and premia.

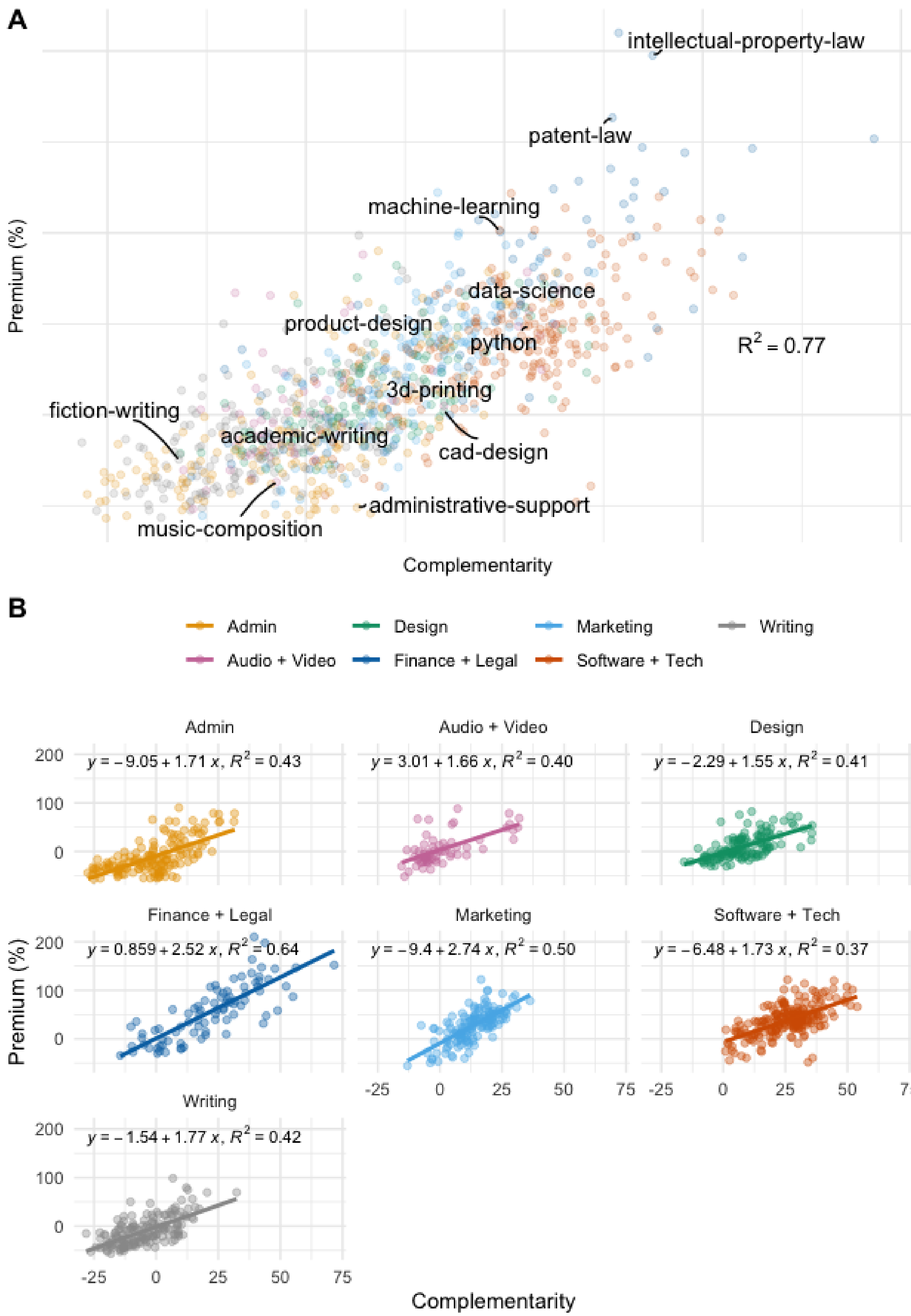

Figure 3 (A) *The most valuable skills are "hub skills" - We clearly see that skills from specific domains, such as "Software & Tech" or "Finance & Legal", exhibit significantly higher complementarity and higher skill premia accordingly.* ***(B)*** *The relationship between a skill's value and its complementarity is different across skill communities. For Finance and Legal or Marketing skills, a better complementarity translates much more strongly into higher skill values than in other skill communities.*

Here, we clearly see that the most valuable skills are skills with a strong complementarity (hubs). They are connected to many other skills of high values and centrality. These skills are not necessarily the most demanded skills, which is represented by the size of the bubbles. However, we clearly observe that skills from specific communities, such as Software & Tech or Finance & Legal, exhibit significantly stronger complementarity and higher skill premia accordingly.

At the same time, the results of our regression analysis confirm our assumption that the value of a skill is, besides market forces of supply and demand, influenced by both the membership to specific skill communities and the complementarity of the skill. However, the analysis also indicates that the benefits of a strong complementarity of a skill are not the same across skill communities, as illustrated in Figure 3B, which depicts the relationship between skills' premia and their complementarity for each of the seven skill communities separately. We see that both the baseline of skill values in each community (intercepts) and the benefit of a stronger complementarity (slope) are different across skill communities. In particular, the community of Finance & Legal and Marketing skills seems to be different to the other groups, as we see that the additional benefit of a stronger complementarity is significantly higher than in the other five communities. In Finance & Legal, furthermore, we observe a high initial level of skill values, which might indicate the strong signalling value of this community of skills.

Revealing the Complementary Value of Skills

Our analysis indicates that the degree of complementarity, in general, influences the value of a skill. The more likely it is for a skill to be combined with high-value complements, the higher its value. In addition to this general component of complementarity, we proposed that there is a specific aspect of complementarity that requires further investigation; similarity. As postulated by Stephany (2021) Stephany (2021), skills have specific synergies that arise from complementarity. The skill network that we created already captures these synergies in action, as it shows us which skills are most frequently combined with each other. One could use the structure of the network as a piece of evidence suggesting that there is no single best, most valuable, skill for everyone to learn, as we observe local hubs and clusters of skills rather than a centralised structure in which the entire network evolves around one "superstar" skill or a group of very central and valuable skills.

To further elaborate on this assumption, that the value of each skill depends on its complements, we examine skill-worker combinations and the value of one and the same type of skill in different combinations. Therefore, we have grouped workers, just like skills, into seven domains, which share the same labels as the skill communities. Workers are attributed to the domain from which they have their largest share of skills. This attribution quite clearly distinguishes workers, as for 78% percent of them at least half of their skills stem from one specific community. We then measure the premium of each of our 962 skills depending on the domain of the worker applying them[2]. The findings of this analysis are summarised in Figure 4.

[2] Here again, we decide to use a sample threshold of twenty observations of skill applications per worker domain. This leads to no observations for some of the worker-skill combinations, e.g., there is no significant application of Finance & Legal skills in the worker domain of Audio & Video.

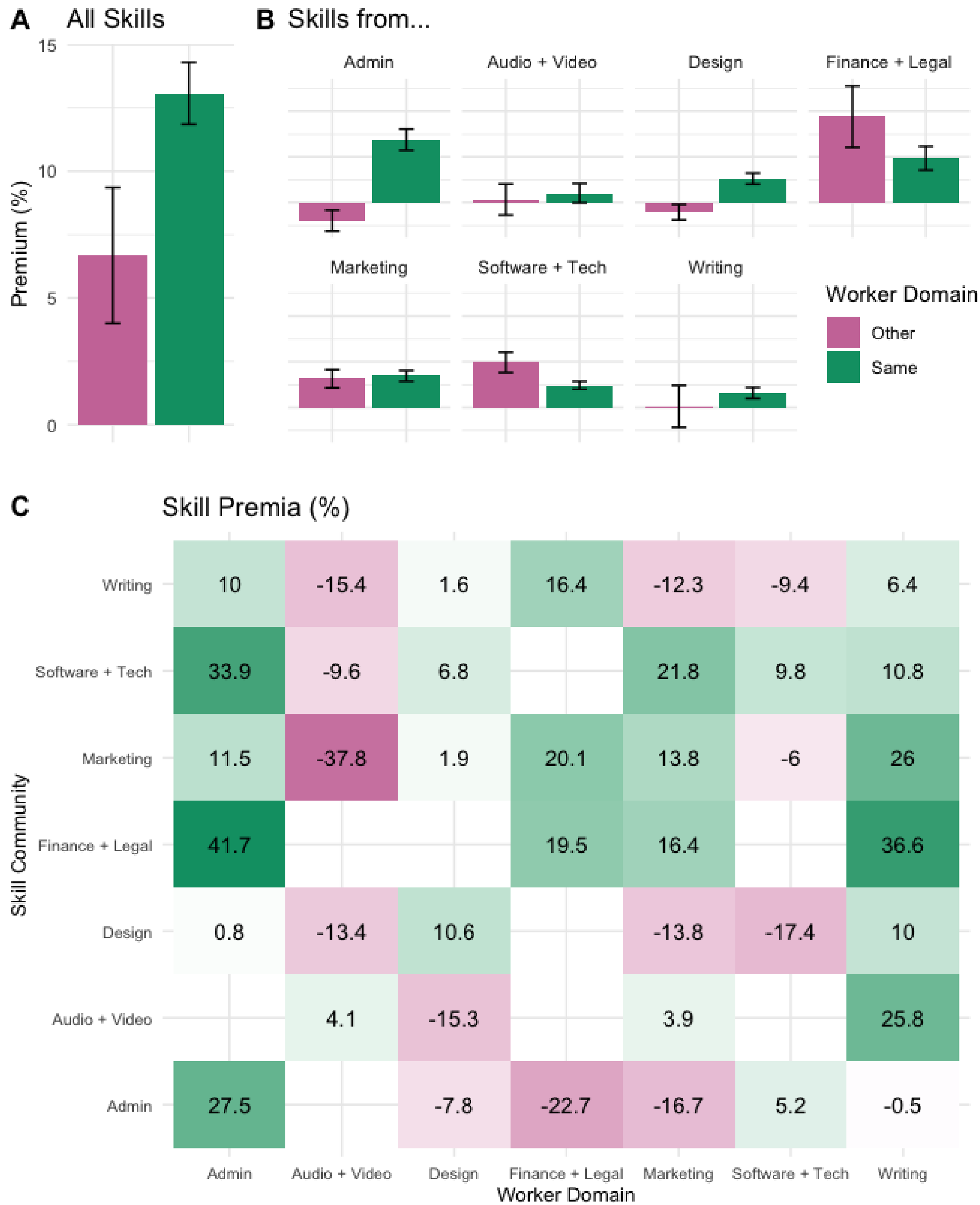

Figure 4 (A) *The premium of a skill depends on the skill background of the worker. In general, the skill premium is highest for workers that have mostly skills from the same domain.* ***(B)*** *One exception are Software and Tech skills, which seem to be more profitable for workers with skills from other domains.* ***(C)*** *The premium of a skill depends on the skill background of the worker (Worker Domain). Skills from the worker's domain are usually profitable, as the diagonal in the heatmap shows. Software and Tech skills are yet an exception, as they are profitable for workers from all domains with an average premium of 26%, and most profitable for workers from Admin with a premium average of 34%.*

The general finding is that skills are more valuable, when combined with other skills of the same type, as shown in panel A of Figure 4. Likewise, the diagonal of the matrix Figure 4C shows the premium of skills from all seven communities, when applied by workers who predominantly have skills from the same domain. However, for some skills this logic does not apply. For example, for skills from the community of Software & Tech., we observe that they are not only profitable throughout all worker domains (with the exception of Audio & Video) but that they are also most profitable for workers who are not active in the domain of Software & Tech. In comparison to tech workers, Software & Tech skills are twice as valuable for workers in Marketing and more than three times as valuable for Admin workers. This could indicate the high general purpose value of many skills from Software & Tech., in particular programming languages, which are beneficial in various combinations outside their original domain.

Working with Skill Values: The Case of AI-Skills

How could our metric of skill pricing be put into action? Here, we illustrate how the evaluation of skills could be applied to examples of AI skills. AI is widely considered to be a major breakthrough technology that is transforming the economy and society (OECD, 2021a). Hence, we assume that skills, which are required to work with AI technologies, should be of high economic value. We hand-labelled 43 skills[3] as AI skill[4], since we believe that they either describe an AI technology, such as Machine Learning, or a prerequisite for working with an AI technology, such as Python. Only 17 AI skills remain in our analysis, as many of them only occur in a few projects (see Table A2 in the Appendix). Our network-based clustering locates all AI skills in the cluster of Software & Tech work with the exception of Image Recognition, which is labelled as Design work.

As complements to breakthrough technologies, AI skills are thought to have high resilience towards automation, in general, which is a central aspect in discussions about the sustainability of work. Ideally, newly learned skills should have a low level of susceptibility towards automation. However, besides automation resilience, economic profitability is certainly another relevant aspect under which the "future readiness" of skill could be evaluated. A skill that is

[3] 'advanced-analytics', 'ai', 'algorithm-development', 'algorithms', 'analytics', 'apache-spark', 'artificial-intelligence', 'artificial-neural-networks', 'automation', 'automation-software-release', 'big-data', 'bot-development', 'c++', 'chatbot-development', 'cloud-computing', 'clustering', 'computer-vision', 'data-analysis', 'data-analytics', 'data-engineering', 'data-science', 'database-architecture', 'deep-learning', 'deep-neural-networks', 'ibm-watson', 'image-processing', 'imageobject-recognition', 'java', 'keras', 'machine-learning', 'machine-learning-model', 'natural-language-processing', 'natural-language-toolkit-nltk', 'neural-networks', 'pattern-recognition', 'python', 'python-script', 'robotic-process-automation', 'robotics', 'supervised-learning', 'tensorflow'.

[4] With the majority consensus among all authors.

highly valuable but most likely to be automated by machines or algorithms in the near future might not be an optimal destination for reskilling interests. However, a skill with low likelihood of automation but little economic reward might be similarly unappealing. For a comparison of our proposed skill evaluation metric, we calculate the automation probabilities of various skills using the automation probabilities developed by Frey & Osborne (2017). Frey & Osborne (2017) use four-digit SOC codes to identify occupations. We use the matching table developed by Braesemann et al. (2021) to map the four-digit SOC codes to the 98 platform occupations. After the matching, we calculate the automation probability for each skill via the following formula:

$$prob(skill\ x) = \sum_{j=1}^{m} prob(occupation_j)/n_j$$

Here the automation probability of $skill\ x$ is given by the average of the automation probabilities of all occupations $j = 1,...,m$ that contain $skill\ x$ weighted by my the share of projects n that are listed under the respective occupation. Our new metric allows us to compare both premium and automation probability for each skill, as illustrated in Figure 5.

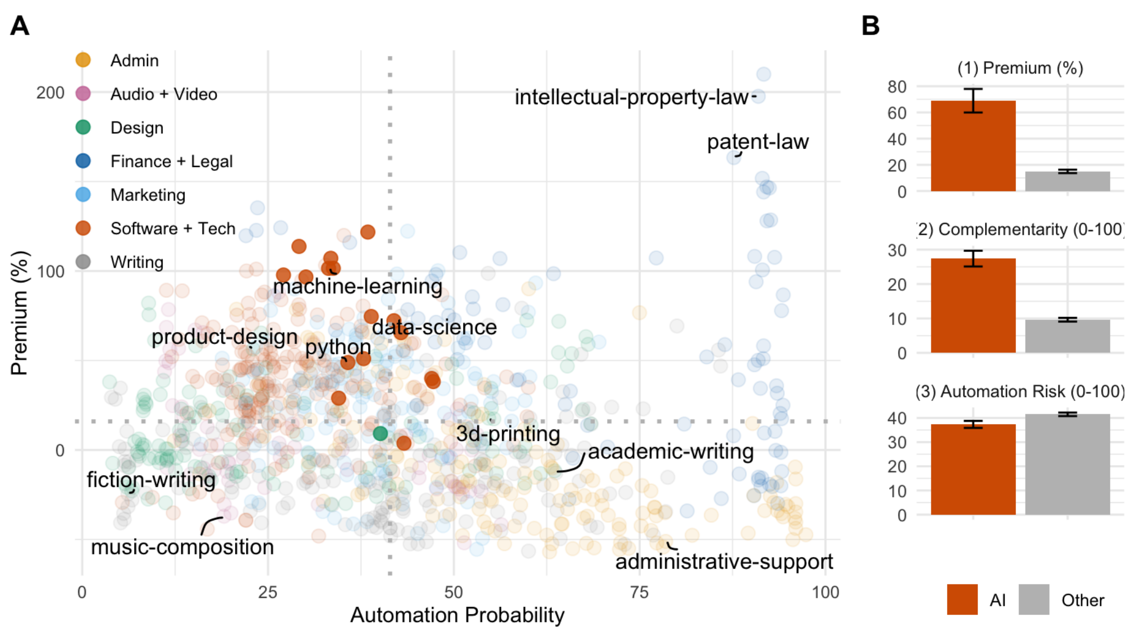

Figure 5 (A) *How does a skill's value relate to its risk of automation? We can compare the two metrics, which are relevant for assessing a skill's sustainability.. AI skills, like many other Software and Tech skills are mostly located in the top-left quadrant of the plot.* ***(B)*** *As the t-tests confirm: they are valuable and have below average automation probabilities, while having high complementarity.*

Our comparison shows that skills do not group randomly in the space of profitability and automation risk. Broadly speaking, we can see that profitability and automation risk are negatively related. In Figure 5A, we dissect the space into four quadrants: high value and low risk (top left), high value and high risk (top right), low value and low risk (bottom left), and low value and high risk (bottom right). Skill domains are distinctly distributed across this space. Many skills from the domain of Software & Tech. are in the top left quadrant, including most of the AI skills. Writing and Design skills are of low economic value but also show low risks of automation. Skills from Admin on the other hand have low economic value and high risks of automation. A group of skills that does not follow the negative relationship between premium and automation risk are certain Finance & Legal skills, which are scattered across different premium levels but all with high susceptibility to automation. Our group of AI skills are clearly different from the rest of all other skills, as they have above average premia and complementarity, and below average automation probabilities (Figure 5B).

Lastly, we compare the development of skill values over time. We use the example of AI skills to illustrate that the premia of skills are not stable over time, as they are influenced by changes in supply, demand, or complementarity. Figure 6 illustrates this change by comparing the premia of selected skills between 2014-2017 and 2018-2021.

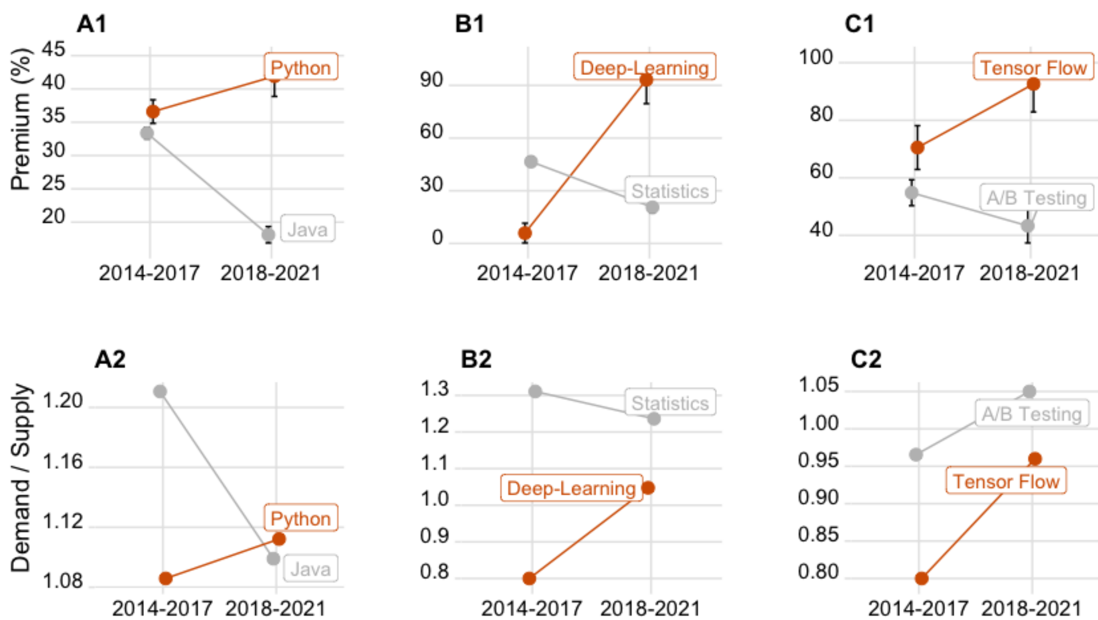

Figure 6 *For a set of AI and non-AI skills we observe that changes in the value of a skill are closely aligned with changing supply and demand. As demand (relative to supply) for Python **(A)**, Deep Learning **(B)**, and TensorFlow **(C)** has increased (lower panel), the premium for these skills rose (upper panel).*

In Figure 6A, we contrast the change in skill premia for two popular programming languages, Python and Java. While both languages start with more or less the same premia in the timespan of 2014-2018, they develop differently over time. While Java loses its premia significantly, Python gains in economic value. A driver of this trend could be a change in popularity, as Python has been rising to become THE data science super skill over the last decade (Grus, 2019). This was not the case for Java, as the comparison of demand (number of projects) versus supply (number of workers) indicates in the lower panel.

A similar logic applies to AI skills in the field of deep learning. In panels B and C, we show the changes in premia of Deep Learning and TensorFlow, one of the most commonly used software environments to programme deep learning applications. In contrast, we selected the skill Statistics and A/B Testing, one of the most frequently used conventional applications of statistics. Here, we see how the premium for deep learning and tensorflow increased significantly over the last 10 years, while the value of Statistics and A/B Testing stagnated. Similarly to the programming languages, this can be explained by a significant rise in the demand relative to supply for Deep Learning and TensorFlow compared to the skills of Statistics and A/B Testing.

Discussion

Conclusion

Technological change does not affect all tasks and occupations alike. New technologies require new skills while making others redundant. Technological change is not "skill-neutral". As a result, the skill composition of occupations changes. To respond effectively, workers need to reskill. However, the precise skill requirements of newly emerging jobs and economic benefits of learning a new skill are often uncertain and constantly evolving. It is therefore difficult for workers, employers, and policy-makers alike to build profitable and sustainable reskilling pathways. To address this uncertainty, we propose a method that attaches an economic value to skills based on market demand and supply as well as their relationship with other skills. Thereby, expanding the understanding of human capital formation in at least two ways. First, we reveal the economic value of individual skills and track their demand for and price over time. Moreover, we investigate what drives the value of individual skills. We find that the value of a skill depends on market forces of supply and demand. Secondly, we shot that a skill's premia is largely determined by the relationship to its complements. Skills that are frequently combined with many other valuable skills (high network centrality) tend to be of high value - they are skill hubs. We also show that the value of a skill is relative to its complements. Often skills are most valuable if they are applied together with a similar type of skill bundle.

For a set of 42 AI skills we illustrate how to apply our evaluation metric. We show that skills associated with AI, which is widely considered to be a major breakthrough technology, have a significantly stronger complementarity in the skill space and, hence, higher skill premia. In addition, we are able to contrast skill premia with automation probabilities - another relevant metric when assessing the sustainability of skills. This two-fold categorisation adds relevant nuance to debates about the future-readiness of learning a new skill. Ideally, reskilling efforts should focus on teaching skills that are both economically profitable and less exposed to risks of computerisation. AI skills fall into this category. Lastly, we track the development of skill values over time. We see that AI skills, such as Deep Learning or Python have been gaining in value significantly over the past years. Our model allows us to ascribe these changes to an increase in demand relative to supply.

This study has limitations. Firstly, we measure the economic value of skills *relative* to each other in a specific setting. This allows for comparisons between skills but statements on the absolute value of skills are not possible. In other words, the exact numerical value of a skill's value is not very meaningful in itself and might not be easily transferred to a different context.

While a statement such as "learning *deep-learning* will increase my hourly wage by 1,31 USD per hour" (see Table A2) would be problematic, stating that *deep-learning* might be a more valuable skill to learn than for instance *data-analytics* is in line with our approach. Secondly, we empirically find that the economic value of a skill is relative as it depends on how it is combined with other skills. While this makes intuitive sense and is highly relevant for reskilling purposes, it was outside the scope of this paper to develop a theoretical framework and empirical strategy to explain the how and why of these mechanisms. Lastly, this work is based on data from online freelancing which comes with several advantages but also pitfalls (see Background section for details). Most notably, online freelancing is limited to fully digital professional services. Therefore, several (often manual) occupations are missing entirely. Moreover, relative to the size of the overall workforce, rather few people work via online freelancing platforms. However, with adequate data access, our methodology could be extended to other data sources, such as online job advertisements or online career platforms, covering large parts of the traditional labour market.

Policy Implications

Our work on categorising and evaluating skills allow for multiple advances in understanding labour market developments. It can help to establish a taxonomy of skills, understand their application and individual complementarity, and enable automated, individual, and far-sighted suggestions on the value of learning a new skill in a future of technological disruption. Hence, policy recommendations are manifold.

First, reskilling institutions, like the European Centre for the Development of Vocational Training, could be the main beneficiaries of this highly individualised data. The high granularity of online generated data allows us to describe skill profiles of individual workers and track their development over time. It also enables reskilling institutions to assess the individual complementarities of learning a new skill. Via online generated labour market data, workers with the need to reskill could insert their current skill profile, be located in the landscape of skills, and receive targeted reskilling advice. This allows them to switch to more sustainable occupations that are closely related to their existing skill set with minimal reskilling effort. Via these individualised reskilling recommendations education providers and vocational training organisations could address the urgent need for individualised solutions in adult reskilling. Furthermore, the continuous "pricing" of skills over time allows reskilling practitioners to monitor the development of skill values and advise workers on which reskilling to "invest" in.

Secondly, official occupational and skill taxonomies could be improved with near real-time online generated data. As technology creates the demand for novel skills, new occupational clusters can emerge quickly and official taxonomies, such as the European Skills, Competences, and Occupations (ESCO) begin lagging behind. This is bad news for both firms and workers, as professional training providers find it hard to "speak" with the same language as market demand. Online generated data, on the other hand, stems from most recent market development and allows an identification of new occupational clusters including in-demand skills. Data-driven and near-real time taxonomies could complement conventional classifications. An immediate contribution to current policy efforts would be the continuous (re-)classification of AI and "green" skills or jobs, as the "twin-transition" has been identified as a catalyst for active labour market policies (OECD, 2021b).

Outlook

This work uses data from online labour markets. The major advantage of this data source is the availability of granular information on demand, supply, and wages, hence the value of skills. On the other hand, this data covers only one segment of the labour market. However, the methodology presented here, could be easily adapted to analyse other data sources covering larger parts of the labour market such as for example job vacancy platforms (e.g. Indeed) or social career platforms (e.g. LinkedIn). Our results have implications for the debate on successful reskilling strategies in times of dynamic technological change. We find that in-demand skills pay off and that they can be identified. This suggests that individualised skill-centred reskilling pathways could represent a promising avenue to mitigate skill mismatches. Yet, the quality of individualised skill recommendations would depend heavily on access to timely and granular data.

The European Commission has recognised the need and potential of a data-driven approach to closing the skill gap by bringing forward various legislative and policy proposals. The Pact for Skills (European Commission, 2020), launched in 2020, for example, aims to maximise the impact and effectiveness of skills investment, with a particular focus on upskilling and reskilling in the vocational training sector. For a successful implementation of the Pact, two aspects are crucial. Firstly, industry needs for specific skills must be made explicit, and secondly, the unique training history of workers needs to be acknowledged. Online generated data of worker profiles presents a promising approach to monitoring occupation taxonomies and skill requirements via online labour platform data. This is very much aligned with Europe's interest in further building out their skill foresight (French Presidency of the & Council of the European Union, 2022) via skill anticipation and the support for career transitions. It can offer targeted and near-real time reskilling advice to workers, regarding both industry needs and the worker skills required to fulfil them. This action could support the Commission's proposals (European Commission, 2021) for recommendations on individual learning accounts and on micro-credentials fostering skill-by-skill learning of professionals instead of traditional certification.

Similarly, the Commission's 2022 Data Act (European Commission, 2022) has identified the importance as well as the complications of accessing business (and platform) data in the interest of the public, while acknowledging the protection of businesses' interests. However, the retrieval and usage of private sector data, such as online labour market or job vacancy data, by public body institutions, is not necessarily enabled under the new legislation. Enforced sharing of private sector data requires the ex-ante proof of a "public emergency", and the current modes of automated data retrieval, such as web-scraping, could be prohibited by the Data Act if they were to be interpreted as coercive or deceptive according to Article 11. In light of this well-intended but potentially contradictory proposal to current EU data legislation, future amendments need to seek an agreement that gives public bodies acting in the interest of the public the right to access data (including via modes of web-scraping).

Our investigations on the complex ecosystem of skill formation show that online data can be a valuable tool for designing sustainable reskilling policies. To leverage the full potential of this resource the legislation needs to make public interest its focal point, allowing data access via web-scraping, while enabling strategic public–private partnerships to release the full potential of online generated data for the benefit of society.

CRediT authorship contribution statement

Fabian Stephany: Conceptualization, Methodology, Software, Validation, Formal Analysis, Investigation, Data Curation, Writing - Original Draft, Writing - Review & Editing, Visualisation, Supervision, Project Administration, Funding Acquisition. **Ole Teutloff**: Conceptualization, Software, Formal Analysis, Investigation, Data Curation, Writing - Original Draft, Writing - Review & Editing. **Vili Lehdonvirta**: Conceptualization, Writing - Review & Editing.

References

Acemoglu, D., & Autor, D. (2011). Skills, Tasks and Technologies: Implications for Employment and Earnings. In *Handbook of Labor Economics* (Vol. 4, pp. 1043–1171). Elsevier. https://doi.org/10.1016/S0169-7218(11)02410-5

Acemoglu, D., & Restrepo, P. (2018). The Race between Man and Machine: Implications of Technology for Growth, Factor Shares, and Employment. *American Economic Review*, *108*(6), 1488–1542. https://doi.org/10.1257/aer.20160696

Aggarwal, I., & Woolley, A. W. (2013). Do you see what I see? The effect of members' cognitive styles on team processes and errors in task execution. *Organizational Behavior and Human Decision Processes*, *122*(1), 92–99.

Alabdulkareem, A., Frank, M. R., Sun, L., AlShebli, B., Hidalgo, C., & Rahwan, I. (2018). Unpacking the polarization of workplace skills. *Science Advances*, *4*(7), eaao6030. https://doi.org/10.1126/sciadv.aao6030

Allinson, C. W., & Hayes, J. (1996). The cognitive style index: A measure of intuition-analysis for organizational research. *Journal of Management Studies*, *33*(1), 119–135.

Altonji, J. G. (2010). Multiple skills, multiple types of education, and the labor market: A research agenda. *American Economic Association, Ten Years and Beyond: Economists Answer NSF's Call for Long-Term Research Agendas*.

Anderson, K. A. (2017). Skill networks and measures of complex human capital. *Proceedings of the National Academy of Sciences*, *114*(48), 12720–12724. https://doi.org/10.1073/pnas.1706597114

Angrist, N., Djankov, S., Goldberg, P., & Patrinos, H. A. (2019). *Measuring Human Capital* (SSRN Scholarly Paper No. 3339416). https://doi.org/10.2139/ssrn.3339416

Autor, D. H. (2015). Why Are There Still So Many Jobs? The History and Future of Workplace Automation. *Journal of Economic Perspectives*, *29*(3), 3–30. https://doi.org/10.1257/jep.29.3.3

Autor, D. H., & Handel, M. J. (2013). Putting tasks to the test: Human capital, job tasks, and wages. *Journal of Labor Economics*, *31*(S1), S59–S96.

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

Beaudry, P., Green, D. A., & Sand, B. M. (2016). The Great Reversal in the Demand for Skill and Cognitive Tasks. *Journal of Labor Economics*, *34*(S1), S199–S247. https://doi.org/10.1086/682347

Blondel, V. D., Guillaume, J.-L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, *2008*(10), P10008.

Borghans, L., Duckworth, A. L., Heckman, J. J., & Ter Weel, B. (2008). The economics and psychology of personality traits. *Journal of Human Resources*, *43*(4), 972–1059.

Börner, K., Scrivner, O., Gallant, M., Ma, S., Liu, X., Chewning, K., Wu, L., & Evans, J. A. (2018). Skill discrepancies between research, education, and jobs reveal the critical need to supply soft skills for the data economy. *Proceedings of the National Academy of Sciences*, *115*(50), 12630–12637. https://doi.org/10.1073/pnas.1804247115

Bowles, S., Gintis, H., & Osborne, M. (2001). The determinants of earnings: A behavioral approach. *Journal of Economic Literature*, *39*(4), 1137–1176.

Braesemann, F., Stephany, F., Teutloff, O., Kässi, O., Graham, M., & Lehdonvirta, V. (2021). *The polarisation of remote work* (arXiv:2108.13356). arXiv. https://doi.org/10.48550/arXiv.2108.13356

Brynjolfsson, E., & McAfee, A. (2014). *The second machine age: Work, progress, and prosperity in a time of brilliant technologies*. WW Norton & Company.

Brynjolfsson, E., & Mitchell, T. (2017). What can machine learning do? Workforce implications. *Science*, *358*(6370), 1530–1534. https://doi.org/10.1126/science.aap8062

Calanca, F., Sayfullina, L., Minkus, L., Wagner, C., & Malmi, E. (2019). Responsible team players wanted: An analysis of soft skill requirements in job advertisements. *EPJ Data Science*, *8*(1), 1–20. https://doi.org/10.1140/epjds/s13688-019-0190-z

Chung, H. M., Kwon, O. K., Han, O. S., & Kim, H.-J. (2020). Evolving network characteristics of the asian international aviation market: A weighted network approach. *Transport Policy*, *99*, 299–313. https://doi.org/10.1016/j.tranpol.2020.09.002

Collins, A., & Halverson, R. (2018). *Rethinking education in the age of technology: The digital revolution and schooling in America*. Teachers College Press.

Dave, V. S., Zhang, B., Al Hasan, M., AlJadda, K., & Korayem, M. (2018). A Combined Representation Learning Approach for Better Job and Skill Recommendation. *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*, 1997–2005. https://doi.org/10.1145/3269206.3272023

De Mauro, A., Greco, M., Grimaldi, M., & Ritala, P. (2018). Human resources for Big Data professions: A systematic classification of job roles and required skill sets. *Information Processing & Management*, *54*(5), 807–817. https://doi.org/10.1016/j.ipm.2017.05.004

del Rio-Chanona, R. M., Mealy, P., Beguerisse-Díaz, M., Lafond, F., & Farmer, J. D. (2021). Occupational mobility and automation: A data-driven network model. *Journal of The Royal Society Interface*, *18*(174), 20200898. https://doi.org/10.1098/rsif.2020.0898

Ding, Y. (2011). Applying weighted PageRank to author citation networks. *Journal of the American Society for Information Science and Technology*, *62*(2), 236–245. https://doi.org/10.1002/asi.21452

European Commission. (2020). *Pact for Skills*. https://ec.europa.eu/social/main.jsp?catId=1517&langId=en

European Commission. (2021). *Action to improve lifelong learning and employability*. European Commission - European Commission. https://ec.europa.eu/commission/presscorner/detail/en/ip_21_6476

European Commission. (2022). *REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on harmonised rules on fair access to and use of data (Data Act)*.

Frank, M. R., Autor, D., Bessen, J. E., Brynjolfsson, E., Cebrian, M., Deming, D. J., Feldman, M., Groh, M., Lobo, J., Moro, E., Wang, D., Youn, H., & Rahwan, I. (2019). Toward understanding the impact of artificial intelligence on labor. *Proceedings of the National*

Academy of Sciences, *116*(14), 6531–6539. https://doi.org/10.1073/pnas.1900949116

French Presidency of the & Council of the European Union. (2022). *Informal Meeting of Ministers for Employment and Social Policy (EPSCO)—French Presidency of the Council of the European Union 2022*. French Presidency of the Council of the European Union. http://presidence-francaise.consilium.europa.eu/en/news/informal-meeting-of-ministers-for-employment-and-social-policy-epsco/

Frey, C. B., & Osborne, M. A. (2017). The future of employment: How susceptible are jobs to computerisation? *Technological Forecasting and Social Change*, *114*, 254–280. https://doi.org/10.1016/j.techfore.2016.08.019

Grus, J. (2019). *Data science from scratch: First principles with python*. O'Reilly Media.

Heckman, J. J., Stixrud, J., & Urzua, S. (2006). The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior. *Journal of Labor Economics*, *24*(3), 411–482.

Hong, L., & Page, S. E. (2004). Groups of diverse problem solvers can outperform groups of high-ability problem solvers. *Proceedings of the National Academy of Sciences*, *101*(46), 16385–16389.

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

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

Ingram, B. F., & Neumann, G. R. (2006). The returns to skill. *Labour Economics*, *13*(1), 35–59. https://doi.org/10.1016/j.labeco.2004.04.005

Kässi, O., & Lehdonvirta, V. (2018). Online labour index: Measuring the online gig economy for policy and research. *Technological Forecasting & Social Change*, *137*, 241–248.

Lazear, E. P. (2004). Balanced Skills and Entrepreneurship. *American Economic Review*, *94*(2), 208–211. https://doi.org/10.1257/0002828041301425

Lehdonvirta, V., & Corporaal, G. (2017). Platform Sourcing: How Fortune 500 Firms are

Adopting Online Freelancing Platforms. *Oxford Internet Institute*.

Neffke, F. M. H. (2019). The value of complementary co-workers. *Science Advances*, *5*(12), eaax3370. https://doi.org/10.1126/sciadv.aax3370

OECD. (2021a). *AI and the Future of Skills, Volume 1: Capabilities and Assessments*. Organisation for Economic Co-operation and Development. https://www.oecd-ilibrary.org/education/ai-and-the-future-of-skills-volume-1_5ee71f34-en

OECD. (2021b). *Designing active labour market policies for the recovery* (p. 13).

Powell, W. W., & Snellman, K. (2004). The knowledge economy. *Annual Review of Sociology*, 199–220.

Ren, Y., & Argote, L. (2011). Transactive Memory Systems 1985–2010: An Integrative Framework of Key Dimensions, Antecedents, and Consequences. *Academy of Management Annals*, *5*(1), 189–229. https://doi.org/10.5465/19416520.2011.590300

Stephany, F. (2021). One size does not fit all: Constructing complementary digital reskilling strategies using online labour market data. *Big Data & Society*, *8*(1), 205395172110031. https://doi.org/10.1177/20539517211003120

Stephany, F., Dunn, M., Sawyer, S., & Lehdonvirta, V. (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. https://doi.org/10.1111/tesg.12455

Stephany, F., Kässi, O., Rani, U., & Lehdonvirta, V. (2021). Online Labour Index 2020: New ways to measure the world's remote freelancing market. *Big Data & Society*, *8*(2), 20539517211043240. https://doi.org/10.1177/20539517211043240

Stephany, F., & Luckin, R. (2022). *Is the workforce ready for the jobs of the future? Data-informed skills and training foresight*. Bruegel.

Waters, K., & Shutters, S. T. (2022). Impacts of Skill Centrality on Regional Economic Productivity and Occupational Income. *Complexity*, *2022*, e5820050. https://doi.org/10.1155/2022/5820050

Weeden, K. A. (2002). Why Do Some Occupations Pay More than Others? Social Closure and Earnings Inequality in the United States. *American Journal of Sociology*, *108*(1), 55–101. https://doi.org/10.1086/344121

Willis, R. J. (1986). Wage determinants: A survey and reinterpretation of human capital earnings functions. *Handbook of Labor Economics*, *1*, 525–602.

Woolley, A. W., Chabris, C. F., Pentland, A., Hashmi, N., & Malone, T. W. (2010). Evidence for a collective intelligence factor in the performance of human groups. *Science*, *330*(6004), 686–688.

Appendix

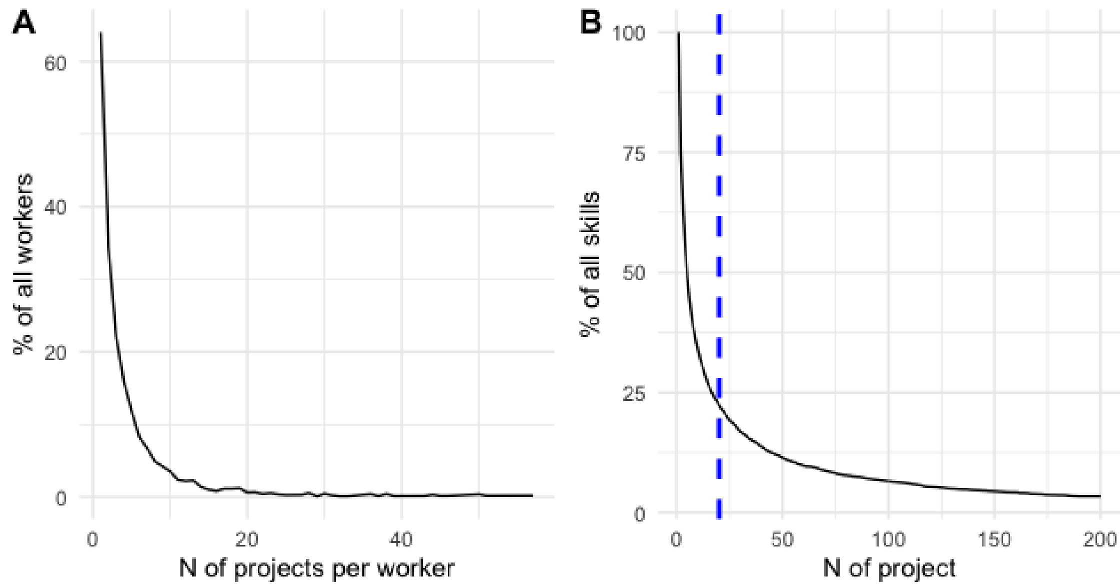

Figure A1 *(A) The minority of workers (4%) has had more than 10 projects and more than half of the workers had only two projects or less. (B) For skills, we see that the largest share of skills only occurred in a few projects, only 22% of all skills were applied to 20 projects or more.*

Table A1 *The top and bottom 20 skills are ranked by their premium. Many of the top ranked skills stem from Finance & Legal while bottom ranked skills are from Admin work.*

Rank	**Skill**	**Premium (%)**	**Price (USD/h)**	**Complementarity**	**Automation Probability**
1	intellectual-property-law	197,59	1,62	43,70	90,96
2	patent-law	163,35	1,52	38,63	87,66
3	international-tax-law	151,77	1,56	71,58	91,49
4	corporate-law	146,99	1,55	42,38	92,10
5	non-disclosure-agreements	146,41	1,11	56,25	92,45
6	international-law	144,14	1,21	47,74	91,68
7	financial-projection	135,29	1,58	38,38	23,58
8	contract-law	128,49	1,34	34,38	92,62
9	trademark-consulting	127,98	1,32	41,40	92,13
10	financial-forecasts	124,19	1,54	31,20	26,70
11	financial-modelling	122,69	1,62	45,13	23,42
12	branding-strategy	122,23	1,82	16,59	43,78
13	cloud-computing	121,75	1,44	25,85	38,43
14	docker	119,86	1,39	44,49	35,20

15	financial-plans	119,72	1,56	41,26	22,03
16	venture-capital-consulting	115,95	1,96	40,05	35,99
17	deep-learning	113,73	1,31	32,67	29,16
18	business-coaching	110,33	2,10	23,81	50,80
19	employment-law	108,16	1,31	52,30	90,47
20	legal-consulting	107,64	1,14	40,85	93,15
…	…	…	…	…	…
943	teaching-english	-49,60	0,75	-19,85	64,30
944	administrative-support	-51,10	0,87	6,46	78,50
945	article-spinning	-51,32	0,78	-13,10	46,31
946	email-handling	-51,78	0,86	1,06	72,43
947	video-upload	-51,87	0,58	-14,24	51,39
948	virtual-assistant	-51,99	0,84	3,65	77,30
949	customer-service	-52,21	0,83	1,32	67,83
950	chat-support	-52,38	0,91	-15,61	56,81
951	customer-support	-52,41	0,89	-0,83	63,19
952	translation-spanish-english	-52,48	0,75	-19,75	44,12
953	news-writing-style	-52,50	0,70	-14,58	45,17
954	active-listening	-52,78	0,83	-26,38	72,98
955	data-entry	-54,26	0,65	8,14	78,18
956	call-handling	-54,71	0,79	-5,16	68,87
957	telephone-skills	-55,02	0,80	0,49	73,51
958	phone-support	-55,14	0,87	-2,40	65,98
959	forum-posting	-55,51	0,67	-13,06	53,08
960	online-help	-55,58	0,92	-22,58	72,29
961	english-tutoring	-56,53	0,69	-18,25	60,05
962	order-processing	-56,92	0,72	-25,20	62,88

Table A2 *Of the 42 initially identified AI skills 17 remain after filtering for a minimum of 20 occurrences. They are ordered by their premium..*

Rank	Skill	Premium (%)	Price (USD/h)	Complementarity	Automation Probability
1	cloud-computing	121,75	1,44	25,85	0,38
2	deep-learning	113,73	1,31	32,67	0,29
3	tensorflow	107,10	1,46	49,71	0,33
4	natural-language-processing	101,52	1,21	28,80	0,34
5	machine-learning	101,45	1,50	24,37	0,33
6	database-architecture	97,77	1,17	40,92	0,27
7	artificial-neural-networks	96,66	1,21	37,04	0,30
8	artificial-intelligence	74,57	1,17	23,80	0,39
9	algorithms	72,21	1,36	24,70	0,42
10	data-science	65,51	1,19	24,91	0,43
11	automation	51,01	1,36	21,28	0,38
12	python	48,84	1,08	28,04	0,36
13	data-analysis	39,91	1,15	18,11	0,47
14	analytics	38,18	1,28	21,54	0,47
15	java	28,94	1,05	29,71	0,34
16	image-processing	9,16	0,95	7,00	0,40
17	c++	3,74	1,06	27,19	0,43

Calculating the price of a skill

The 962 linear regression models - one for each skill - explain a project's rate with

1. the year the project has been carried out in (measure by yearly dummies, 2014 - 2020),
2. the occupational category the project falls into (measured by twelve categories[5]),
3. the experience of the worker (measured in number past projects),
4. the occurrence of a specific skill (measured by a skill dummy).

[5] "Sales & Marketing", "Web, Mobile & Software Dev", "Writing", "Design & Creative", "Translation", "IT & Networking", "Data Science & Analytics", "Admin Support", "Customer Service", "Legal" "Accounting & Consulting", "Engineering & Architecture".

Each of the 962 most popular skills is considered individually as an explanatory feature in the linear regression. The beta coefficients of each of the 962 skill regressions allow us to determine the added value of an individual skill according to the following formula:

$$log(project\ wage_k) = \beta_0 + \beta_1 * year_k + \beta_2 * occupation_k + \beta_3 * experience_k + \beta_4 * skill_j + e_k,$$
$$k \in 1,..., n\ and\ j \in 1,..., 962$$

As the dependent variable, the project wage in USD per hour, is log-transformed, we need to perform an exponential transformation on the coefficient of the skill dummy β_4 to calculate the added value of the individual skill in USD per hour. Performing this model 962 times for each skill provides us with added values in USD per hour for each skill.

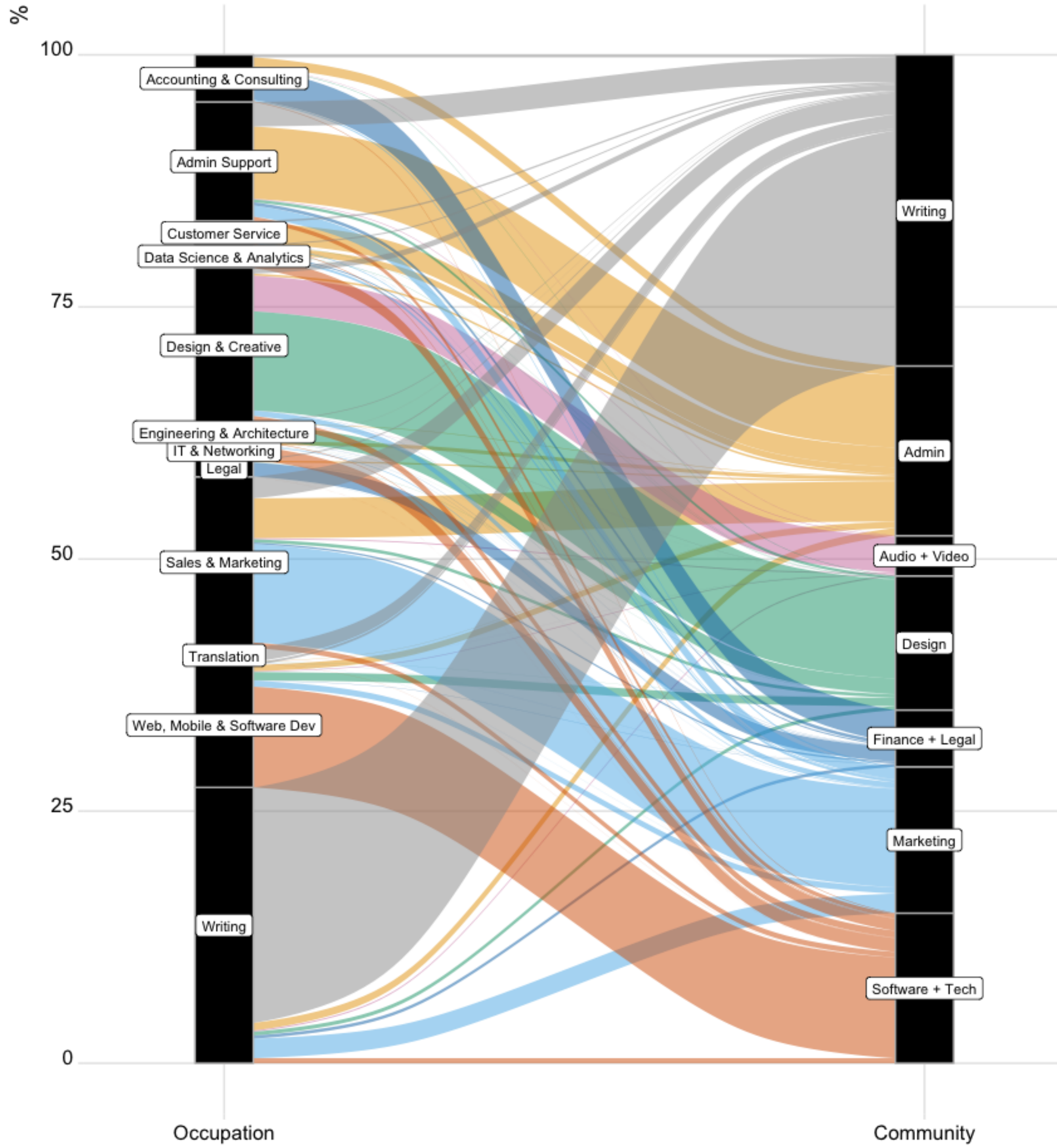

Figure A2 *Skill application clusters largely overlap with the occupation taxonomy provided by the freelance platform, i.e., the majority of skills in the application cluster of "Design" also fall into the occupation category "Design & Creative".*